Universität Regensburg

A Comparison of Nudging and Boosting for Privacy during Web Browsing

Masterarbeit im Fach Medieninformatik

am Institut für Information und Medien, Sprache und Kultur (I:IMSK)

Vorgelegt von:	Anna-Marie Ortloff
Adresse:	
E-Mail (Universität):	
E-Mail (privat):	
Matrikelnummer:	
Erstgutachter:	PD Dr. David Elsweiler
Zweitgutachter:	Prof. Niels Henze
Betreuer:	PD Dr. David Elsweiler
Laufendes Semester:	4. Semester M.Sc. Medieninformatik
Abgegeben am:	10.11.2020

Contents

List of Figures

List of Tables

Zusammenfassung

In dieser Arbeit wurde untersucht, wie sich *Nudges* und *Boosts* auf die Privatsphäre beim Surfen im Internet auswirken. Es handelt sich dabei um zwei verschiedene Ansätze, wie Verhaltenänderungen herbeigeführt werden können. Dazu wurde eine naturalistische dreiwöchige Studie durchgeführt, bei der Daten über das Surfverhalten von 69 Teilnehmern mit einer *Browser*-Erweiterung aufgenommen wurden. Die Teilnehmenden waren in drei Gruppen eingeteilt, und je nach Gruppe waren sie in der zweiten Studienwoche entweder *Nudges*, *Boosts* oder, in einer Kontrollgruppe, keiner Intervention ausgesetzt. Die Ergebnisse zeigen die Schwierigkeit, bei naturalistischen Daten von einer relativ geringen Anzahl von Teilnehmenden trotz des, durch die Erhebungsmethode bedingten Rauschens in den Daten, Effekte festzustellen. Es konnte, je nach gewählter Methode der Analyse, nur teilweise eine Änderung der Privatsphäre festgestellt werden. Zwei Variablen standen stellvertretend für die nicht direkt messbare Privatsphäre beim Surfen im Internet. Die durchschnittliche Anzahl der Anfragen an Drittanbieter war während der ersten Woche höher als sowohl während der zweiten, als auch der dritten Woche der Studie. Graphisch betrachtet, traf dies vor allem für diejenigen Teilnehmer zu, die *Boosts* ausgesetzt waren. Ein ähnliches Muster konnte auch für die durchschnittliche Änderung der Anzahl an Cookies pro Webseite festgestellt werden. Graphisch betrachtet, war diese für Teilnehmer in der *Nudge*-Gruppe während der ersten Woche höher als für beide anderen Gruppen, und während der zweiten und dritten Woche dann niedriger. Signifikante Änderungen des Verhaltens im Internet selbst, im Sinne von vermehrt veränderten Einstellungen, oder dem Besuch anderer Arten von Webseiten, konnten nicht festgestellt werden.

Abstract

This thesis explores the effect of boosts and nudges, two different approaches meant to achieve behavioral change, on browsing privacy. Sixty-nine participants partook in a three-week naturalistic study, whereby data on their browsing behavior was collected by a browser extension. The participants were divided into three groups, which were exposed either to boosts, nudges, or in the case of the control group, to nothing, in the second week of the study. The results reflect the difficulty in detecting effects amid the noise in naturalistic data from a relatively small sample of participants. Depending on the method of analysis, a change in browsing privacy could only be detected partially. Two measures were used as proxies for browsing privacy, since it is not possible to measure this directly. The average amount of third party requests was higher during the first week of the study, than during the second, as well as the third week. In a visualization, this was mostly the case for the participants in the boost condition. There was a similar pattern for the average change in cookies. Visually, it was higher during the first week for participants in the nudge condition, than for both other conditions. During the second and third week, the average number of changes was lower. The browsing behavior itself did not change significantly: For example, it was not possible to detect more changes to settings or a different amount of visits to certain categories of websites under the influence of boosts or nudges.

Task Statement

People spend a lot of time online and online services collect large amounts of data from them. The General Data Protection Regulation (GDPR), which has been in effect since May 2018, regulates privacy online and offers more choices and possibilities for users to take control of their data, but these options are not necessarily used (Utz et al., 2019). Possible consequences of the loss of privacy online are targeted advertising and discrimination based on user profiles (Datta et al., 2015; Mikians et al., 2012). Boosting and nudging are two competing strategies which can be used to induce behavioral change in users (Grüne-Yanoff & Hertwig, 2016), and which have also been employed in the domain of privacy (Acquisti et al., 2017). Nudges exploit users' cognitive biases, while boosts aim to support them in their decision making process, for example by providing additional information (Grüne-Yanoff & Hertwig, 2016).

Work in the domain of privacy often suffers from only measuring behavioral intentions and not the behavior itself (Lowry et al., 2017). This is especially worrying because users' intentions and actions concerning privacy do not necessarily match. This phenomenon is called the privacy paradox (Norberg et al., 2007). To overcome this limitation, this thesis aims to compare the effect of nudges and boosts on online browsing behavior by deploying a browser extension to capture naturalistic behavior and expose participants to either boosts or nudges during a three-week study period.

The steps which are necessary to reach this goal are the following:

- Collect data to use in boosts
- Design nudges and boosts based on literature and previous work

- Implement a browser extension for Chrome/Firefox to use in a naturalistic study. This includes:
  - Collecting information on participants' browsing behavior while preserving their privacy as much as possible
  - Categorizing visited websites as a measure to preserve anonymity
  - Accumulating privacy related information on website visits
  - Implementing a database and an Application Programming Interface (API) to connect to the database from the extension
  - Presenting nudges or boosts to participants
- Recruit a sufficient number of participants to take part in a study
- Conduct a study with the extension
- Analyze and interpret the results

1. Introduction

Users spend a large amount of time online (Kemp, 2019). This provides online services with many opportunities to collect data from and about them. While personal data is also being collected and used offline, such as through the means of loyalty cards, the internet offers more extensive possibilities for data collection, makes it easier to share this data, and offline and online data can also be combined (Geronimo, 2017). When online privacy is compromised, users can suffer targeted advertising (Datta et al., 2015) or discrimination based on information available about them (Mikians et al., 2012). For example, females receive less ads encouraging them to start high paying jobs than males (Datta et al., 2015). Users are concerned about their privacy online (Kokolakis, 2017), but they do not always act upon their concern. They may disclose information online, and later come to regret that disclosure, after it is too late to prevent it (Wang et al., 2011). This discrepancy between intentions and behavior is termed the privacy paradox (Norberg et al., 2007).

Changing privacy related behavior could provide a solution for this dilemma, by enabling users to preserve their privacy more, if they so wish. Two possible strategies to achieve behavior change are boosts and nudges. Both approaches build upon the theory of bounded rationality (Grüne-Yanoff & Hertwig, 2016), but they differ otherwise. Boosts aim to provide users with additional information or a better environment to support them in their decision making process, while nudges attempt to take advantage of users' cognitive biases to subconsciously influence them to behave differently. Both approaches have already been applied to promote privacy conducive behavior (e.g. Acquisti et al., 2017; Zimmerman et al., 2019b; Egelman et al., 2009; Ortloff et al., 2020).

Much previous work on using nudges or boosts to promote privacy has been conducted on laptop or desktop computers and concerning online behavior, but often,

studies were not carried out in a naturalistic setting. This is generally a problem in the domain of privacy (Lowry et al., 2017). It is hard to actually measure privacy related behavior in practice, which is why studies about privacy often measure self-reported behavior, not observed behavior (Baruh et al., 2017). In light of the privacy paradox, it is not clear how valid such findings are.

Similarly, boosts and nudges are often compared, but usually only theoretically, not in practice (Grüne-Yanoff & Hertwig, 2016). Zimmerman et al. (2019b) did so in a study using different nudges and boosts to try to get users to visit less privacy invasive sites to answer their health questions. However, this was a lab-based study in a controlled environment in the domain of health-related search (Zimmerman et al., 2019b), so the observed effects may not reflect everyday user behavior.

Thus, this master thesis aims to explore the possibilities of heightening users' browsing privacy through boosts and nudges. To provide a realistic measure of user behavior, a naturalistic study was conducted over three weeks. The following research questions will be assessed in the course of this thesis:

- **RQ 1:** Do boosts/nudges change users' behavior and preserve their privacy more?
- **RQ 2:** Do boosts change users' knowledge about privacy?

The remainder of this thesis is structured as following: Chapter 2 presents the foundation of previous work on which this thesis is based. Chapter 3 details the experiment which was conducted to answer the research questions introduced above, as well as preliminary work undertaken to prepare for this experiment. Statistical analysis of the collected data is performed in Chapter 4, while in Chapter 5 the results of these analyses are interpreted with respect to the research questions presented above. The relevance of the findings is then discussed in Chapter 6, taking into account limitations of the utilized methodology and connecting the findings to previous work. Finally, in Chapter 7, conclusions are drawn from this study, and possible avenues for future work are outlined.

2. Related Work

This chapter summarizes some of the related work in several domains which are relevant for this work. It gives a brief overview of definitions of privacy in general, and the current situation concerning online privacy. Methods of online tracking as a threat to online privacy are outlined, including the consequences of the loss of online privacy. Finally, nudges and boosts are presented as two methods to achieve behavioral change.

2.1. Privacy

Privacy is a complex construct, thus defining it is not straightforward; it has even been termed "elastic" (Margulis, 2011). One well-supported theory of privacy, that of Altman (1975), encompasses a wide array of interactions, because it is focused on social interactions in general (Margulis, 2011). In it, privacy is considered to be "the selective control of access to the self, involving dialectic, optimization, and multimodal processes" (Altman, 1977, p. 67). Dialectic processes encompass privacy being seen as a process of interaction with others, in which phases of openness and phases of closedness alternate (Altman, 1977). Optimization in conjunction with privacy means that rather than assume there cannot be too much privacy, in that more privacy is always better, Altman (1975) posits that the optimal level of privacy varies. There can be both too much privacy, this is termed "isolation", and too little privacy, "crowding" (Altman, 1977, p.67). Multimodal refers to Altman's claim that various methods and behaviors may be used to reach a personally optimal level of privacy (Altman, 1977). This does not only apply to different individuals, but may also differ across cultures (Altman, 1977).

Another widely known and supported approach, is that of Westin (1967), which focuses on information privacy (Margulis, 2011). Privacy is defined as "the claim

of an individual to determine what information about himself or herself should be known to others" and is relevant at multiple levels, the personal, group and organizational levels (Westin, 2003, p. 3). Westin (1967) postulates four states, or means of achieving privacy, namely solitude, intimacy, anonymity, and reserve, which are discussed by Margulis (2011). Solitude means not being monitored by others, and intimacy is when a close and open relationship is possible within a small group. Anonymity means not being identified or monitored in public, while reserve is about being able to reduce the disclosure of what is considered one's own information and data. Other actors have to respect this decision to limit disclosure. Similarly, the four functions, or purposes of privacy in Westin's theory are also discussed by Margulis (2011). Personal autonomy means not wanting to be manipulated by others, and emotional release means being able to step back from the responsibilities that a social life imposes on us. Self-evaluation is the process of "integrating experience into meaningful patterns and exerting individuality on events"(Margulis, 2011, p. 10). Finally, limited and protected communication refers to setting boundaries between oneself and others on the one hand, and having a protected setting to confide in those which are trusted on the other hand.

These two theories of privacy are reflected in different definitions throughout the literature. For example, the two kinds of privacy which Pötzsch (2009) distinguishes, are somewhat similar to the two theories presented above. Altman's theory is similar to what Pötzsch (2009) terms as privacy of the personal sphere, while Westin's privacy definition is close to what she calls information privacy, which focuses on the aspect of control over one's personal data. A more practically applicable taxonomy of information privacy, although it is called data privacy, comprises of four dimensions: purpose, visibility, granularity, and retention of collected data (Barker et al., 2009). This taxonomy is intended for use in a more technological context, e.g. for database management systems related research (Barker et al., 2009). The control aspect is also highlighted by Margulis (1977) (as cited by Margulis (2011)): "Privacy, as a whole or in part, represents control over transactions between person(s) and other(s), the ultimate aim of which is to enhance auton-

omy and/or to minimize vulnerability." (Margulis, 1977, p. 10). This definition was derived from literature on privacy (Margulis, 2011), and thus represents an integration of different definitions of privacy.

Since this thesis will conduct research about users' online browsing behavior with respect to privacy, a comparably narrow definition of privacy is appropriate. Thus privacy in this thesis means for a user to be able to control access to and distribution of their data by other parties, similar to the definition introduced by (Margulis, 1977) above.

2.2. The Current State of Online Privacy

The GDPR, which became effective in Europe on May 25th 2018, made the public more aware of privacy. Though this law certainly offers users more possibilities to control their data online (Sobolewski et al., 2017), it is not clear whether they take advantage of this, or whether the requests for consent with data collection are instead just a source of annoyance for users (Utz et al., 2019). The GDPR may have been somewhat effective in reducing the amount of third party activity, even without users taking action, but this change may have also been impacted by other factors (Sørensen & Kosta, 2019). Even if users do take actions and request their data, these requests are not honored by companies in many cases (Urban et al., 2019). The GDPR gives users the right to interact with the data collected about them and forces services to ask users for consent explicitly (Sobolewski et al., 2017). However, websites partially collect data or install trackers before the consent notice is displayed (Sanchez-Rola et al., 2019).

Nonetheless, during the process of data collection for the study described in this thesis, the German federal court ruled that preselecting more privacy-invasive choices in cookie notices, which requires users to take action to prevent cookies from being set, is illegal (Tageschau, 2020). Previous work showed that notices meant to inform users about data collection can be used to nudge users into accepting more purposes for the data that is being collected (Machuletz & Böhme, 2019).

Cookie notices are just one example of the difficulty of achieving both privacy and usability at the same time, since these two goals seem to contradict each other (Cranor & Garfinkel, 2004). Another example is the current implementation of the principle of notice and choice (Mysore Sathyendra et al., 2017) or notice and consent (Barocas & Nissenbaum, 2009), whereby users should be informed about what happens to their data, so they can make informed choices. In practice, privacy policies are the most common way to implement notice and choice, but users often do not understand them (Reidenberg et al., 2014) or do not read them at all (Obar, 2016). This means that privacy policies are not an effective tool to inform users.

Internet users are concerned about their privacy online, but that does not mean that they take actions to protect it (Kokolakis, 2017). This discrepancy is called the privacy paradox (Norberg et al., 2007) and has been extensively studied in previous work, both providing evidence supporting (e.g Taddicken, 2014; Carrascal et al., 2013), and contesting the notion of a privacy paradox (e.g. Tsai et al., 2011; Lutz & Strathoff, 2014). Kokolakis (2017) summarizes some possible explanations for the privacy paradox. One of these is privacy calculus, which assumes that users perform a trade-off between the benefits and risks of disclosing their data online (Jiang et al., 2013; Kokolakis, 2017). Users making such a trade-off have been observed both using quantitative (Jiang et al., 2013) and qualitative methods (Ortloff et al., 2020). Other possible explanations are based on the theories of cognitive biases and bounded rationality (Baek, 2014; Brandimarte et al., 2013; Kokolakis, 2017), among others.

2.3. Web Tracking

Data can easily be collected online, and it is easy to share these data or connect them with data collected offline (Geronimo, 2017). Web tracking means that personal information is collected about users' online behavior (Bujlow et al., 2017). Tracking encompasses the "capturing, tracing, observation, and analysis of users and their behavior in order to gain a comprehensive picture of them"(Pugliese, 2015, p. 367). The behaviors which are monitored range from users' search queries, the pages they

visit, the products they buy to which people they are in contact with (Bujlow et al., 2017). Ermakova et al. (2018) provide a structured review of relevant literature concerning web tracking.

The purposes of web tracking vary, and contrary to popular belief, it is not only used to target users for advertising (Bujlow et al., 2017). Other purposes include personalization in general (Sanchez-Rola et al., 2016), e.g. to provide customized content or search results, personalization of prices (e.g. Hannak et al., 2014), financial credibility judgement or web analytics, which informs web providers about their visitors (Bujlow et al., 2017). Additionally, governments can use web tracking for surveillance purposes, and tech companies may be legally required to provide data. For example, as is evident from Google's transparency report, from July 2019 to December 2019, they received around 11200 requests from the German government to reveal user identities, concerning around 18500 accounts[1]. For requests which had available compliance percentages, 70% of these requests were granted.

Before moving on to different methods used for web tracking, some terminology should be clarified. Third parties and first parties are frequently mentioned. When a user visits a certain website, this website is the first party, it is the one that the user directly interacts with. A third party in this context, is any other site, which receives data about the user, while they are interacting with the first party site (Pugliese, 2015). Third parties may be included because they provide images, other content or other functionality such as advertising, analytics, or social functionality like comment infrastructure (Mayer & Mitchell, 2012).

Bujlow et al. (2017) provide a summary of tracking methods. The earliest tracking methods were session-only methods, which enabled tracking only during a single browser session (Bujlow et al., 2017). These methods include the use of session identifiers stored in hidden fields on forms, or the explicit authentication of users by signing in to a service (Bujlow et al., 2017). Log-in could be enforced by restricting

[1]Current data are available for download at `https://transparencyreport.google.com/user-data/overview?user_requests_report_period=series:requests,accounts;authority:DE;time:&lu=user_requests_report_period&user_data_produced=authority:DE;series:compliance`, downloaded data set is included in additional material

access to functionality on a website to authenticated users, but this form of tracking was transparent, in that users knew they could be identified when signed in (Bujlow et al., 2017).

While they do not apply this categorization themselves, other work divides the remaining methods of web tracking into two groups (e.g. Pugliese, 2015; Mayer & Mitchell, 2012; Ermakova et al., 2018). In stateful tracking, the user is identified by data which is stored on the client-side, e.g. on their device. Stateful tracking includes the cache-based or storage-based methods described by Bujlow et al. (2017). Cache-based methods allow attackers to discover which sites were previously visited by a user by checking if elements from certain sites can be loaded from cache or have to be loaded from scratch (Bujlow et al., 2017). One method of storage-based tracking is the use of http-cookies, which are small data files placed on a user's device (Bujlow et al., 2017). There are different types of cookies and, while first party cookies can be necessary or useful, for example in authentication, third party cookies are considered to be a larger risk for privacy. They are cookies set by a different website from the one a user is currently visiting (Mayer & Mitchell, 2012; Pugliese, 2015). Cookies can be combined with methods utilizing invisible pixels, which can be used to send third party cookies to trackers' servers (Fouad et al., 2020). Cookies have since evolved to also include flash cookies, which can be accessed from multiple browsers and provide the capacity to store more data (Bujlow et al., 2017), and so called zombie cookies or super cookies, which are stored in multiple storages and can regenerate themselves, when they are deleted (Pugliese, 2015).

In stateless tracking, users, or rather browser instances, cannot be directly identified, but their identity can be narrowed down by measuring certain characteristics of the browser instance. When such characteristics are unique, which was the case for almost 90% of fingerprints in an evaluation of around 120,000 fingerprints (Laperdrix et al., 2016), it is possible to track a browser instance across multiple sessions. The data used in stateless tracking are revealed during communication and they may change at any time (Pugliese, 2015). These methods include different variants of fingerprinting (Bujlow et al., 2017). Fingerprinting works by using

characteristics such as the unique identifier of a device, versions of operating system or browser which are automatically transferred in requests, or installed fonts or display settings, which can be obtained through active queries for them (Bujlow et al., 2017).

Other tracking methods exist, and are continuously developed. They include exploiting biometric behavioral data, such as keystrokes dynamics, mouse movement or touch interactions (Pugliese, 2015). Recent work shows that cookie synchronization is very prevalent (Papadopoulos et al., 2019). This is a technique which aims to circumvent existing measures of protection, such as the same origin policy, and enables trackers to combine their data on a user, thus expanding their reach (Papadopoulos et al., 2019).

An early list of measures to counter web tracking includes the disabling of JavaScript and the use of adblockers as implemented measures, as well as the blocking of third party requests as a further possible measure, among others (Krishnamurthy et al., 2007). These approaches are examined with respect to their impact on functionality (Krishnamurthy et al., 2007). As is often the case, heightening privacy may limit usability, in this case cause websites to break.

Bujlow et al. (2017) provide a more current list of protection measures against online tracking. Many methods make use of blacklists containing trackers, and block certain content, such as content from third parties (Bujlow et al., 2017). Such methods include browser extensions used for adblocking, like Adblock Plus or uBlock Origin or more specifically tailored to protect privacy, such as Ghostery (Mazel et al., 2019). Methods blocking only third party content can be circumvented by forwarding a user directly to the third party site through popup windows or redirection (Bujlow et al., 2017). A large percentage of trackers are also not identified correctly by such lists (Fouad et al., 2020). Different privacy preserving extensions do not make use of blacklists, but rather use heuristics to detect trackers, e.g. Privacy Badger, or block all JavaScript, e.g. NoScript (Mazel et al., 2019). Other methods to protect privacy work by filtering and changing requests by routing them through a proxy, such as Privoxy, or by hiding the IP address, e.g. using anonymous proxy

servers, like Tor, or VPN services (Bujlow et al., 2017). Changing the browser settings to send a Do Not Track header is not very effective, because adhering to it is voluntary, so many websites do not (Gervais et al., 2017). Using the private browsing mode, which has different names, and is implemented differently, for different browsers is another possibility, but it does not live up to its name and there are misconceptions among users concerning what it means (Gao et al., 2014). Private browsing protects against attackers which have access to a user's computer, since they cannot see that user's browsing history etc (Bujlow et al., 2017), but contrary to popular belief, it does not prevent tracking (Gao et al., 2014). Privacy focused search engines, such as DuckDuckGo, protect privacy by not sending along the referrer header to pages visited from their search engine result page, among other things (Bujlow et al., 2017).

Finally, there are also tools which do not aim to protect users from tracking, but rather want to make them aware of the extent of web tracking. These include browser plugins such as Lightbeam for Firefox (Bujlow et al., 2017), which visualizes third parties and first parties, and which was intended for end users. It should be noted that Lightbeam was discontinued in 2019 (Joni & Neiman, 2019), and is now being maintained on Github[2]. Another example of privacy awareness related tools is $heriff[3], a browser extension for Chrome and Firefox, which enables users to check for price discrimination (Bujlow et al., 2017). MindYourPrivacy is another tool used to bring attention to privacy infringement by displaying sites which leak user data to trackers, among others (Takano et al., 2014). However, this tool was only used in research, and not deployed for use of the public, to the best of my knowledge. On the AmIUnique website[4], users can check whether their browser fingerprint makes them uniquely identifiable among visitors of this website. It should be taken into account that the percentage of fingerprints which are unique becomes lower, when comparing to a population larger than that of visitors of such a privacy related site (Gómez-Boix et al., 2018).

[2] https://github.com/princiya/lightbeam-we
[3] http://sheriff-v2.dynu.net/views/home
[4] https://www.amiunique.org/

2.4. Consequences of Loss of Online Privacy

There are numerous consequences to the undermining of privacy on the web. Tailored advertising has been reported to include ads that concern sensitive topics such as sexual orientation or health (Wills & Tatar, 2012). It must to be noted that this may not be due to information induced from behavior (Wills & Tatar, 2012). However, to people feeling embarrassed when encountering these ads, it may not make a difference (Wills & Tatar, 2012). Targeted advertising can also be discriminatory, such as in the case where ads encouraging users to seek highly paying jobs were displayed significantly more often to males than females (Datta et al., 2015).

Another kind of discrimination which has been frequently investigated is price discrimination, which means prices that are different because of user characteristics (Mikians et al., 2012). Price steering (Hannak et al., 2014) or search based discrimination (Mikians et al., 2012) is when users are lead towards different groups of items depending on their willingness to pay. Price discrimination can be based on the type of operating system or browser, a user's history of purchases, whether they have an account on a site (Hannak et al., 2014), or on location (Mikians et al., 2013). Signs of search based discrimination were also found when comparing trained personas, with one trained as budget-conscious, and the other as affluent (Mikians et al., 2012).

The large amount of data collected on individuals online can make it easier to steal identities using both data actively disclosed on the web and accumulated by web tracking (Bujlow et al., 2017). In a broader context, web tracking can be used to sway elections, as was investigated for the 2016 US election (Isaak & Hanna, 2018). News coverage on such cases makes the loss of online privacy public. Users' perceived privacy and security can influence their trust in (online) businesses and cause them to loose trust when they think their data is not handled with care and kept secure (Flavián & Guinalíu, 2006). Their privacy and security concern can also influence their purchasing behavior (Jibril et al., 2020). This can lead to financial losses for businesses (Wirtz et al., 2007). On the other hand, when users are allowed to control tracking, they allow tracking in some cases (Melicher et al., 2016). This ties in

with the aforementioned definition of privacy, whereby control over one's data is a crucial aspect of privacy.

2.5. Nudging and Boosting

Previous work on getting users to change their privacy related behavior frequently references the concept of nudges (Acquisti et al., 2017). Boosts are a different approach to inducing behavior change and are considered an alternative to nudges (Hertwig, 2017; Grüne-Yanoff & Hertwig, 2016). Both these approaches are based on the theory of bounded rationality, but differ in other aspects (Grüne-Yanoff & Hertwig, 2016).

These differences start with their respective theoretical foundation. Nudges are grounded in the Heuristics and Biases research program (Grüne-Yanoff & Hertwig, 2016). "Heuristics and Biases" interprets bounded rationality as a consequence of humans' flawed decision making processes (Kahneman, 2003). Systematic cognitive biases, such as overestimating the commonness of events which are easy to remember, are assumed to be the cause of bad decisions (Tversky & Kahneman, 1974; Kahneman, 2003).

There is some discussion on the exact definition of a nudge, with various definitions being criticized as being too vague (Grüne-Yanoff & Hertwig, 2016). Thaler & Sunstein (2008) describe nudges "as any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives"(p. 6). According to a more specific definition by Rebonato (2012), as described by Grüne-Yanoff & Hertwig (2016), a nudge tries to influence users to behave in a way that helps users reach their ultimate aim by using documented cognitive biases, but without using financial compensation as a motivation. Nudges do not change features for which "people have explicit preferences"(Grüne-Yanoff & Hertwig, 2016, p. 153), but rather change features that influence decision making implicitly and are not of direct importance to users. Finally, it should be easily possible to rescind any change in behavior induced by a nudge (Grüne-Yanoff & Hertwig, 2016).

Boosting draws on the theory of simple heuristics (Grüne-Yanoff & Hertwig, 2016), which assumes that bounded rationality stems from humans' use of heuristics to make not optimal but satisficing decisions (Gigerenzer, 2006). Decisions may be considered bad decisions from a certain perspective, but the simple heuristics framework does not attribute this to cognitive flaws, but rather to the circumstances in which decisions are made (Gigerenzer, 2006). These situational constraints, for example time limits, limited information or resources, or general uncertainty, are responsible for less than optimal decisions (Gigerenzer, 2006). Heuristics are a practical tool to help make decisions while balancing optimal results and constraints (Gigerenzer, 2006).

Boosts have been defined as "interventions that target competences rather than immediate behavior"(Hertwig & Grüne-Yanoff, 2017, p. 977), as opposed to nudges, which try to directly change behavior. They try to change the circumstances surrounding decision making. Competences or abilities acquired through boosts can be general in nature or only applicable to a certain domain (Hertwig & Grüne-Yanoff, 2017). To achieve this goal boosts can make use of strategies improving either individuals' thought processes directly or the environment in which they make their decisions (Hertwig & Grüne-Yanoff, 2017). Boosts try to enable people to make informed decisions which are good for them (Hertwig & Grüne-Yanoff, 2017). This means that the boost's goal needs to be known to the individual who encounters it (Hertwig & Grüne-Yanoff, 2017).

Some nudge definitions, being very general in nature, encompass all attempts to change behavior on a voluntary basis, where no financial incentives are present (Hertwig & Grüne-Yanoff, 2017). It becomes hard to distinguish between nudges and boosts, because there is an overlap between the two. Non-educative nudges can be classified clearly as nudges, according to the definition presented above (Hertwig & Grüne-Yanoff, 2017). Long-term boosts, which give access to new abilities applicable to a wide range of context, belong clearly to the boost framework (Hertwig & Grüne-Yanoff, 2017). However, it is hard to distinguish educative nudges, which provide information, for example in the form of a label or a warning, from

short-term boosts, which promote a competence only useful in a specific situation (Hertwig & Grüne-Yanoff, 2017). In the scope of this thesis, such interventions are considered to be more boosts than nudges, because they do not specifically exploit cognitive biases and this is seen as a central aspect of the nudge framework, as is evident from the more detailed definition of nudging above.

For the sake of completeness, it should be mentioned that both nudges and boosts focus on individual decision making processes. However, in some situations it may not be sufficient to only change individual behavior, but rather necessary to change norms and mechanisms governing social interaction to achieve the best outcome for society as a whole (Reijula et al., 2018). Reijula et al. (2018) call this approach the design approach.

There are some ethical concerns about nudges which are summarised by Renaud & Zimmermann (2018). These include the opacity especially of those nudges, which exploit subconscious decision-making processes. This means that human autonomy and agency are restricted. Other worries are that choice architects may not know what is best for nudgees, or may not agree with them, so nudges could go against the interest of those being nudged. Moreover, nudges may also have unforeseeable side-effects and may be mismatched to a certain group of users, who have different needs from the majority. Counterarguments of proponents of nudging are also presented by Renaud & Zimmermann (2018). Nudges are justified by the inescapability of choice architectures and, since some choice has to be made, it might as well be one that is intended to be beneficial. Nudging is also stated to relieve the cognitive load of making choices. Additionally, arguments concerning autonomy are questioned since it is not clear whether autonomous decision-making is always the solution for ethical issues. It is also claimed that since nudges explicitly do not include coercion, autonomy is not limited after all. In general boosts suffer less from ethical concerns than nudges, but the latter can also be implemented in an ethical way. Hertwig (2017) offers guidelines to determine when nudges are considered unethical and how this can be remedied.

2.6. Examples of Using Nudges and Boosts to Promote Privacy

Nudging as well as boosting has been investigated in the domain of security and privacy, but these approaches have sometimes been called by different names (Renaud & Zimmermann, 2018). For example, Calo (2014) distinguishes between three kinds of interventions aimed at behavior change: code, nudge, and notice. The former two represent forms of nudges, while the latter is essentially a boost, considering the definitions laid out in earlier sections. Thus, when presenting examples of nudges and boosts, these may not be called by that name in their source text.

According to Renaud & Zimmermann (2018), nudges in the security and privacy domain have mainly been used either for privacy preservation, e.g. leading people to install apps that do not request unnecessary permissions, or to increase security, e.g. to improve password strength. Acquisti et al. (2017) review previous work on nudging users to make better privacy related decisions in their Section 3. This includes making privacy policies more accessible, for example by showing privacy indicators in search engines (Tsai et al., 2011) or app stores (Kelley et al., 2013). Different forms of privacy indicators have been investigated frequently, for example using either the number of symbols (Egelman et al., 2009) or color in a traffic-light metaphor (Zimmerman et al., 2019a) to indicate the privacy level of a website. More unobtrusive forms of nudging were also investigated, such as reranking or filtering of results based on privacy (Zimmerman et al., 2019b). Some nudges also make use of humans' social nature and tendency to be influenced by their peers' behavior (Thaler & Sunstein, 2008). Amazon Mechanical Turk users were successfully nudged away from acceptance of a cookie, when informed that others like themselves had declined it as well (Coventry et al., 2016). Nudges of this kind were also proposed for use with social media (Ziegeldorf et al., 2016).

Boosts have also been used to make privacy polices more accessible. An early example of this was the Privacy Bird extension, where users were warned if a website's privacy policy did not conform to their expectations (Cranor, Guduru, & Arjula, 2006). Since this extension provided information on privacy practices, when users configured their settings, and also showed summaries of privacy policies in cases

of mismatch (Cranor, Guduru, & Arjula, 2006) it can be considered a boost. Privacy Bird was based on the Platform for Privacy Preferences Project (P3P), but this standard was never supported by a majority of websites and there was no means to ensure the correctness of P3P policies, so they were often not equivalent to human-readable policies (Cranor et al., 2008; Cranor, 2012). Support of P3P has since been discontinued (Cranor, Dobbs, et al., 2006; Microsoft, 2016). A similar vein of work presented summaries of information from privacy policies directly in users' current context (Ortloff et al., 2018, 2020). Boosts have also been used to promote secure actions on the internet, including actionable tips (van Bavel et al., 2019), even though the authors of this work refer to their approach as nudging. Zimmerman et al. (2020) evaluate several features of websites about health information with respect to privacy and correctness of information. Results from their study could also be used in boosts. They suggest that .gov and .org top level domains are good sources for information and do not infringe privacy as much as other kinds of websites (Zimmerman et al., 2020).

While boosting and nudging for privacy have mostly been investigated on desktop devices, there is also some work concerning mobile devices. It focuses largely on choosing apps (Choe et al., 2013; Alohaly & Takabi, 2016) or assisting users in deciding which permissions to grant to apps (Liu et al., 2016; Lin et al., 2014; Almuhimedi et al., 2015). Apart from interactions with apps, warnings concerning privacy and security during browsing on mobile devices have also been proposed (Maurer, 2010).

3. Experiment Design

Previous work has noted the importance of measuring actual user behavior in privacy research, and not only self-reported behavioral intentions (Lowry et al., 2017; Sotirakopoulos et al., 2011), since the latter are often not transferred to actions. This section describes the design of a naturalistic experiment which aims to measure the influence of privacy boosts and nudges on browsing behavior. For all tests conducted during the design phase of the experiment, significance was assumed at the .05 level and data analysis was conducted using Gnu R (R Core Team, 2019).

3.1. Proxies for Browsing Privacy

Browsing privacy is not a concept which can be directly measured, but several different proxies have been used for it in previous work (Mazel et al., 2019). When examining the effects of different protection techniques on privacy, some of the proxies used include those that were examined during the process of constructing boosts, see Section 3.2 for more information. These include the number of images, number of requests, both in general, and only considering first and third party requests, and number of cookies, among others (Mazel et al., 2019). For this study, two metrics, which have been used a lot, were chosen to represent browsing privacy: the number of new cookies on a given site (e.g. Mayer & Mitchell, 2012) and the number of third party requests (e.g Englehardt & Narayanan, 2016; Fruchter et al., 2015). Since the number of cookies was measured by taking the difference between current cookies and cookies from the previous website visit, this metric was adjusted to be the change in cookies. If cookies were deleted, this number could also be negative. Even though both of these metrics are not exclusively related to tracking, as cookies can be used to provide functionality and third party requests may also serve to load media, they are deemed appropriate proxies because they are

quite correlated to tracking (Mazel et al., 2019). Consequently, browsing privacy is assumed to be higher, when the number of third party requests and the changes to cookies are lower.

3.2. Choice of Boosts

The process to select appropriate boosts to use in the study consisted of multiple steps. First, an open source dataset was examined to look for correlations of website characteristics with tracking. Since the dataset only contained top level domains, several sites and subpages were crawled to determine whether found patterns were the same. The results were not conclusive, but recent literature showed promising directions for boosts and preliminary boosts were derived from literature. These were then tested for comprehensibility and effectiveness in a between-group survey, utilizing Prolific as a crowd-sourcing platform.

3.2.1. Examining Tracking and Website Characteristics in the Whotracksme Dataset

The publicly available Whotracksme dataset[1] was used to get a general understanding of which website characteristics correlate with tracking on this website. At the time of investigation data were available from May 2017 to January 2020.

The Whotracksme dataset contains a number of variables which are described in detail on github[2](Karaj et al., 2018). The sites.csv data from this source contains information on website characteristics, as well the number of trackers present on a specific site. This makes it possible to find out which website characteristics correlate with the presence of trackers. Since the variables used are all measured at the interval level, and the dataset is sufficiently large that normality can be assumed, Pearson correlations were calculated (Field et al., 2012). Table 3.1 shows the correlations of different variables with the number of trackers on a site, sorted by pearson's r, in descending order.

[1] see `https://github.com/cliqz-oss/whotracks.me`

[2] see `https://github.com/cliqz-oss/whotracks.me/tree/master/whotracksme/data` for a description of the variables

variable correlated with number of trackers	p	r	lower CI	upper CI
number of hosts	<.001	0.98	0.978	0.979
number of companies	<.001	0.97	0.971	0.972
number of xhr requests	<.001	0.6	0.593	0.601
number of iframes	<.001	0.59	0.582	0.591
proportion of pages with cookies	<.001	0.54	0.539	0.549
tracked	<.001	0.54	0.537	0.546
number of images	<.001	0.43	0.424	0.435
referer_leaked	<.001	0.43	0.42	0.432
number of requests through beacon API	<.001	0.42	0.413	0.424
proportion of pages with unique identifier in query string of tracker	<.001	0.4	0.396	0.407
number of stylesheets	<.001	0.35	0.34	0.351
referer_leaked_header	<.001	0.35	0.347	0.36
number of custom fonts	<.001	0.34	0.338	0.35
number of scripts	<.001	0.27	0.264	0.276
avg number of requests to tracker with tracking	<.001	0.19	0.188	0.2
proportion of pages with external blocking	<.001	0.17	0.165	0.178
referer_leaked_url	<.001	0.13	0.123	0.137
number of requests loaded via video or audio elements	<.001	0.12	0.109	0.122
avg number of requests to tracker	<.001	0.1	0.092	0.105
https	<.001	0.06	0.058	0.071
avg number of failed requests to tracker	<.001	0.05	0.042	0.055
requests associated with plugins, such as Flash	0.008	0.01	0.002	0.016
bandwidth usage of tracker	0.004	-0.01	-0.016	-0.003
popularity	<.001	-0.03	-0.039	-0.025

Table 3.1.: Pearson correlations of number of trackers with other variables from Whotracksme dataset

Cohen (1988) considers values of r above $|.5|$ to mark a large effect, and values above $|.3|$ to mark a medium size effect, as cited by Field et al. (2012). Naturally, the number of hosts and companies are extremely correlated with the number of trackers, since companies are the companies associated with the trackers and hosts are the domains a tracker used, but the number of trackers is also highly correlated with the proportion of pages where a cookie was sent by the browser. Similarly, the correlations with number of iframes, requests made from scripts (xhr requests), number of images, stylesheets, and fonts and number of requests to the beacon API, among others, were also above the threshold of $.3$ for medium sized correlations.

There is however, a limitation to this analysis. The Whotracksme dataset contains information about websites, but there is no separate information available for subpages of the same site. Website characteristics may vary between different sites. For example, the checkout page on Amazon may differ from the start page on the same website. Previous work showed that subpages generally have more tracking than landing pages (Urban et al., 2020). Consequently the website characteristics which correlate with tracking may differ for subpages.

3.2.2. Webcrawling

Since there is no information available in the Whotracksme dataset for website characteristics on subpages of a domain, as a second step, similar data were crawled from a set of websites and their subpages. The websites chosen were the most visited websites in Germany in February, according to the Alexa ranking (Alexa, 2020). The websites were limited to Germany because the study described in this thesis would take place in this country, so any boosts designed from the collected information should be relevant to Germany. The data used were simply the most recent available data at the time of investigation.

A first approach was to access the 500 most visited websites and evaluate them, then follow their links to subpages, eliminating double entries in the process, evaluating those, and finally to follow the links from the subpages and evaluate those pages which were two steps away from the starting point. This procedure was in-

tended to ensure that data for several subpages would be collected. Additionally, it aimed to evaluate a more diverse set of pages, not only the most visited ones, to forgo a bias stemming from the fact that these pages are frequently analyzed and may be optimized to conform with or subvert certain tracking prevention tools (Gervais et al., 2017). A trial account for Alexa was set up and used to manually access the 500 most popular websites. These were then used as input in a Python (version 3.7) script utilizing Selenium (Selenium, 2020) with a Firefox webdriver and a proxy to visit the websites and access their website characteristics. The collected characteristics were modeled after the variables from the Whotracksme dataset, focusing on those that correlated with the presence of trackers and those for which it was possible to implement collection quickly. This resulted in collecting the following variables for each of the crawled websites:

- full url
- domain url
- number of cookies
- number of session cookies
- number of iframe elements
- number of img elements
- number of visible words
- number of stylesheets
- number of video elements
- number of audio elements
- number of total requests
- number of requests for image
- number of requests for CSS
- number of requests for scripts (JS)
- number of requests for fonts

- number of requests for audio
- number of requests for video
- number of links

Originally, this script used parallel threads to save time, but on trial runs, it nevertheless took very long, and used too much working memory. This resulted in crashes of the system on which it was running, even when reducing the amount of threads used. Since some of the chosen websites had as many as 3000 links, following them all used too much memory and time, and was deemed outside the scope of this preliminary work.

In the end, the set of starting websites was reduced to the top 50 most visited websites in Germany (Alexa, 2020). Additionally, the number of links followed at each level in the hierarchy of pages was reduced to five, so the maximum number of pages evaluated per website in the starting set amounted to 31. This amount of subpages was not reached for all of the websites in the starting set, since crawling some pages resulted in errors and was not possible due to blocking of non-human access to the site. Even with this reduced number of sites, the script could only be run for one website from the starting set at a time, and the laptop running the script had to be rebooted between runs, to clear the working memory.

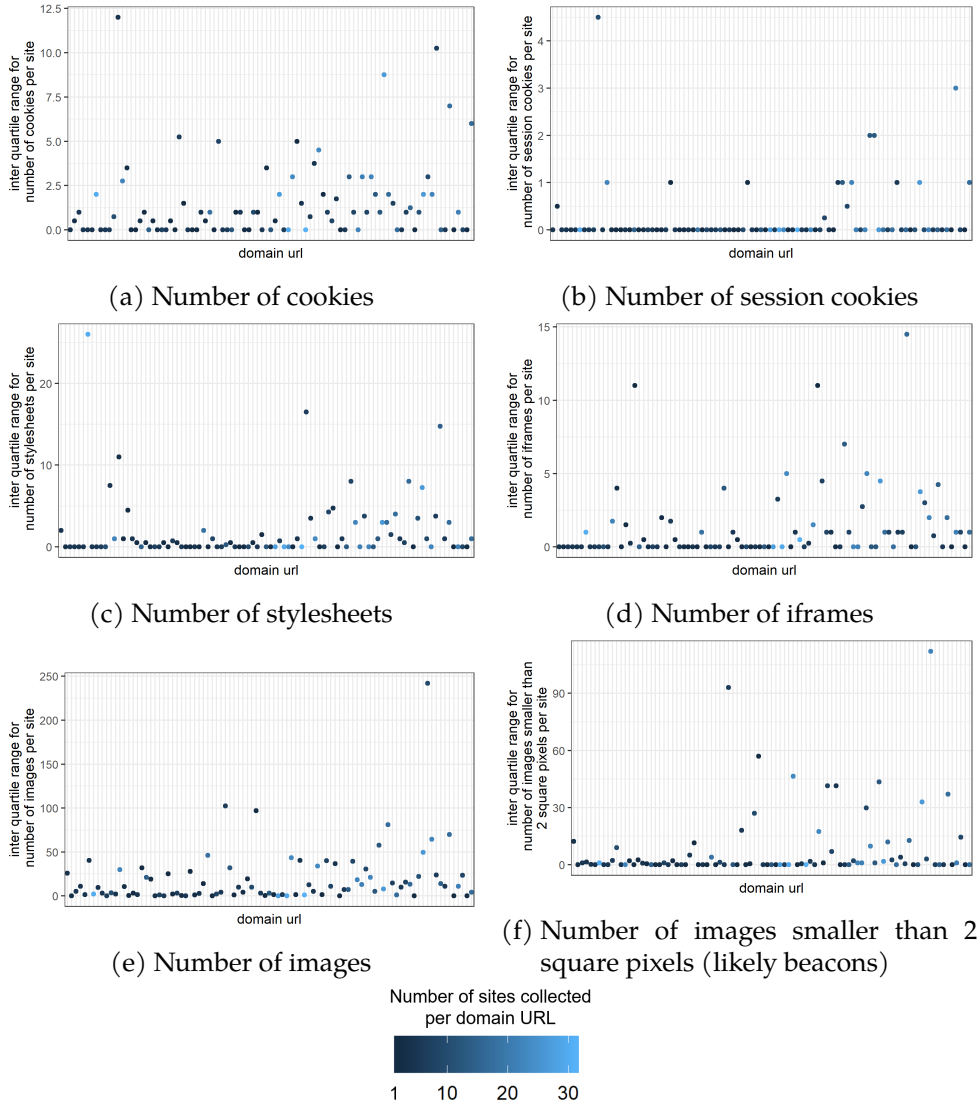
The final dataset consisted of 916 websites, augmented by the information listed above. It included 181 unique domains, with a mean of 5.06 subpages ($SD = 6.93$). The minimum number of subpages per domain was 1 and the maximum was 31. During data collection, the evaluated sites were labeled with their domain, by first using the urllib.parse-function, and then accessing the netloc component of the parsed result. Taking a closer look at the data, some domains differed merely in their prefix, some of them having a www-prefix, while others did not. Removing the www-prefix for all websites resulted in a dataset of 178 unique domains, with a mean of 5.15 subpages ($SD = 7.02$). The minimum and maximum of subpages per domain stayed the same after this correction.

For each of the website characteristics collected, the spread for each domain was evaluated graphically. Since the number of available subsites per domain varied

and was frequently quite small, a robust measure of spread, the IQR was used. The visualizations for the six website characteristics related to presence of trackers in the Whotracksme dataset are presented in Figure 3.1. While there is little spread for some of these variables, namely the number of session cookies (see Figure 3.1b) and the number of iframes (see Figure 3.1d), other variables show a greater spread for some website domains. This makes it apparent that sites within the same domain can differ from each other with respect to these characteristics.

To examine co-occurrence of tracking with certain website characteristics, a correlation matrix of the collected website characteristics for the crawled data was used (see Figure 3.2). In the Whotracksme dataset, cookies were one of the features which were most correlated with tracking ($r = .54$). The number of cookies have also been used to examine the level of browsing privacy achieved with different techniques (Mazel et al., 2019; Mayer & Mitchell, 2012) and one threat for browsing privacy is tracking. This further justifies using the number of cookies as a proxy for tracking in this analysis.

When comparing the correlations with tracking in the Whotracksme dataset and the correlations with cookies in the crawled datasets, some correlations are reproduced to a certain extent. In the crawled data, correlations between the number of cookies and the number of requests for images, for scripts, and the number of iframes are among the higher correlations, at least above the level of |.3| for a medium sized effect for all of the aforementioned. The number of links and total number of requests are also above this threshold in the crawled dataset, but these variables were not included in the Whotracksme dataset. Collecting information on requests to the beacon API was not trivial and could not be implemented in the limited timeframe, however, non-visible tiny images can be used as tracking pixels (Fouad et al., 2020) and data on them was collected. However, their correlation with the number of cookies ($r = .29$) was not quite a medium sized effect. It has to be taken into account however, that the crawled dataset is smaller, and might be less representative than the Whotracksme dataset, since it does not contain data from as many unique domains. Thus it is normal for the correlation coefficients to differ somewhat. Many



(a) Number of cookies

(b) Number of session cookies

(c) Number of stylesheets

(d) Number of iframes

(e) Number of images

(f) Number of images smaller than 2 square pixels (likely beacons)

Figure 3.1.: IQR of different variables for different domain urls. The color of the points depicts from how many websites the statistic of spread was generated.

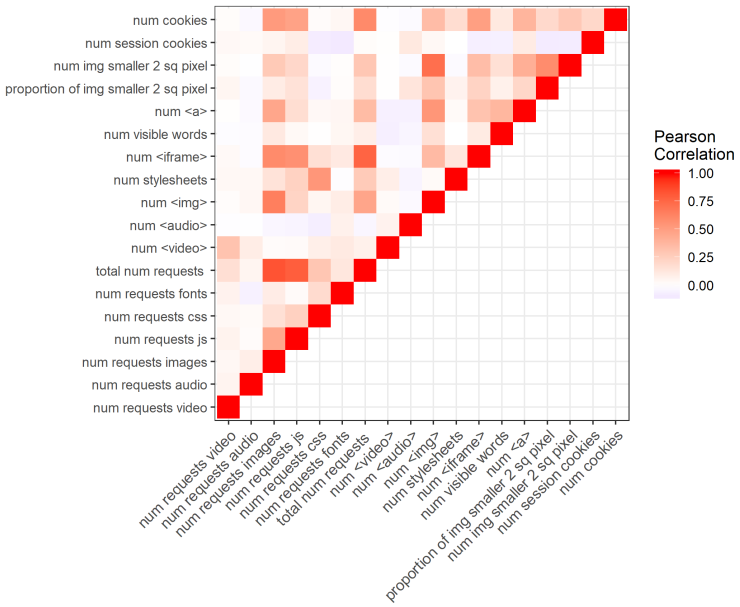


Figure 3.2.: Correlation matrix for website characteristics for crawled data

of the website characteristics, such as the number of audio files or video files or the number of iframes, are hard to assess for end users, so they are not ideal for use in boosts.

A website characteristic which is more accessible to users is the broad category of a website. Websites, or more specifically domains (use of the term as above), were categorized using several available services to obtain objective categorizations. Three different services (Cyren[3], Brightcloud[4], and Symantec[5]) were used to categorize the 50 websites from the starting pool. Interrater agreement was calculated for the three services to see whether their judgments were similar. The assigned categories were unified by hand so that similar categories, which often differed in use of plurals or punctuation characters, were considered the same. In some cases, one or more of the services assigned up to three categories to a single website. To simplify, the most frequent category across the three services was adopted as the main category and used in the calculation of interrater reliability. Fleiss' Kappa for 3 raters was calculated on the unified simplified categories, resulting in $\kappa = .76$. This is a substantial level of agreement (Landis & Koch, 1977).

[3] https://www.cyren.com/security-center/url-category-check
[4] https://brightcloud.com/tools/url-ip-lookup.php
[5] https://sitereview.bluecoat.com/#/

Of the three services, only Symantec did not require the frequent solving of captchas to gain access to the information. To assess whether it would be appropriate to use this service to categorize the remaining websites, interrater agreement was calculated for Symantec and each of the two other services. There was substantial agreement both between the Symantec categorization and the Brightcloud categorization ($\kappa = .78$) and between Symantec and Cyren categorizations ($\kappa = .80$) (Landis & Koch, 1977). The more accessible Symantec service was used for the rest of the websites, until a total of 187 websites were categorized. In cases where Symantec assigned more than one category, categorizations were gained from the two other services and the Symantec-version of the category which was assigned most frequently was adopted as the website category. When there was no single most frequent category, the author decided between the two or three most frequent categorizations by accessing the website, comparing its purpose to available definitions of the categories and selecting the most appropriate one.

Three measurements of tracking were considered: The number of two types of cookies and the number of very small images, which can be used as beacons to track users. For each of these three variables, their means and standard errors for each category were calculated and plotted. The result can be seen in Figure 3.3.

Some categories, e.g. e-mail, have low values across all three categories, while others, e.g. pornography, have high values across all three categories. Others differ for the three categories. For example, websites from the news category exhibit a low number of session cookies, but high numbers of non-session cookies and images which are smaller than two square pixel.. The number of sites used to generate these summative statistics differed widely for each category. Some categories were assigned to only one website (e.g. finance, gambling) and, as a result, error bars could not be calculated. The maximum number of websites for a single category was 145 for the search engines/portals category, while the median of websites per category was 11. All in all, the data for website categories from the crawled sample were not deemed conclusive enough to use in boosts.

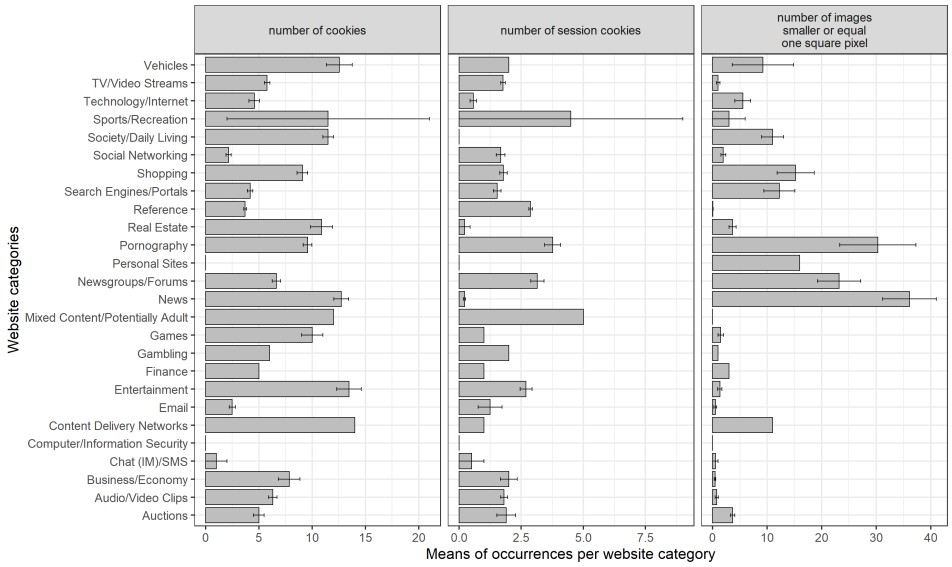


Figure 3.3.: Differences between website categories for three measures of tracking. Bars represent means with standard errors. In cases when a category comprises of only one website, error bars cannot be calculated.

3.2.3. Preliminary Boosts building on Previous Work

Differences in websites from different categories have also been examined in previous work. OpenWPM, an open source web privacy measurement tool was used to investigate online tracking on a larger scale (Englehardt & Narayanan, 2016). An analysis of about one million sites showed that websites belonging to the news category and the arts category have the highest number of involved third parties on average, while websites from the science category, reference category, and pornography category have the least number of third parties on average (see Figure 4 in Englehardt & Narayanan, 2016). A different analysis, also conducted on a large scale, with about 1.5 million sites, showed similar results for cookies (Urban et al., 2020). News websites, and websites from the arts and entertainment category had the highest average number of cookies, while websites from the education category had the lowest (Urban et al., 2020). An additional finding was that more cookies are set on subpages than on landing pages of websites (Urban et al., 2020). It must be noted for the sake of completeness, that though the categories used in these two papers seem comparable, they were acquired from different sources, namely Alexa (Englehardt & Narayanan, 2016), and the McAfee SmartFilter Internet Database

(Urban et al., 2020). Informational statements using this information were constructed and can be found in Table 3.2.

Other previous work has investigated different methods to increase privacy during web browsing. Some of these methods were browser specific (Kontaxis & Chew, 2015), while others were more general and applicable to multiple browsers. Among the reported methodology is using private browsing mode to delete cookies after the browser is closed (Tsalis et al., 2017), disabling third party cookies to reduce the number of third parties and cookies on websites (Englehardt & Narayanan, 2016; Krishnamurthy et al., 2011), and using third party tools, such as adblockers, to prevent tracking (Gervais et al., 2017). Ur et al. (2012) confirm that browser history can be used to inform personalized advertisements. Preliminary boosts were constructed using this information from previous work and can be seen in Table 3.2. Since users in the study would be able to use either Chrome or Firefox browsers, browser-specific information (e.g. from Kontaxis & Chew, 2015) was not used. Similarly Firefox provides a setting to automatically delete browsing history on closing the browser, while Chrome only has the possibility to do this manually. Because of that, this boost would probably be less effective for Chrome users than Firefox users, and for the sake of comparability, it was left out of the further study. Some of the papers contained concrete numbers for certain aspects of the reduction of tracking (e.g. Englehardt & Narayanan, 2016; Gervais et al., 2017). These numbers were unified to a reduction percentage and used in the preliminary boosts.

3.2.4. Pretest of Boost Comprehension

Since it is crucial to the effectiveness of boosts that they are understood, a pre-test was conducted on the crowdsourcing platform Prolific. A survey was designed with Qualtrics, including the boosts marked with x in the "used to test pre-study" column of Table 3.2. To ensure adequate payment of participants on Prolific, eight acquaintances of the author and advisers pretested the Qualtrics survey. One participant was excluded, because that person took over 50 minutes to complete the survey, and stated that they were distracted by phone calls multiple times during

boost name	boost	used to test pre-study	used in pre-study	used in final study
news	News and media websites have more cookies and third parties per page than other kinds of websites.	x	x	
enter-tainment	Entertainment websites have more cookies and third parties per page than most other kinds of websites.	x	x	x
education	Education websites have less cookies and third parties per page than most other kinds of websites.	x	x	x
third party reduction	Blocking third party cookies in the browser settings leads to a reduction of the number of third parties per page by about 30%.	x	x	x
cookie reduction	Blocking third party cookies in the browser settings leads to a reduction of the number of cookies in general.	x		
private browsing	By using private browsing mode (Firefox) or incognito mode (Chrome), cookies are deleted automatically after the browser has been closed.	x	x	x
adblocking	By using an adblocker, the number of third parties per page is reduced by 40%, even without changing the blocker settings.	x	x	x

Table 3.2.: Translated English version of boosts used in different phases of study preparation and in the study; the original German version can be found in Annex A.

the test. Additionally, they did not actually fill out the questions in the survey. Of the remaining participants, four were male, and three were female. Their average age was 42.86 years ($SD = 16.98$) and they took on average 9.78 minutes to finish the survey. In accordance with advice by one of the advisers on this thesis, the expected study duration for Prolific participants was calculated to be about 20% less (7.82 minutes), since these participants are more used to taking surveys. This value was rounded up to 8 minutes, to account for the different number of questions in different survey conditions and it was decided to award participants 1.25 € for participation, a little more than the current German minimum wage of 9.35 €/hour (Nienaber, 2018). Especially he control condition took longer, since it contained some questions from all six other conditions, while only including two boost comprehension questions less than the manipulation conditions. Calculating with slightly longer durations was meant to ensure that participants were paid at least the German minimum wage on average.

Suggestions by the testing participants were incorporated into the final survey. This included rephrasing some questions and eliminating one boost due to confusion with a very similar boost. This resulted in seven conditions for the final study. In the next step, two batches of 14 participants each were tested with Prolific. Comments from two participants were incorporated. There was a minor alteration in the wording of tasks, a question (previously optional) was set to required, and an attention check (previously forgotten) was added to the condition displaying the adblocking boost.

The final version of the survey, which was administered to participants on Prolific, first presented participants with information about the study. After securing informed consent and asking the participant to enter their prolific id, participants were randomly allocated to one of seven conditions. In each of the six manipulation conditions, they saw a boost (see Table 3.2) and were asked two questions referring to their comprehension of this information. Following recommendations from Prolific, an attention check question was included after this (Prolific Team, 2018). It required participants to enter the word "UMFRAGE" into a textfield. On the next

page of the survey, they first had to imagine browsing the internet in a normal way and then answer questions, which required them to use the knowledge from the boost.

For the three boosts concerning certain types of websites, these questions required participants to determine which of five URLs had the most or least number of third parties respectively. For one of these questions, only one URL belonged to the category of website mentioned in the boost, and the others were from different website categories. It was the participants' task to find the odd one out and correctly apply their knowledge from the boost to determine whether that website presumably had less third parties (in the case of education websites) or more third parties than other websites (in the case of the news or entertainment websites). For the other of these questions four URLs were from sites of the category mentioned in the boosts and one was not. Accordingly, the participants' task was the other way around.

For the three boosts concerning browsing behavior, questions showed screenshots from Chrome and Firefox browsers. In the case of the private browsing boost, screenshots were of a site being visited in private browsing mode or in normal browsing mode, and participants were asked to decide whether cookies set on this site would be retained after closing the browser or not. In the case of the adblocking boost, screenshots were of the installed browser extensions, with one image including an adblocker and the other not including one. For the third party reduction boost, screenshots were of a part of the browser settings, where third party cookie blocking was either enabled or disabled. For both of the latter two boosts, participants had to estimate how many third parties would be present on a site with the shown extensions or setting active, if the default were 10.

In the control condition, participants did not see a boost, but were nevertheless asked to imagine browsing the internet in a normal way and required to fulfill the attention check. On the next page of the survey, participants in the control condition saw a subset of the same questions the participants in the manipulation conditions saw. However, to prevent their workload from being much higher than in the other conditions, these participants saw only one URL-question (either asking for the least

or most privacy invasive URL) for each of the website category related boosts and only either the Firefox or Chrome specific questions for the other three boosts. These question subsets were selected randomly.

To further check whether participants had been paying attention and to gauge their retention of the presented boosts, all participants were then presented each of the boosts in turn and asked whether they had seen them during the survey. They were also asked to rate the boosts on a seven-point likert scale from not at all helpful to very helpful concerning making their browsing more private. Finally, demographic information was collected and the participants were debriefed.

A total of 127 participants took part in the final pre-study on Prolific. They took on average 6.56 minutes to complete the study and can thus be considered to have been adequately paid. The conditions received about the same number of participants, with most of the conditions being assigned to 18 participants, except the adblocking (17) and entertainment (19) conditions. Of the 127 participants, 70 identified as male, 56 as female, and 1 as diverse. They were on average 30.79 years old ($SD = 10.77$), with the youngest participant at 18 years and the oldest at 67 years of age. All of the participants reported speaking German as a native language, or fluently and almost at the level of a native tongue, as was required to participate in this pre-test. Most of the participants were either working (66) or students (43), but 17 also reported currently not being employed, and one person was retired. The participants had a relatively high level of education, with 67 of them having at least a bachelor's degree. As for the remaining participants, 14 of them had finished vocational training, 33 had the entrance qualification to higher education (German: "Allgemeine oder Fachhochschulreife"), 17 had finished a medium level of school (German: "Realschule"), and two had finished a basic level of school (German: "Hauptschule/Volksschule").

All participants except those in the adblocking condition of the first batch of participants recruited in Prolific, where the relevant question had not yet been included, passed the attention checks in the format suggested by Prolific. The additional attention checks at the end of the study were also largely successful. However,

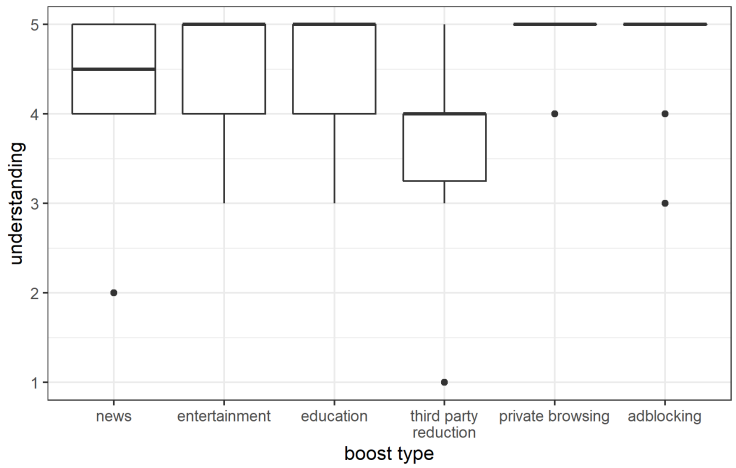


Figure 3.4.: Understanding of boost information

a total of 35 boosts were wrongly claimed to have been seen by participants, while seven boosts were claimed not to have been seen in the study, even though they had been presented to participants. These errors in the attention checks were somewhat evenly distributed across the conditions, with each condition having some errors. Since the errors were fairly common, it could not be justified to reject participants based on such errors, so these participants were retained in the analysis.

The participants in the manipulation conditions were asked to rate their understanding of the boost presented to them, as well as how well they thought they would be able to explain the information in the boost to someone else. Both these questions were intended to measure the level of understanding the participants had of the boosts. Self-reported understanding of boosts was high, see Figure 3.4. The median value of understanding for all of the boosts was at least "somewhat well". The news boost was understood slightly less well than most of the others, and the third party reduction boost was understood the least.

The participants felt it was less easy to explain the boost information to someone else than it was to understand it themselves. Still, the median ease of explaining was above a medium level of ease for four of the six examined boosts. This can be seen in Figure 3.5. Again, the third party reduction boost was deemed harder to explain than the others and in this case, the median value for the entertainment boost was also lower than the other four.

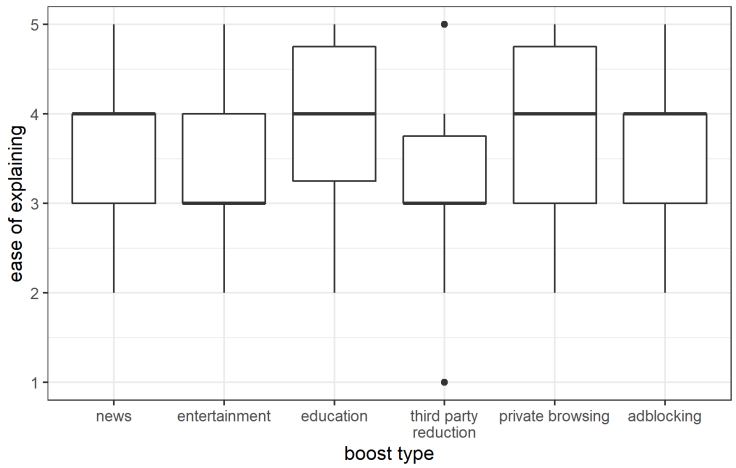


Figure 3.5.: Ease of explaining boost information

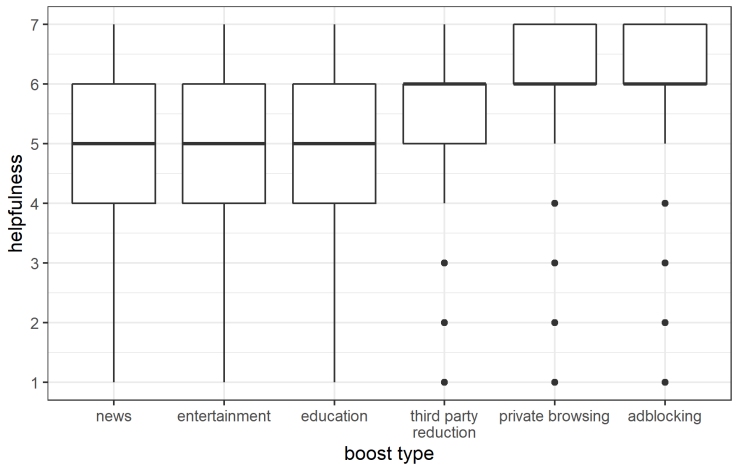


Figure 3.6.: Helpfulness of boost information to increase privacy during browsing

All of the boosts were presented to all of the participants at the end of the study and asked how helpful they found them. On average, all of the boosts were rated to be at least somewhat helpful in making behavior more private. However, helpfulness ratings were a little lower for the three boosts concerning different types of websites than for the three boosts about aspects of the browser. Among the latter, the third party reduction boost was rated slightly less helpful than the private browsing and adblocking boosts, as is evident from Figure 3.6.

To examine whether the boosts were successful in conveying knowledge, the answers of participants in a boost condition were compared to answers to the same type of question of participants in the control condition. Participants in the control

condition answered some of the same questions as participants in a boost condition, but the number of questions about each boost was reduced in the control condition, to lower the workload for these participants.

For the three website category boosts, the percentage of correct answers was calculated for both the manipulation and the control condition for each boost. Because of the limited number of questions, possible percentages were 0%, 50% and 100% for the manipulation conditions, and either 0% or 100% for the control condition. Because of this limitation, the data cannot strictly be considered metric, so a Wilcoxon rank sum test for ordinal data was conducted. For the news boost, the performance difference between the control condition ($M = 13\%$, $SD = 34\%$) and the boost condition ($M = 47\%$, $SD = 47\%$) was significant, $W = 85, p = .017$. The performance difference between the control condition ($M = 0\%$, $SD = 0\%$) and the boost condition ($M = 97\%$, $SD = 11\%$) was also significant and even clearer for the entertainment boost, $W = 0, p > .001$. Likewise, the participants' performance concerning the education boost differed significantly ($W = 11, p < .001$) between the control condition ($M = 6\%$, $SD = 24\%$) and the manipulation condition ($M = 89\%$, $SD = 21\%$).

The percentage of correct answers was also calculated for the private browsing boost, but for this question 0%, 25%, 50% and 100% were possible in the manipulation condition, and 0%, 50% and 100% were possible for the control condition. Again, a Wilcoxon rank sum test for ordinal data was conducted and the difference between the control condition ($M = 6\%$, $SD = 24\%$) and the private browsing condition ($M = 85\%$, $SD = 21\%$) was significant, $W = 99$, $p = .03$.

The analysis for the third party reduction and adblocking boosts was slightly more complicated since the comprehension questions asked participants how many third parties are present on average on a certain site, with different settings in the browser. Assuming the presence of ten third parties with default settings, the correct answer would have been seven in case third party cookies are blocked, and six when an adblocker is used. To measure the size of the deviation from this correct value, the absolute value of deviation from the correct value was calculated. This

means, that if participants answered either four or eight for the case with an active adblocker, this would be an absolute deviation of 2. These values were averaged over all the relevant questions answered by the specific participant.

Since neither the average absolute deviations for the adblocking condition ($W = 0.79, p < .001$), nor the corresponding control condition ($W = 0.87, p < .001$) were normally distributed, when examined using a Shapiro-Wilk test, another Wilcoxon rank sum test was utilized. It detected a significant difference in performance between the control condition ($M = 2.72$, $SD = 1.04$) and the adblocking condition ($M = 1.5$, $SD = 1.75$),$W = 348, p = .002$.

The average absolute deviations for the third party reduction condition were normally distributed according to a Shapiro-Wilk test ($W = 0.95, p = .12$), but they were not normally distributed for the corresponding questions in the control condition ($W = 0.85, p < .001$). For this reason, a Wilcoxon rank sum test was calculated. It did not detect a significant difference between the control condition ($M = 3.86$, $SD = 2.07$) and the third party reduction condition ($M = 3.67$, $SD = 3.86$), $W = 638, p = .91$.

Hence all the boosts, except the third party reduction boost, proved to be effective since participants in the manipulation conditions provided more correct answers or answers that were closer to the correct answer than participants in the control condition. To examine some possible reasons for the failure of the third party reduction boost the distributions of the answers were plotted. In general the mean absolute deviation was slightly less in the control condition of the third party reduction boost, although not significantly so, than in the manipulation condition. These plots were faceted by the question type (showing either a setting where third party cookies are activated or not activated) and the browser depicted in the images in the question. Figure 3.7 shows that for Chrome browser, when third party cookies are activated, almost all participants correctly selected 10, while for Firefox, participants' selections show a greater spread. This might be due to Firefox not simply providing the option of activating or deactivating third party cookies in custom settings. Instead, more nuanced settings were available. To make the two images for Firefox

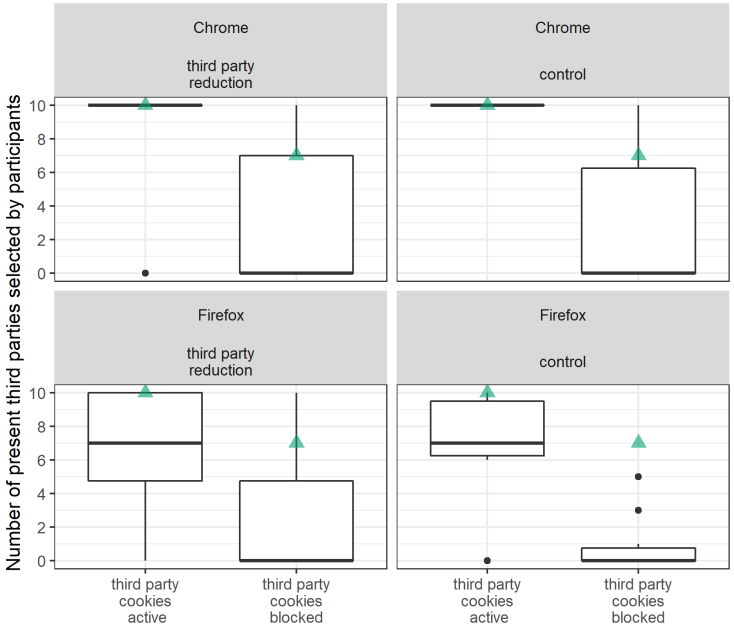


Figure 3.7.: Distribution of answers to comprehension test questions concerning the third party reduction boost. Triangles mark the correct answer for each question.

as similar as possible, both aimed to use the custom settings. However, there was no option to simply allow all cookies in this setting. The setting which seemed the least privacy-protective was chosen for the image which was meant to represent third party cookies being activated. This meant only cookies from websites which a user had not visited would be blocked. Of course, this is different from the available settings on Chrome to simply not block third party cookies, and to not explicitly state to block any other kinds of cookies. This may explain the larger spread of answers for Firefox in this type of question. Participants may have expected a certain percentage of third parties to be removed by this kind of blocking.

For Firefox, there is also a larger difference between the manipulation condition and the control condition for the setting where third party cookies are blocked. In the control condition, more participants seemed to expect that blocking third party cookies meant reducing the number of present third parties to zero, when it only provides a reduction of 30%. For Chrome, this difference is not as pronounced, although for both conditions, the number of third parties that participants expect

to be present are slightly closer to the correct number according to the boost, than for Firefox.

Based on the findings of the boost pre-study, the effective entertainment, education, private browsing and adblocking boosts were implemented in the main study. It was decided to retain the third party reduction boost, even though it was not shown to be effective in the pre-test. Nevertheless, participants in the pre-study felt this boost was helpful, and it is possible that presentation issues with the Firefox version of this boost's comprehension question could be responsible for participants in the manipulation condition performing poorer on the comprehension questions, than the participants in the control condition. Additionally, it is possible for participants in the main study to directly use the information in this boost. The implemented browser extension can also track their cookie blocking settings, enabling the detection of a change in behavior which can be connected to the third party reduction boost.

On the contrary, even though the news boost was effective, it was decided not to employ it in the main study for ethical reasons. For one, it was not understood as well as the other two similar boosts referring to website categories. More importantly however, acting on this boost could mean that participants try to avoid visiting news websites. Another possible reaction could be for participants to try to limit their visits to only one or very few different news websites. This reliance on only one source of news could lead to an aggravated filter bubble situation for these participants. Filter bubbles are already in place for many users, since they tend to look at articles which mirror their own opinion (Flaxman et al., 2016). It does not seem ethical to encourage participants to limit their access to different kinds of news, so it was decided not to use this boost in the main study.

3.3. Choice of Nudges

In literature about nudges, social nudges are investigated in domains as diverse as energy conservation (Allcott, 2011), environmentally friendly traveling (Riggs, 2017), tax compliance (Hallsworth et al., 2017), and reducing food waste (de Visser-

Amundson & Kleijnen, 2020). They build on the social nature of human beings, who are influenced by the behavior of their peers around them (Thaler & Sunstein, 2008). Social nudges towards privacy have also been successful (Coventry et al., 2016), getting people to decline cookies when they thought a majority of their peers were also doing so. Social nudges have also been proposed in the form of comparison-based privacy (Ziegeldorf et al., 2016), a format for nudges which utilizes a user's peer group for nudges concerning their privacy. Examples of nudging towards privacy throughout the literature have often worked with visualizations, using color (Zimmerman et al., 2019a) or number (Egelman et al., 2009) of privacy indicators to visualize the level of privacy.

In the current study, a more general social nudge comparing a participant's behavior to all of the participants in the study was used. A textual representation of a participant's rank within the group was combined with the visualization of privacy on a scale from red, for comparably privacy invasive, to green, for comparably not as privacy invasive. On this scale, the average value including all the participants, as well as the current participant's value, were marked. However, a large proportion of the population, up to 8% of European males and 0.5% of females, are affected by red-green deficiency, a common visual deficiency, whereby the colors of red and green cannot or only with difficulty be distinguished (Albrecht, 2010). To account for this condition in participants, a smiling and frowning emoji were additionally presented together with the color coded scale. An example of a nudge can be seen in Figure 3.9. While previous work suggests that social nudges are not very popular with users (Schöbel et al., 2020), at least in some cases, there is evidence that the effects of social nudges persist even after the nudge itself is not present anymore (Brandon et al., 2017). Normally, this kind of continued effect would be associated with boosts. As such, social nudges are deemed appropriate for use in this study, despite their possible unpopularity.

The social nudges in this study utilized five different measures of privacy, against which the participants' score was compared. Number of cookies and third party requests are directly derived from the proxies for browsing privacy. Visiting more

behavior	privacy points	browser availability
encountering one third party request	-1	both
encountering one new cookie	-1	both
visiting a website in private/incognito mode	+2	both
visiting a website with Do not track active	+1	both
visiting a website while blocking third party cookies	+3	both
visiting a website while rejecting trackers	+2	Firefox
visiting a website while rejecting all cookies	+5	Firefox
visiting a website while all cookies are nonpersistent	+1	Firefox

Table 3.3.: Behaviors and the privacy points which are awarded/deducted for them

websites means encountering more cookies and third party requests, so the numbers presented to the participants were averaged over the number websites visited to be able to compare participants with different amounts of internet usage. A low number in these measures represents less privacy invasive behavior. Privacy points provide a more abstract representation of browsing privacy. They were awarded for privacy supportive behavior, such as blocking third party cookies or using the private browsing mode and deducted, when encountering cookies or third party requests. Accordingly, a high number of privacy points means less privacy invasive behavior. The number of points awarded or deducted was defined more or less arbitrarily. Behavior, for which evidence supported its privacy protectiveness, received more privacy points than other behavior for which evidence was not as conclusive. The number of points awarded for specific behavior can be seen in Table 3.3. As such there were three measures including privacy points, one referring to privacy points concerning website visits, one referring to privacy points concerning browser settings, and one combining the two of them.

3.4. Implementation of Study Apparatus

To evaluate the effects of nudges and boosts on browsing privacy in situ, a web extension for Firefox and Chrome browsers was implemented. They are currently the most used browsers on desktop devices in Germany, with about 47.3% of browser market share attributed to Chrome and 22.5% of marketshare allocated to Firefox (Statcounter - GlobalStats, 2020). The next most popular browser for desktop devices is Safari at 10.9% marketshare. The former two browsers enable extensions to be implemented using a unified API (MDN contributors, 2020a), which is not possible with Safari. Thus, other browsers were not supported in this study, because they do not conform with this system of developing browser extension. Further less popular browsers, such as Edge, or Opera, were not supported, because while the extension API provides a common platform for implementation, there are a host of incompatibilities between Chrome and Firefox (MDN contributors, 2020b), and the compatibility differs between browsers in general (MDN contributors, 2019). So, adapting an extension to all the browsers requires amounts of work which are outside the scope of this thesis. Different aspects of the implementation of this extension and the other parts of the study apparatus are described in the following.

3.4.1. Web Extension

Participants during the study installed and used a web extension for three weeks. This extension was based on previous work by Ortloff et al. (2020), which is available on Github[6]. It retains the main structure of that extension, some code concerning making the extension compatible for multiple browsers, and the CSS files, which were only slightly adapted and augmented for this study. Code from a web extension version of the discontinued lightbeam addon was used to capture third party requests and first parties. Specifically, the *SendThirdParty* and *SendFirstParty* methods from the capture.js file were used and adapted in the extension. The original code is available under a Mozilla Public License 2.0 in a Github Repository being

[6]Github repository of code used by Ortloff et al. (2020): `https://github.com/Maxikilliane/CPP-browser-extension`

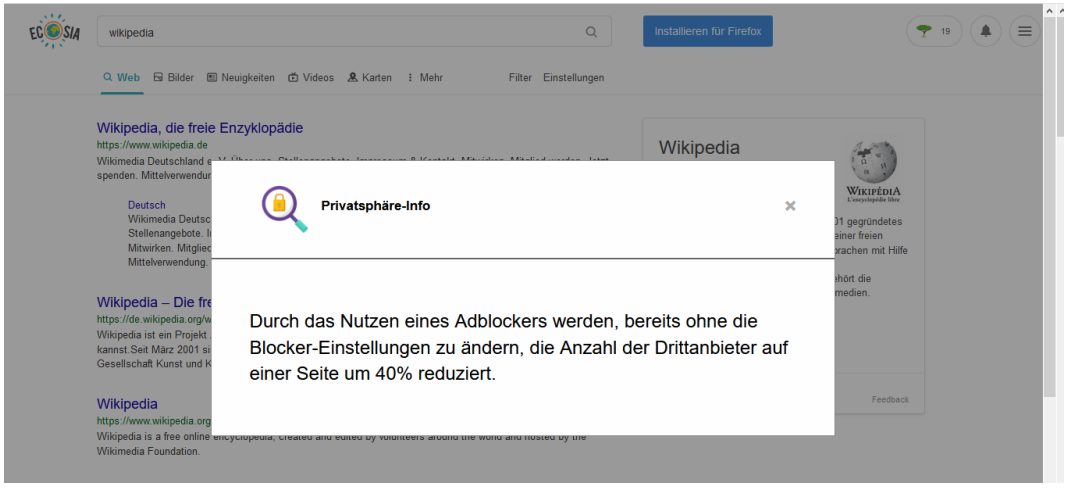


Figure 3.8.: Screenshot of a boost

maintained by Princiya Marina Sequeira[7] (Sequeira, 2019). It was adapted to make it compatible with the Chrome browser, and to fit with the purpose of the study extension. This meant that information was saved for the first and third parties that was different from the Lightbeam extension.

The interface of the extension consisted of two components. Boosts or nudges were shown in the form of a modal dialog, which was inserted into the current page by a content script. This happened on the first website a participant visited every day during the intervention period of the study. A boost consisted of an information shown to the participants, as described in Section 3.2. A screenshot of this kind of intervention can be seen in Figure 3.8. Nudges showed the participant their performance in a certain metric compared to other participants, both visually, and in textual form, as is depicted in Figure 3.9. It presented this information both for the whole duration of the study up to the previous day, as well as only for the previous day. Participants in the control condition did not see any kind of modal dialog.

Additionally, participants received information about their participation in the study in a pop-up, which was accessible by clicking on the extension's icon in the browser toolbar. This was the extensions main user interface and it always showed the participant's randomly generated label, as well as contact information of the au-

[7] Github repository of web extension version of Mozilla Lightbeam addon: `https://github.com/princiya/lightbeam-we`

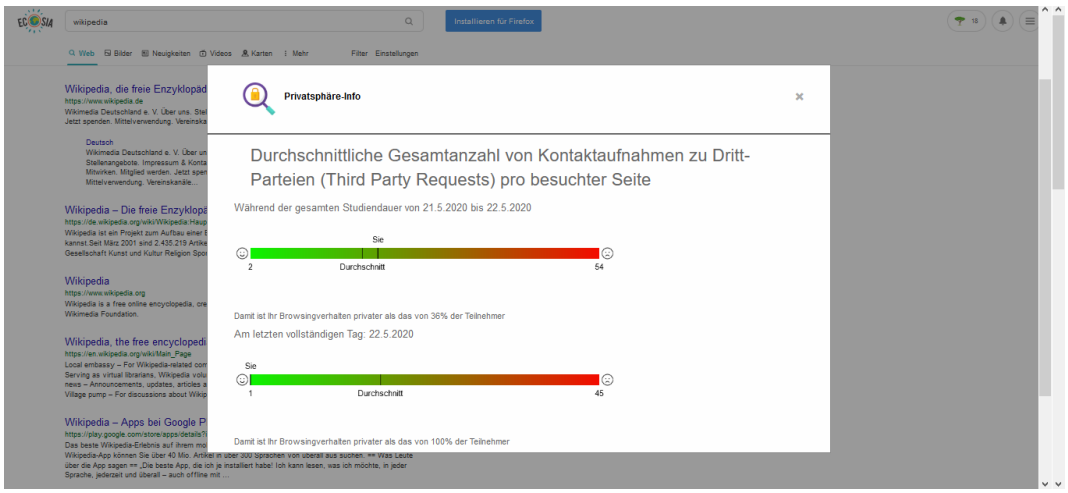


Figure 3.9.: Screenshot of a nudge. Zoom levels are lower than the default settings in this screenshot, to enable the capture of the whole nudge. Since this screenshot was taken during a test of the study extension, the study duration dates are not correct.

thor and the study duration. Depending on the phase of the study, it either showed no additional information, the previously shown nudges or boosts together with brief explanations of some terms used in them or, at the end of the study, all possible boosts and nudges. At the end of the study, a participation code was also made available to the participants through this interface.

3.4.2. Backend and Database

A representational state transfer (REST) API was implemented with Python, version 3.7, and its framework Django, version 3.0.6, using the Django REST framework, version 3.11.0 . This API was used to enable communication of the web extension to the database, which is described below. It was available at `http://132.199.143.90:8090/api/` and provided the following API endpoints which were used during the study. They are listed starting with the "api"-part of the URL shown above. The domain is omitted in the list below for sake of brevity.

api/participants/ was used to save the current participant's generated data (label) to the database

api/website_categorizations/?domains=<domains> was used to retrieve categorizations for the domains passed as a comma separated list to the domains query parameter

api/website_visits/ was used to save information concerning website visits of a participant; this includes saving the number of privacy points awarded or deducted for that website visit on the backend

api/popup_sessions/ was used to save when participants access the extension's user interface

api/modal_sessions/ was used to save when participants view modal dialogs

api/random_participant was used to get a random participant from the database at the beginning of the study

api/study_nudges/?participant_id=<id> was used to get nudge information concerning a specific participant if that participant's id was passed as a parameter

api/participation_code/ was used to get a random participation code at the end of the study

Note that additional API end points were implemented, but only those that were used during the study are described here.

Data generated during the experimental procedure were saved in a PostgreSQL (version 9.5.21) database. The database included some tables generated by Django for administration purposes, and others, which were used to save experimental data concerning the browser extension. The latter tables are visualized in an ER diagram in Figure 3.10, which was generated using the DBeaver software in the community edition, version 7.1.0 (*About DBeaver*, 2020).

Data concerning the participants' study participation were saved in the *participant* and *study_condition* tables. The *participation_code* table was queried at the end of the study to return a random non-assigned participation code and contained the same number of participation codes, as there were participants in the study. The participants' browsing behavior was saved in the *privacy_point_history_entry* and *website_visit* tables, and their interactions with parts of the extension were saved in

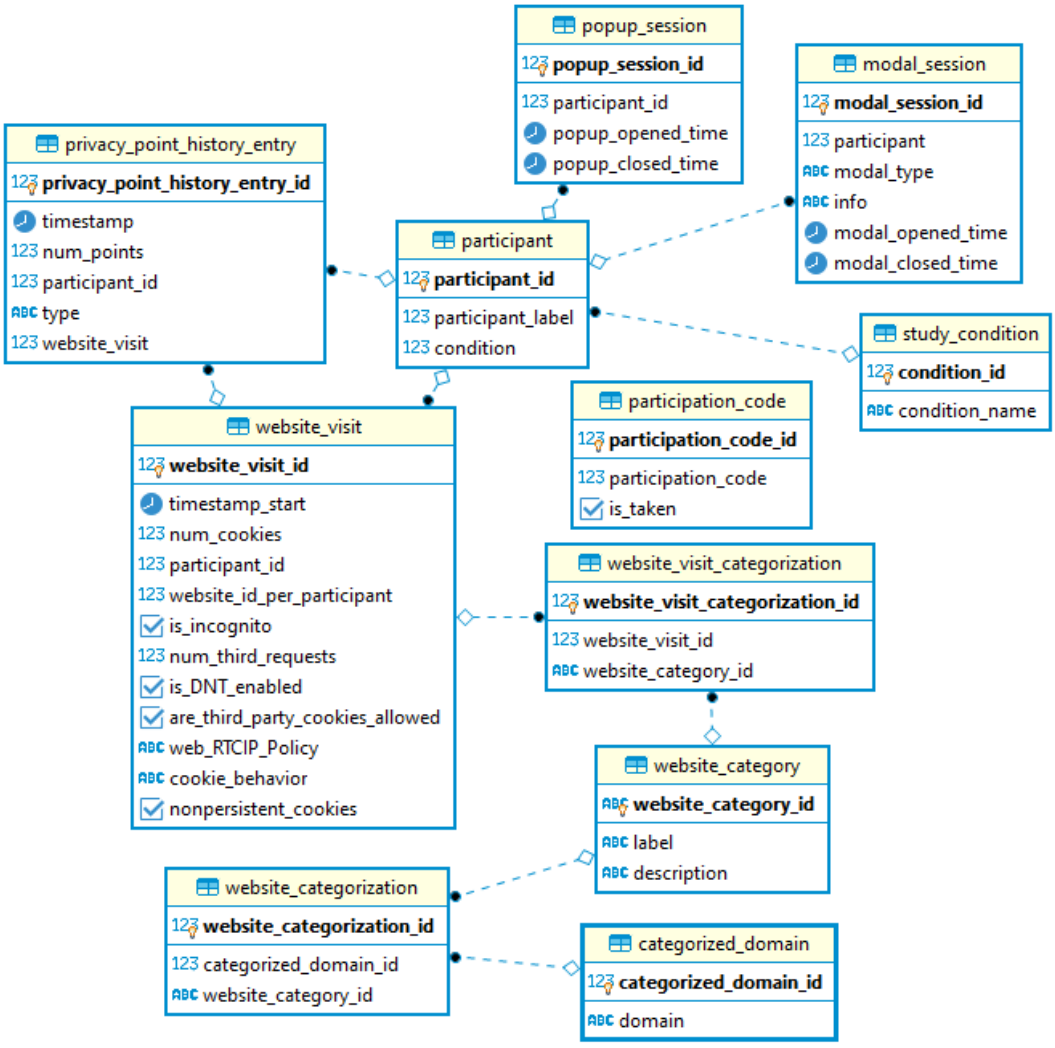


Figure 3.10.: ER diagram of the database, as generated by DBeaver. Tables automatically generated by Django are excluded here.

the *popup_session* and *modal_session* tables. Website categorization was handled in a multiple step process to ensure that participants' website visits could not be connected to a domain. The domains which had been categorized using Webshrinker, and their corresponding categories were saved in the *categorized_domain* table, connected to the categories saved in *website_category* through the *website_categorization* table. This was necessary so previously categorized domains did not require a new request to the Webshrinker API, which only had 30,000 requests available for the duration of the study. On the other hand, website visits referenced the *website_category* through the *website_visit_categorization* table.

3.4.3. Categorizing Websites

An external API was used to obtain website categorizations. To choose an API, two possibilities were evaluated. First, the whoisxml API was tested, since their plan with 50,000 requests for $25 was affordable and could be restocked without monthly limits. As 100 requests were free for testing, a random sample of 100 domain-urls was chosen from the 187 domains previously categorized in the prestudy. The returned categories were then compared to the categorizations of the same domains by Symantec, from the prestudy. To calculate agreement, the categories were unified in a process similar to the one in the prestudy, by using the category descriptions to match categorizations and then specifying one of the versions to be converted to the other. Whoisxml assigned multiple, up to three, categories to many of the domains. To calculate a sort of maximum agreement, if one of these categories matched the category assigned by Symantec, it was chosen. In cases where none of the categories matched, the first one was chosen. There was only slight agreement, $\kappa = 0.12$ for 96 websites, between the two categorizations. The four missing websites are the result of failed categorizations, so there were four domains which could not be categorized. When manually inspecting the Whoisxml categorizations, many seemed random and not even close to the actual content of the website, even for popular and English language sites, like "google.com". This site was classified

as belonging to the categories *Internet and Telekom*, which is appropriate, but the other categorizations *Health* and *Pets and Animals* do not seem fitting at all.

Consequently, a second API, Webshrinker, where requests can only be purchased for a single month, and which is also slightly more costly at $20 for 30,000 requests, was tested. Similarly, this API also offers 100 free try-out requests, which were used with the same sample of domains. There are two taxonomies available to be used for categorization. The first, IAB Categories, which encompasses over 400 categories, was too detailed and extensive for the purpose of this study, while the second option, Webshrinker categories, seemed more appropriate with 40 categories (Webshrinker, n.d.). Thus the second option was used in this test. The unification process was similar to the test of the Whoisxml-API and multiple categories were treated the same as well. The categorizations of Symantec and Webshrinker were more similar however, with a moderate agreement, $\kappa = .51$ for 98 domains. The two missing domains out of 100 requests stem from having two domains twice in the sample, once with "www"-prefix, and once without. The API returned all URLs without that prefix, resulting in duplicates, which were excluded from the analysis. A total of seven websites could not be categorized by the API. When leaving those out of the data, agreement is slightly higher, $\kappa = .55$ for the remaining 91 websites.

A manual inspection of the categorization also yielded more sensible categories than for Whoisxml. To achieve this, a sample of 50 domains was drawn from the websites categorized with Webshrinker, which were not left uncategorized by the API. These domains were manually categorized using the Webshrinker categories by accessing the website in question and then assigning a category. Only one category was assigned to make the evaluation easier. Maximum agreement was calculated in the same way as described above. The agreement between the manual categorization and the automatic categorization by Webshrinker was substantial for the 50 websites, $\kappa = .80$. As a consequence Webshrinker was chosen to be used in the study. Descriptions of the Webshrinker categories can be found in Annex B.

3.4.4. Tracking of Participant Behavior

Several kinds of participant behavior were tracked during the course of the study

Interaction with modal dialogs For each time a modal dialog was displayed, time stamps were saved for the time it was opened and closed respectively, as well as the nudge or boost which was displayed and the participant-id for whom this occurred.

Interaction with exension user interface For each time a participant clicked on the study icon in the browser bar, time stamps for the start and end of the interaction were saved.

Website visits For each website a participant visited, a time stamp marking the start of the website visit, the number of third party requests from that website, the change in number of cookies on that website, the categories assigned to that website, a website id, which was unique per participant, as well as four different privacy related settings were saved. These privacy related settings were whether Do not track was enabled or not, whether a website was visited in private browsing mode or not, the WebRTC IP handling policy, and the cookie blocking policy. For Firefox, this last setting provided nuanced options, and it was additionally tracked whether cookies were set to be persistent or non-persistent. For Chrome, this setting was approximated by tracking whether third party cookie blocking was enabled or not.

Privacy points These were saved according to the number of third party requests, cookies and the privacy settings concerning a website visit. The logic behind their assignment is described in Table 3.3.

Care was taken to ensure that it was not possible to detect which participant had visited which website. To this end, website visits were categorized, by sending a request to the Webshrinker API through the backend. Then, only the categories of the website were passed on to be saved along with the website visit. In case of sensitive categories, such as pornography, the categories were anonymized to uncategorized, which was also assigned when a categorization was not possible. For each partici-

pant, a list of websites which they had visited was kept on their machine, in the local storage assigned to the extension. Websites there were matched to their categories, to reduce the amount of categorization requests necessary, and were assigned an id on the participant side. That way it was possible to detect whether a participant had visited the same website multiple times, since the website id was the same, while still maintaining ignorance of the actual websites visited.

3.4.5. Pretest of the Experiment System and Questionnaires

The experiment system was tested with 4 different participants from 19th May 2020 to 22nd May 2020. During this test, each phase's duration was one day. A multitude of bugs was uncovered during this test, many of which concerned the Chrome browser. These were fixed and a new version of the extension was deployed to the test participants immediately. This was repeated several times in an iterative process.

To estimate the time necessary to fill out the questionnaires which were to be part of the main study, two subjects, who had also previously tested the Qualtrics survey, tried out the questionnaires. Since one of the participants was much faster than the other, and this participant had also been much faster during the test run of the Qualtrics survey, the other participant's times were used. This served to let participants in the main study know how much time to allocate for filling out questionnaires at the start and end of the study, and was also used to estimate the number of course credits to award to participants studying Media Informatics or Information Science at the University of Regensburg.

The installation instructions for Firefox and Chrome were also tested with a single participant. Based on their feedback, some changes of phrasing were implemented and an illustration was shuffled further back in favor of explanatory text before it.

3.4.6. Problems during the Study

Despite testing with multiple participants, some problems with the extension occurred during the study. It is not possible to display modal dialogs like the one

used in the intervention on all websites, since this calls for the use of a content script. A very general match pattern was used in the extension to match sites on which to inject this content script (MDN contributors, 2020c). All HTTP, HTTPS and WebSocket URLs were supported, but some URLs, e.g. with the file scheme, or, for example, the chrome extensions page were not supported. Trying to present the modal on such a site resulted in an error in these cases, which, while not having a detrimental effect on the functionality of the extension in general, did result in errors being logged on Chrome, since the extension was installed in developer mode. Some participants using Chrome were concerned about these errors, which they saw when installing the extension. To alleviate these concerns, an explanatory e-mail was sent out to Chrome users.

The assignment of participant ids and conditions also did not work quite as expected. During the installing process of the extension, sometimes an id was assigned, and then overwritten by a second id. This problem did not appear during the pretest of the experiment system, and was only discovered, because there were more participant ids assigned, than surveys filled out at the beginning of the study. This lead to the numbers of participants for each condition not being exactly equal.

During the study, several participants reported problems by e-mail. In some cases, the extension uninstalled itself upon closing the browser, so that new participant ids were assigned on reinstalling the extension. In one case, this occurred so often within a single day and browser session, that the participant terminated their participation. Unfortunately, it was not possible to find out why this behavior occurred. Some other participants reported that their browser's performance was worse during the study. As a response and to learn how often such problems occur, the final survey was adjusted to include questions about encountered problems and to enable participants to indicate multiple participant labels if necessary.

When analyzing these self-reported problems, in general the percentage of problems among Chrome users (15 users, 41.7%) was higher than for Firefox users (4 users, 16.7%). Similarly, MacOs users (6 users, 50%) were also more likely to encounter problems than Windows users (13 users, 27.7%). The one participant us-

operating system	browser	usage of privacy extensions	total users in this condition	% any problem	% reinstall necessary	% slower	% louder	% browser crash	% multiple ids
Mac OS	Chrome	No	8	62.5	12.5	50	37.5	0	37.5
Mac OS	Firefox	Yes	2	0	0	0	0	0	0
Mac OS	Firefox	No	2	50	0	50	0	0	0
Windows	Chrome	Yes	13	38.5	0	38.5	7.69	15.4	15.4
Windows	Chrome	No	15	33.3	0	13.3	6.67	0	13.3
Windows	Firefox	Yes	7	14.3	14.3	0	0	0	14.3
Windows	Firefox	No	12	16.7	0	0	0	0	8.33

Table 3.4.: Summary of problems during study, cross tabulated by operating system, browser, and the use of privacy extensions. Since participants using Linux did not encounter problems, and since there were no Chrome users on MacOs who used privacy extensions, these rows are omitted from the table. Problems which only occurred for a single participant did not receive an extra column.

ing Linux did not report any problems. It was also examined whether having privacy protective extensions installed lead to problems with the study extension. This might be the case since the study extension requested quite invasive permissions, even though these were explained thoroughly to participants. However, contrary to these expectations, of the 23 users of privacy extensions in the sample, only 26.1% (6) encountered problems during the study while 35.1% (13) of those without further installed privacy extensions had problems. At the same time, two participants reported their browser crashing during the study period and both of these crashes occurred under Windows, on Chrome, and with Adblock Plus installed. A summary of problems by browser, operating system and the use of privacy extensions can be found in Table 3.4.

It should be noted that there is a high likelihood of unreported problems. For one, several participants did not complete the final survey and did not respond to multiple e-mail reminders. These participants may also have encountered problems, which caused them to stop participation. Also, some participants may have considered such problems as louder ventilation too minor to report. To decrease the work-load of the final survey, conditional questions were included, which, when an-

swered negatively, let participants skip sections of the survey. This was also the case for a question asking whether participants experienced problems with the study extension. Finally, some participants did not report problems, even though they declared multiple participant ids in the final study. This was also rated as an instance of a problem.

Finally, there was a time measurement bug for the participant with the id 104 (P 104). Their local date was set to be later than the actual date, by more than a week. Since the extension used client side dates to save website visits and other measures and also to determine whether to show interventions, such an error changes the correct sequence of study phases. Based on their local date, the participant would have been shown the intervention already during the first week of the study. However, this participant was allocated to the control condition, so this was not a problem, and their data were retained. They were adjusted based on the dates on which the participant filled out the survey at the end of the study to change the dates to the real dates during the study. Such problems could be resolved in future versions of the extension by regularly querying the backend for the current date and using this, at least to manage the phases of the study.

3.5. Participant Requirements and Recruiting

Participants in this study were required to be over 18 years old, to ensure the validity of their informed consent. Additionally, since the study was conducted in German, participants had to have a sufficiently good command of the German language, and should be able to speak and understand it fluently. Finally, they had to have access to a desktop or laptop computer on which they were allowed and able to install a browser plugin for three weeks. They should also be willing to use a single browser, either Firefox or Chrome, for the duration of the study.

Since work on this thesis was ongoing during the Corona situation and regulations including social distancing were in place (deutschland.de, 2020) participation and recruitment were set up to be completely remote. Recruitment started at the end of March 2020 and was on-going throughout April until just before the study started

in mid-May. A summary of the study was posted in forums at the university, which serve the purpose of recruiting participants for student experiments. The same summary was also sent to friends and acquaintances of the author through various messaging services, such as Telegram, Whatsapp, and in some cases, e-mail. The first adviser of this thesis additionally published the call for participants on social networks, such as Facebook and Twitter, since the author did not have an account on these. People willing to participate were asked to enter their name, e-mail address and which browser they would use during the study in a form. They were also encouraged to forward the survey to others who might be interested, promoting a snowball sampling approach.

3.6. Procedure

Two days before the study started, on Friday, May 22nd, an e-mail was sent to all the addresses collected in the recruitment form. It contained a brief explanation of the study procedure, as well as the information that the study would start in two days, on Monday, May 25th. A document explaining the study in more details, including risks and benefits for the participants was also sent along with this e-mail, so that participants had time to read it without pressure.

On the day the study started, participants received further instructions per e-mail. Informed consent was obtained by emphasizing once again, that the document which had been sent along with the earlier mail had to be read, understood and accepted before continuing with the study. It was stressed that installing the study extension meant accepting the terms of the study as presented in the consent document. Depending on their stated browser preference, participants either received details and illustrated instructions to install the extension on Chrome, or on Firefox. They were encouraged to report any problems and questions they encountered to the author. Indeed, some participants using Chrome reported their concern about errors displayed in the extension tab after installing. These errors did not hamper the study procedure however, so the author was able to reassure the worried participants. When the study extension was correctly installed, it requested

a random participant id through the aforementioned API, which also allocated the participant to one of three study conditions.

After installing the study extension, participants were asked to fill out a number of questionnaires, measuring their affinity for technology interaction (ATI) (Franke et al., 2019), their privacy concerns, measured with Internet Users' Information Privacy Concerns questionnaire (IUIPC) (Malhotra et al., 2004, a German translation used by Harborth & Pape (2019) was provided through personal communication), their general knowledge about privacy, as measured with the Online Privacy Literacy Scale (OPLIS) (Trepte et al., 2015; Masur et al., 2017), and their self-reported privacy behavior. The last questionnaire was adapted from Zimmerman et al. (2019a), where items to measure self-reported privacy behavior were provided in English. These were translated to German by the author. During the questionnaire tests described in Subsection 3.4.5, minor changes to wording were carried out based on feedback from the testers. A second kind of privacy knowledge, which was more specifically geared towards this study, was also measured. It consisted of the items used in the control condition of the boost pre-study. However, due to the problems with the Firefox version of the questions concerning the third party reduction boost, the image for this question was changed. This meant that the images displaying blocked and active third party cookies differed more than the corresponding images for Chrome, but also that the image depicting active third party cookies was hopefully understood better as such. After all questionnaires were filled out, the participants were prompted to use their browser as they normally do during the course of the next three weeks.

The main study consisted of three phases, each one week long, which are displayed in Figure 3.11. The first week served as a control, to measure browsing behavior for all participants without the influence of an intervention. Every time a participant visited a website, the data concerning website visits described in Subsection 3.4.4 were collected and requests were categorized as third and first party requests and cached in the local storage on the participant's machine. A request was categorized as a first party request, if the hostname property of the target URL

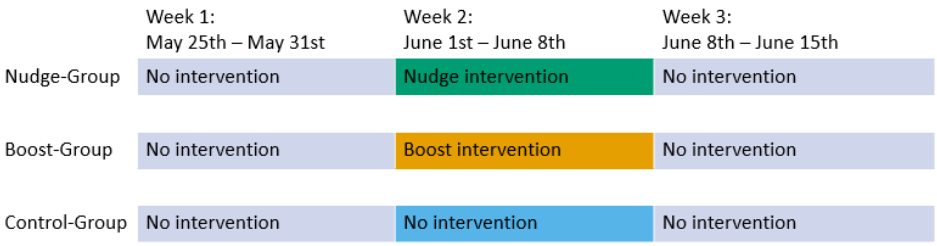


Figure 3.11.: Procedure of the main study

was the same hostname property of the current tab. If this was not the case, it was categorized as a third party request. This procedure was implemented using code snippets from the Lightbeam extension.

When a threshold of 25 sites was reached, or when a new date started, the cached third party requests were allocated to the website visits during which they occurred and both third party requests and cookies were accumulated to obtain a count per website visit. The visited websites were categorized and then sent to the backend with all the information described in Subsection 3.4.4. Regular checks of the questionnaire data showed that not all participants who had signed up for participation took part in the main study. To remind these missing participants about the study, two e-mails were issued to the participants during the first week, one after two days, and the other after four days. This resulted in eight additional answers. Starting a bit late was not deemed a cause for concern, since participants were not exposed to any interventions during the first week of the study. As long as participants started their participation during the first week of the study, this was assumed to be sufficient.

During the second week, the intervention phase, browsing behavior was collected using the same procedure as during the first week. However, participants in the boost and nudge conditions received daily reminders in the form of either boosts or nudges. As described in Section 3.2 and Section 3.3, there were five different nudges, and five different boosts. One of these was displayed to the participant once a day, on the first website they visited after the date changed, in the form of a modal dialog. The presented intervention (either a boost or a nudge) was chosen from the set of previously not yet displayed intervention at random. Once all interventions had

been displayed at least once, the population to be sampled from was refilled, so that each intervention was displayed in a modal dialog at least once and at most twice. During this phase, participants could also access their previously presented nudges or boosts in the extension's user interface, which was accessible by clicking on the extension icon in the browser bar. In the case of the nudges, these were updated daily with current information.

A third group, which did not see any interventions during the study, was utilized in this design to ensure that differences in browsing behavior did not occur simply because of changing circumstances outside of the study. This was especially important during the time when the study took place, since it was possible for restrictions based on the Covid-19 situation to be changed or repealed at any time during the study. These external factors could also lead to behavior changes in participants, such as participants generally spending less time at their computers in the face of relaxed restrictions. The control group was implemented as a comparison to identify behavior change which is unrelated to the boosts or nudges. The importance of control groups in privacy research is stressed by Lowry et al. (2017).

During the third week, the interventions for the boost and nudge groups stopped, and the respective information was also removed from the user interface. Nudges and boosts are often said to differ in the way their effects last when the nudge or boost itself is not present anymore (Hertwig, 2017). The last week of the study intended to capture behavior after nudges and boosts were not in effect anymore. Consequently, browsing behavior continued to be collected as described above.

On June 13th, participants received another e-mail reminding them to check their e-mail on June 15th for instructions regarding the end of the study. On June 15th, participants received an e-mail with further questionnaires. The questionnaire measuring privacy knowledge concerning boost information was administered again to evaluate whether the intervention resulted in changes to this metric. Finally participants answered questions on their experiences during the study, and on participant demographics. When sending an answer to this last questionnaire, they received debriefing information about the study, explaining the between-groups design, the

nudges and boosts. The e-mail also contained instructions to access a participation code which allowed them to sign up for course credit compensation or a voucher lottery. By entering a code given in the last questionnaire in a form in the extension's user interface, participants were able to access a participation-code, as well as all possible boosts and nudges, related to their own behavior. This aimed to enable participants in the control condition to also benefit from the presented information if they were interested, since this group had not seen any information during the study. Participants were thanked for their participation and provided with illustrated instructions on how to deinstall the extension. Since there were responses missing from several participants, two follow-up mails were sent to those participants, whose e-mail address was not in the list of participants signed up for either course credits or the voucher lottery. These e-mails were sent two and four days after the end date of the study. Student participants at the University of Regensburg were compensated with course credits, while all participants were able to sign up for a lottery, in which two vouchers were given away. Participants were required to enter their participation code along with their e-mail address, when signing up for compensation, to make sure that they had actually participated in the study, while at the same time maintaining their anonymity. Since the participation codes were assigned at random, it was only possible to verify participation with them, but not to identify a participant. Since both of the winners of the vouchers were local to Regensburg, it was decided to award vouchers for local businesses in that city, to support the local economy (*Der Altstadt-Zehner – ein Gutschein, 1.000 Möglichkeiten*, 2020).

4. Data Analysis

Several preliminary analyses were conducted to describe the sample of participants and their behavior, to investigate the effect of external events which took place during the study, and to replicate findings from previous work, which were used in boosts. The notion of a privacy paradox was also explored for these data. Finally, multiple steps were taken to answer the question, whether the boosts or nudges have an effect on browsing privacy, and whether boosts foster knowledge about privacy.

As stated before, all data analysis in this thesis was conducted using the freely available software Gnu R (R Core Team, 2019), mainly the packages *nlme* for multilevel modeling (Bates et al., 2015), and *car* (Fox & Weisberg, 2019) for regression and diagnostics. The database was accessed through functions from the *RPostgreSQL* package (Conway et al., 2017) and cleaning and ordering of the data was performed using the tidyverse packages (Wickham et al., 2019). Visualizations were generated using the *ggplot*, which is also included in the tidyverse (Wickham et al., 2019). For all analyses performed on these data, significance was assumed at $p = .05$. Any statements from participants were translated from German by the author of the thesis.

When conducting regression analyses, several different metrics were used for case-wise diagnostics. These are summarized here. Potential outliers in a model can be identified by looking at standardized residuals. These should not be larger than |2| in more than 5% of cases, and not larger than |2.5| in 1% of cases (Field et al., 2012). Such potential outliers can further be examined by comparing their leverage, also called a hatvalue, to the average leverage, which is defined as $\frac{(k+1)}{n}$, where k is the number of predictors in the model and n is the total number of cases (Field et al., 2012). Values above three times the average leverage could mean that a

case has undue influence on the overall model (Stevens, 2002). Cook's distance, or Cook's d for short, is also used to identify outliers. Values above 1 are cause for concerns (Cook & Weisberg, 1982). A final metric used, were covariance ratios (CVR), which should be between certain boundaries (Field et al., 2012) defined as follows, whereby the meaning of n and k is the same as above:

$1 - 3 \times \frac{k+1}{n} < CVR < 1 + 3 \times \frac{k+1}{n}$

4.1. Participants

During recruitment, 76 people filled out the survey and entered their contact data. Of those, 68 participants filled out the surveys at the beginning of the study. This is a response rate of 89%. At the end of the study, 60 participants filled out the questionnaires, resulting in a drop-out rate of 12% from the start of the study to the end. The data of the dropped out participants was nevertheless used in analyses where appropriate. Additionally, there was one participant (p 122), who did not fill out surveys, but whose browsing data are nevertheless present.

Of those who finished the study, 30 identified as female, 29 as male and 1 as diverse. The youngest participant was 19, and the oldest 60, and the average participant age was 25.52 (SD=8.8). Most of the participants (55) called German their mothertongue, one spoke German almost at the level of a native speaker, three stated that their command of German was excellent, while one claimed good working knowledge. The participants' level of education was fairly high, with one participant with a Phd, 4 with a Master's degree or equivalent, 12 with a Bachelor's degree and 38 with an entrance qualification to higher education. Three participants named finished vocational training and two named graduation from an intermediate secondary school (German: Realschule) as their highest level of education. This is also somewhat reflected in the participants' occupations: 7 are working, and 53 are studying or are currently in an apprenticeship. Of those in the workforce, 1 works in a computer associated field and 6 do not work in obviously computer associated fields. Among these are two people working in a pharmacy, two working in public service, an English instructor, and a research assistant. Among the students,

38 study something computer associated, with the majority of those studying either media informatics or information science in some capacity, 13 study other subjects, and 2 are in an apprenticeship in a technical, but not directly computer related area.

During the study, 36 of the participants used Chrome, and 24 used Firefox. Windows was the most widely used operating system, used by 47 participants, while 12 used MacOs, and 1 person used Linux. Of the participants finishing the study, 21 were in the boost condition, 18 in the control condition and 21 in the nudge condition. Of all those where browsing data was available, the conditions were not quite as equally distributed, with 25 people in the boost and nudge conditions, but only 19 in the control condition.

Some participants used multiple devices of the same type during the study. Mobile device use was not relevant for this study, since the extension was not available for mobile devices, but 17 participants used a desktop or laptop device other than the one where the study extension was installed. In eight cases, this was due to work related reasons, when participants used a different device for their job. Some of them specified that it was forbidden for them to install an extension such as the study extension on that device. Relocation was the reason to use a different device for four participants, who either were in a location away from their desktop computer, and then used a laptop during that time, or otherwise were somewhere else without their main device. Three participants named device characteristics as the reason for using a different device. They needed a different operating system, or a faster computer for certain activities, such as work on a project, or gaming. The final two reasons did not fit into any of these categories. One participant stated simply "Laptop not enough"(P 95), and this statement was somewhat ambiguous. Another participant used multiple devices at the same time and did not switch to the study device every time for "just looking something up for a moment"(P 23).

Most of these 17 participants used the study device more than their other device. Two used the study device over 90% of the time, and nine used it between 60% and 90% of the time. Three stated that they used both devices roughly the same amount of time, while three used the study device only between 10 and 40% of the

time. A second question was used to validate this question. It asked directly for the percentage of time the study device was used. According to this question the study device was used between 30% and 98% of the time, and on average 67.8% of the time ($SD = 19.6\%$). This was validated by comparing the answers to the two questions. All the answers for the second question but one were in the same range, and can be considered valid. One participant (P 95)for whom this was not the case stated using the study device between 10 and 40% of the time in one question, but 50% in the other.

Most of the participants in the study (50) used their usual browser, but three used a different browser to participate, since their usual browser was not supported, and seven used a browser, which they normally used, but not as much as they did during the study. The three which used a different browser did not notice severe differences between their usual browser, and the one they used in the study, but mainly some performance and usability issues between Chrome and Safari. One person also noted that their passwords were saved in the other browser. So in general, participants were used to the browser they used in the study.

Multiple browsers were used by 15 participants during the study for various reasons. Brave was used by one person, Tor and Firefox by two people respectively, three people each used Chrome, Opera and Edge, and four people used Safari in addition to the study browser. Multiple mentions were possible. Some participants named multiple reasons for using a different browser. Better privacy was a reason to use different browser, namely Tor, Opera or Brave, for four participants. It has to be noted, that three of these named another browser, either Chrome or Firefox beside these, and it is unclear, whether the reason to use these was also privacy. Four participants used a different browser because of functionality associated with the other browser in their default settings or their settings prior to the study, such as PDF-files or links being opened in a certain browser. A different browser was also used in the case of performance issues, or missing functionality in the study browser, for example some software not working in that browser. This was the case for four participants. Four participants used another browser, since that browser

was adjusted to their normal usage patterns, e.g. their passwords were saved there, or they always use certain websites there. One participant gave a reason which did not fit into any of these categories, and which was also hard to understand: "because of the study, and because I also use it on my mobile"(P 80).

As with device usage, the percentage of time that the study browser was used, was measured with two questions. Most of the participants used the study browser either over 90% of the time, or between 60% and 90% of the time, one participant used it around the same amount of time as other browsers, and one person used it only between 10 and 40% of the time. The study browser was used between 20% and 97% of the time ($M = 78.5\%$, $SD = 26.0\%$). Most of the answers to the second question corresponded to those of the first question. There were four answers, which did not fit exactly. Three of those were very close to the next category, they all indicated that they used the study browser over 90% of the time, but named their usage percentage as 90% (2 people) or 85% (1 person), so the participants' general estimation seems to be accurate in these cases. However, one participant (P 113) stated that they used the study browser between 60 and 90% of the time in one question, but only 25% of the time in the other question, which is very different.

In total, about half of the participants (32) used only one browser and one desktop or laptop device during the study, the one with the study extension. A different device was used by 13 participants, and a different browser was used to some extent by 11 participants. Only four participants used both a different device and a different browser.

Many of the participants were familiar with browser extensions before the study, but 19 had no browser extensions installed apart from the study extension. The participants had between 0 and 17 browser extensions installed at the end of the study, apart from the study extension ($M = 2.67$, $SD = 3.40$).

The following assessments of participants are provided for N=68. There was a tendency towards affinity for technology among the participants. The average over all nine 6-point Likert items of the affinity for technology scale (Franke et al., 2019) was 4.10 ($SD = 1.07$), with a minimum of 1.78, and a maximum of 5.78.

Privacy knowledge, as measured by OPLIS was high compared to general percentile ranks, which take into account the whole population, not direct peers with respect to age and gender. On average, participants were in the 74.6th percentile (SD=20.4), with the least knowledgeable participant at a percentile rank of 27, and the most knowledgeable participants at the percentile rank of 100. This is not surprising, considering that many participants study or work in a domain related to computers.

The Likert items making up IUIPC were measured from 1 to 7. Malhotra et al. (2004) do not give instructions for how to interpret the different constructs, which are part of privacy concern, but for the purpose of this thesis, an average was used to provide a single value for the level of privacy concern. First, the averages of each of the three constructs awareness, collection and control were calculated. An average of these averages made up the value of privacy concern, when it is used in this study. Participants' privacy concern, as measured by IUIPC was moderate to high ($M = 5.59$, $SD = 0.71$).

Participants used between 0 and 7 ($M = 2.67$, $SD = 1.91$) different measures to protect their privacy. Some of them also used browsers or search engines, which protect privacy more than the most common ones. Tor is used daily or weekly by 6 participants, and monthly by 10. Brave is used daily by two participants, and Opera by two participants. Firefox can be considered more privacy protective than Chrome, and 25 participants use it daily, 5 weekly, and 13 monthly. As for search engines, Duckduckgo is being used daily or weekly by 12 participants and monthly by 2. Qwant is used weekly by 1 person, and Ecosia, which takes a middle ground concerning privacy, was used daily or weekly by 9 people, and monthly by 13.

4.2. Descriptive Statistics on Participants' Behavior throughout the Study

To get a sense of general browsing behavior, several features of the participants' browsing behavior during the study were visualized. Even though some participants filled out the final survey some time after the end of the study on June 15th,

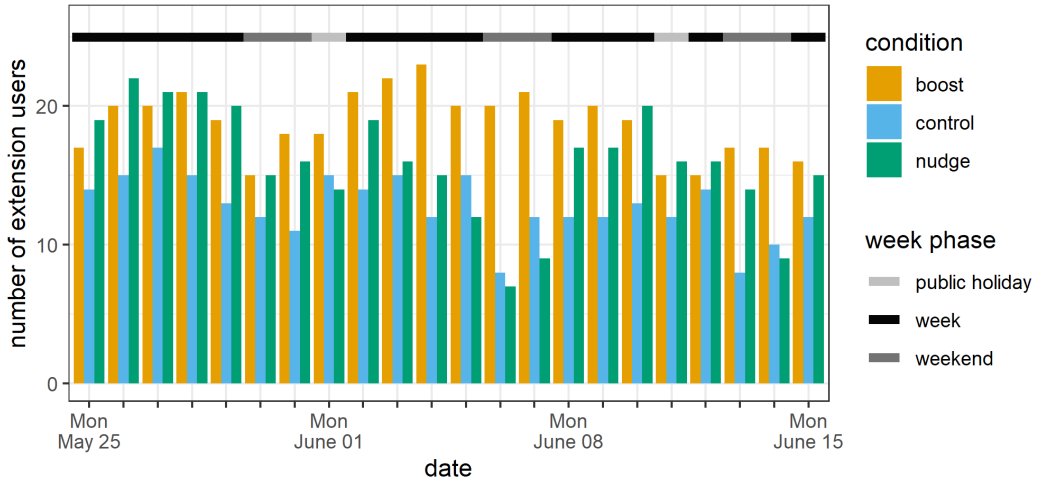


Figure 4.1.: Number of active internet users per condition per day

and thus usage data exists after that time, only browsing behavior up to that date are visualized here. After June 15th, the number of remaining participants grows steadily lower, and is not representative of behavior during the study anymore.

The number of participants for whom the study extension collected data on each given day is visualized in Figure 4.1. It has to be taken into account that while the total number of participants in the boost and nudge conditions were equal, the number of participants in the control condition was smaller, so these bars are always lower. Visualizing the numbers as percentages per condition would have obscured the total number of users however. At the beginning of the study, participation numbers rise over the first two or three days. This is likely because some participants did not start the study on the first day, but later. Over all three conditions, the number of daily internet users declines slightly towards the end of the study. This may be due to some participants terminating their participation due to too many problems with the extension, or for other reasons. For the nudge and control conditions, there is also a clear pattern of less participants using the internet on weekends, than on weekdays. In the boost condition, this pattern is less obvious, except during the first weekend.

There were two public holidays in the period of the study, Whit Monday on June 1st, and Corpus Christi on June 11th. On Whit Monday, there were less internet users in the boost and nudge condition, than during the directly surrounding week-

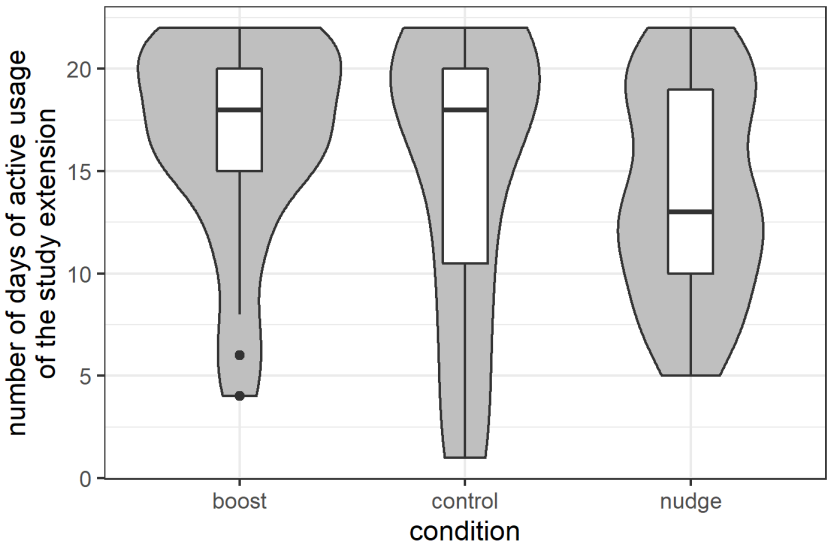


Figure 4.2.: Boxplots and violin plots of the number of days of active internet usage per participant during the study, per condition

days, but this was not the case for those in the control condition. On Corpus Christi, there were also less internet users than on the weekdays before, for all three conditions. However, on the next day, the number of active participants did not increase again. This may be, since the public holiday was on a Thursday, and people may also take Friday off to have a long weekend, at least for those participants who are in the workforce.

The number of days of active usage, meaning days where a participant visited at least one website are shown in Figure 4.2. Overall, participants were active on between 1 and 22 days ($M = 15.1$, $SD = 5.7$). While in the boost and control condition, a large amount of the participants visited websites on more than 15 days during the three-week study, the nudge condition exhibits a bi-modal distribution, with many participants being active on most of the days of the study, but also many participants being active on just a little more than half of the days of the study.

The number of website visits of each participant per day are visualized in Figure 4.3. The y axis on this graph is logarithmically transformed and loess smoothing (local regression) was applied to visualize trends. When a person did not visit any websites on a given day, their data is not included in this smoothing, and is instead displayed at the very bottom of the graph, so that these data points are only half vis-

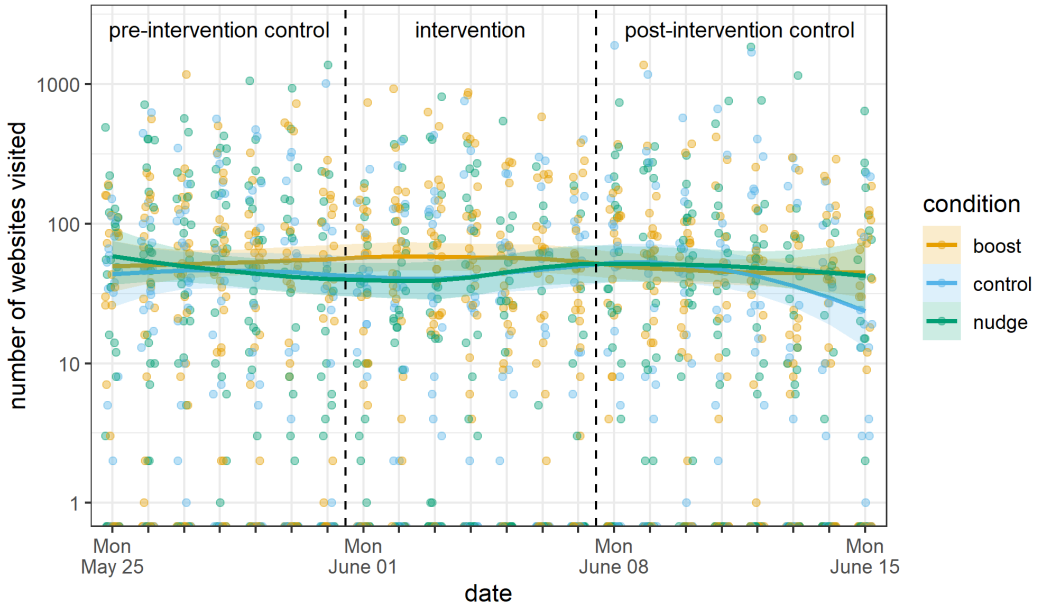
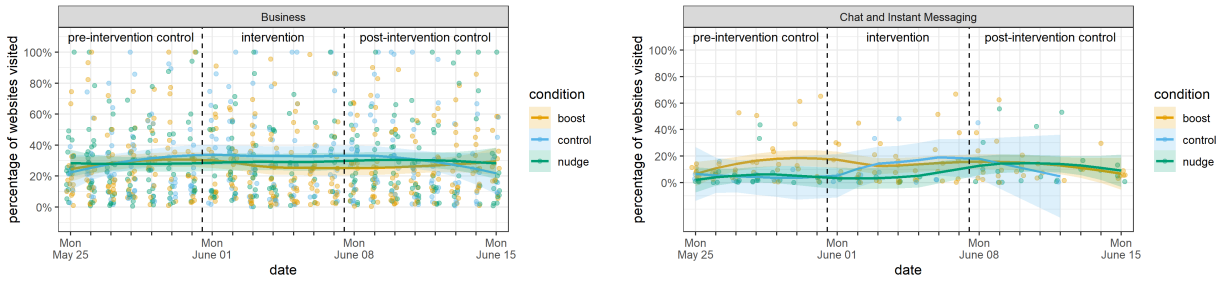


Figure 4.3.: Daily number of website visits, y axis is logarithmic

ible. The number of website visits for the three conditions are relatively similar, and relatively stable during the whole duration of the study. Website visits for the participants in the boost and control conditions decrease slightly at the middle of the first week of the study, and then increase again from the middle of the second week of the study, until they are at the level of those in the boost condition again. For the participants in the control, the total number of website visits begins to decline on the last Thursday in the study.

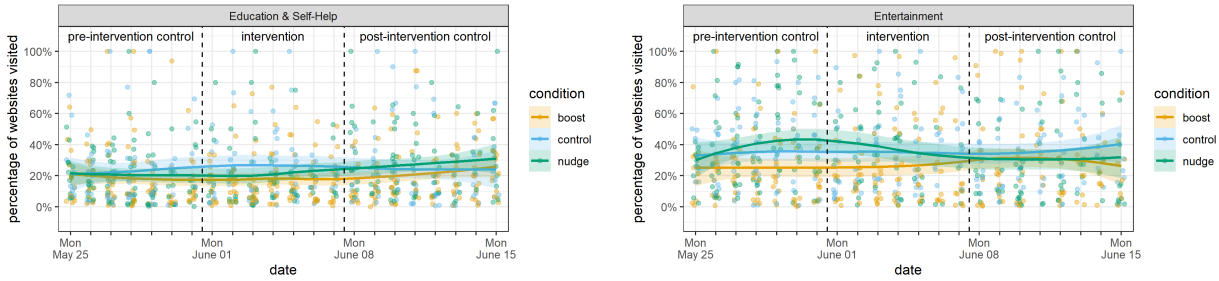
It was also investigated whether certain categories of websites were visited in different amounts during different phases of the study. The first category assigned to a website visit was considered to be the main category, and these categories were taken into account. Only those categories, which registered more than 1000 visits, were included in the analysis. All other categories were subsumed as *Other*.

Figure 4.4 shows graphs depicting the percentage that visits to each of these categories constituted among the total number of website visits by a participant. Again, loess smoothing was applied to make trends visible. For most of these categories, the average percentages of website visits per category are roughly the same for all three conditions, and more or less constant across the duration of the study. For example, for *business* websites, this means about 30%, for *search engines and portals*,



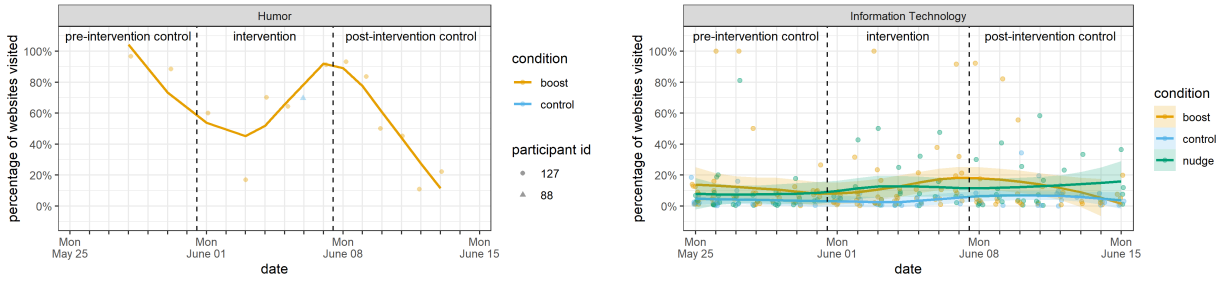
(a) Business

(b) Chat and Instant Messaging



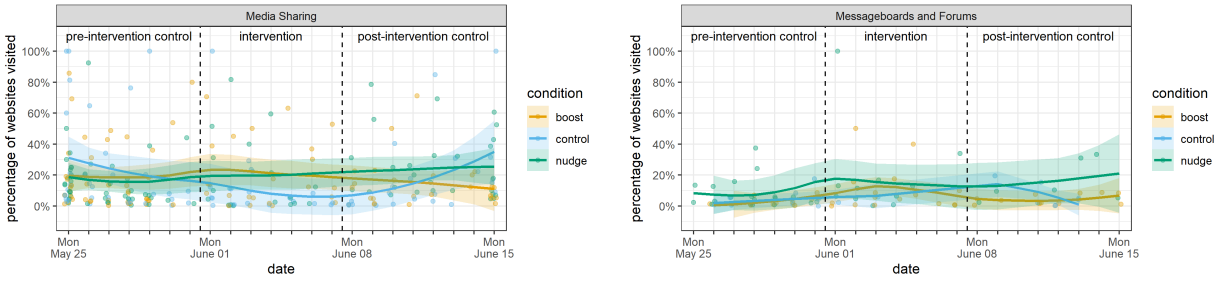
(c) Education & Self-Help

(d) Entertainment



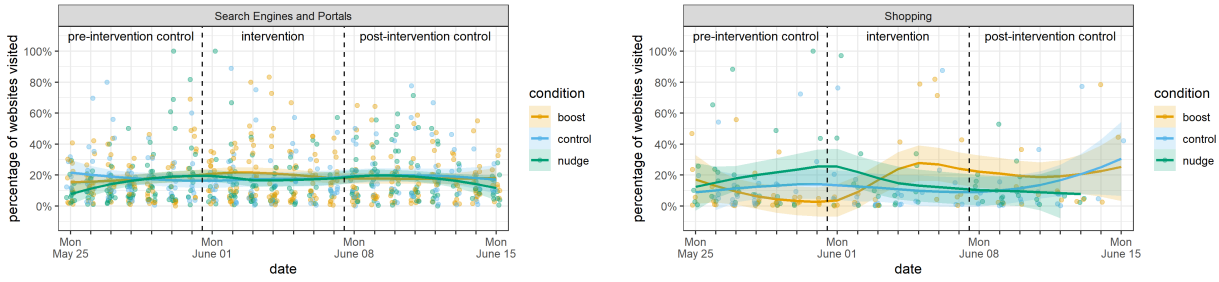
(e) Humor

(f) Information technology



(g) Media Sharing

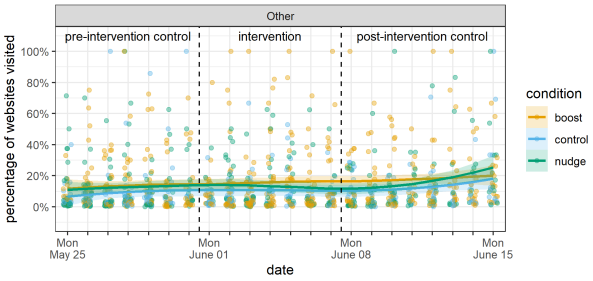
(h) Messageboards and Forums



(i) Search Engines and Portals

(j) Shopping

Figure 4.4.: Daily percentage of website visits for different categories



(k) Other

Figure 4.4.: Daily percentage of website visits for different categories (Continued)

the average percentage is around 20%, and for *information technology* websites, the average percentage is between 10% and 20%. Contrary to the number of active participants per given day, weekdays and weekends do not show differing patterns.

Some of the website types are visited by more of the participants, like business websites, education websites, or search engine and portal sites, while others are visited by fewer of the participants, like chat and instant messaging websites or messageboards and forums. While the latter also show a similar pattern of more or less constant percentages, there is more variability in those website categories which were visited by fewer participants. This is likely because a single participant's visits on a given day have more influence on the whole trend, than for those website categries which were visited by a larger amount of participants on a given day.

There are some types of websites where this pattern does not hold. For example, although websites from the *humor* category were visited more than 1000 times, almost all these visits came from a single participant in the boost condition (P 127), and only on one day, another participant visited sites of this category. Thus, these data cannot be considered representative for visit patterns to humor websites in general. Another example is shopping websites, where the loess approximation for the boost condition shows a low number of visits for the first half of the study, and then a higher number of visits for the second half of the study, while the number of visits from participants in the nudge condition start out around 20%, and then decline during the second week. Website visits to shopping websites are more or less constant for participants in the control condition, but increase in the last half of

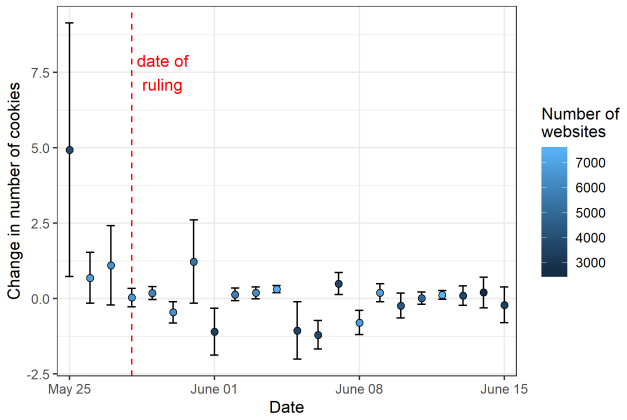
the third week. However, there is only a small amount of data points per condition, so these trends are not stable.

Since the website categories of entertainment and education are mentioned in boosts during the study, they shall also be examined in some more detail. The average number of visits to education websites is about 20% over the course of the study. The percentage of visits is a little higher for the control group from the second half of the first week through the second week of the study than for the other two conditions. During the third week, the percentage in the nudge condition is a little higher. Throughout all three study weeks, the percentage of visits to education websites is lowest for participants in the boost condition, although all in all, the percentages are similar for all three conditions. The percentages diverge more for entertainment websites, and they are generally higher than for education websites, fluctuating between just over 20% and 40%. The percentages of visits to entertainment websites remain more or less constant for the control group at approximately 35% and the boost group at about 25%, although the percentages increase slightly for participants in boost group at the end of the intervention week. There is more variability concerning the percentage of visits to entertainment sites for the participants in the nudge condition. These increase from 30% to slightly over 40% throughout the first week of the study and then decrease to 30% again through the second week and remain constant in the third week of the study.

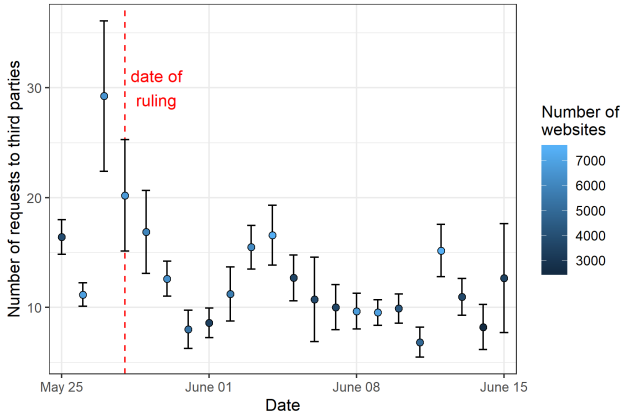
In summary, with some exceptions, the participants in this study exhibit similar browsing behavior across conditions and phases of the study.

4.3. Effect of Current Events

During the first week of the study, on May 28th 2020, the German federal court ruled that preselecting privacy-invasive choices in cookie notices was illegal (Tageschau, 2020). This might influence the proxies for browsing privacy used in this study, and thus was investigated graphically. The results are presented in Figure 4.5. There is an obvious peak in the change to number of cookies on the first day of the study, see Figure 4.5a which stems from the process used to gain information on the number of



(a) Change in number of cookies per website



(b) Number of requests to third parties per website

Figure 4.5.: Means and 95% confidence intervals of proxies for browsing privacy before and after May 28th 2020, color represents the number of websites visited on a certain day and used in the calculation of the summary

cookies. The change in cookies per website was measured by comparing the number of cookies on a current website to the number of cookies on the previous website. After installing the extension, there was no previous website for the first website visited, so all the cookies already present in a participant's browser were attributed to this first website visit.

In general it is hard to compare data from before and after the ruling, since the ruling was published only three days into the study, and so there was much less data from before the ruling than after. This was also the reason why it was chosen to examine the data only graphically, and not with methods of inferential statistics. However, the mean change in cookies per day seems to be lower after the ruling, than before. Nevertheless, when excluding the first day of the study, and when taking into account that some participants did not start the study on the first day, but rather later into the week, this difference can also be attributed to normal variance in the users' behavior.

For the number of third party requests per website (see Figure 4.5b), there is no real noticeable trend, although on the day before the ruling, the number of third party requests is higher per website than on any other day during the study phase. While there is some fluctuation, it cannot be attributed to any obvious events, such as the change between weekends and weekdays. The labeled first days of each study phase were Mondays, but weekdays do not seem to be different from weekends, the two days before the labeled days. In general, from the data available, the ruling does not seem to have made a difference.

4.4. Website Type and Privacy

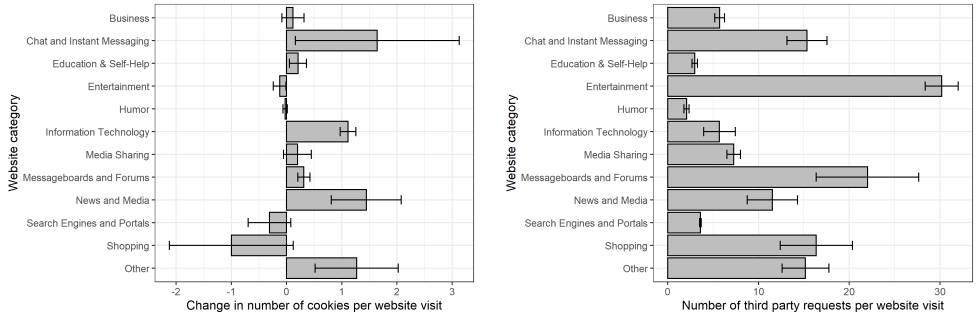
Information to use in boosts was gained from literature. It was not possible to track the use of adblockers during the study, and the number of users who elected to block all third party cookies was too small to evaluate meaningfully. Similarly, the outcome variables necessary to investigate the usefulness of the boost concerning private or incognito browsing were not tracked in this study. However, it shall be investigated whether it is possible to replicate the finding that news and entertain-

ment websites are more privacy invasive than other sites, and education sites are less privacy invasive than others. While the boost concerning news websites was not used in the main study, it too, will be investigated here for the sake of completeness.

In general, when looking only at the first category, which is considered to be the main category of a website, websites from 30 different categories were visited during and after the study. For this analysis, data was also taken from after the official end of the study. The access to the categorization API was available until June 24th, and some participants did not uninstall the study extension until as late as a week after the official end of the study. The number of visits per website category varied from a minimum of 2 (drugs category) to a maximum of 33197 (business category), but the mean number of visits per category was 4236 (SD=8429). The high number of websites categorized as business is most likely because business was a general main category which was further specified with other categories concerning the type of business.

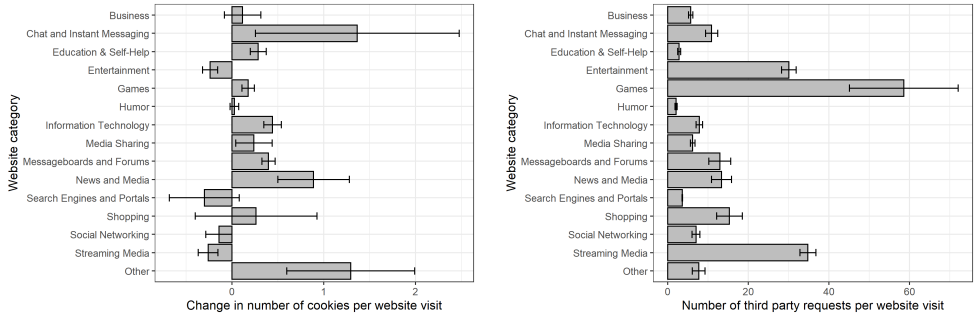
When not only taking into account the first category assigned to each website, websites of 36 different categories were visited at least 2 times (categories drugs, advertising, religion and proxy and filter avoidance) and at most 33197 times (business category). The mean amount of times websites of a certain category were visited was 5426 (SD=9793). It has to be noted that since a website can be assigned as many as three categories in this scheme of assessing categories, some websites appear multiple times in the analyses based on this categorization.

Bar charts with 95% confidence intervals were used to visualize the number of third party requests and change in cookies per website of a certain type, see Figure 4.6. They were chosen because visualization measures which show distributions in more detail, such as boxplots, or violin plots, were cluttered and unclear due to many outliers, some of which were very far from the mean. In addition there was a large amount of websites which had values of zero for the browsing privacy proxies. It has to be noted that the confidence intervals were calculated assuming independence between websites of different website visit ids, but this is not the case, since



(a) Number of cookie changes per website main/first category

(b) Number of third party requests per website main/first category



(c) Number of cookie changes per website for all assigned categories

(d) Number of third party requests per website for all assigned categories

Figure 4.6.: Means and 95% confidence intervals, upper row shows them main categories, lower row for all categories assigned (some websites may be represented multiple times in the data)

the same website may have been visited multiple times by different participants. Due to the measures taken to protect the participants' privacy, it is not possible to know this, so the confidence intervals should be interpreted with caution. All categories, except the three mentioned in possible boosts, which were visited less than 1000 times, were included in a summary category termed *Other*.

Looking at the charts for the change in cookies, education websites do seem to have a low number of cookie changes, but so do entertainment websites. However there are other categories, which are related to entertainment, such as humor, messageboards and forums, games, or chat and instant messaging. Some of these have higher numbers of cookie changes, but others do not, so it is not clear how to compare this finding to previous work, which used different tools to categorize websites (Urban et al., 2020). News websites are among those categories with the highest positive number of changes in cookies.

When examining the number of third party requests, education sites are again among the least privacy invasive categories, similar to previous work. Websites belonging to the entertainment and related categories have the highest average number of third party requests. The deviation is with news websites, which is in a medium range concerning the number of third party requests, both considering only main categories, and considering all categories assigned.

So in general, this analysis confirms the trend of previous work that news and entertainment websites are more privacy invasive considering cookies and third party requests than education websites. While it was not replicated that news websites were more privacy invasive than entertainment websites, it also has to be taken into account that this is a much smaller sample than the ones analyzed by Englehardt & Narayanan (2016) and Urban et al. (2020).

4.5. Relationship between Privacy Knowledge, Privacy Concern, and Privacy Behavior

Measures for privacy knowledge (OPLIS), privacy concern (IUIPC), as well as self-reported actions to protect privacy were taken at the beginning of the study. To

find out whether the privacy paradox applies for the participants in this study, the influence of PRIVACY KNOWLEDGE and PRIVACY CONCERNS as independent variables on self-reported privacy related actions as the dependent variable was investigated. It might seem intuitive, that more protective actions are undertaken with heightened privacy concern, however, according to literature supporting the privacy paradox, this is not the case (Taddicken, 2014; Kokolakis, 2017). Privacy knowledge is hypothesized to influence the number of protective actions as well, since without knowledge about privacy, it is hard to judge threats, and know which actions to take.

Since the number of privacy actions was measured as a count variable, that being the number of actions undertaken by participants to protect their privacy, more wide-spread regression approaches were not appropriate since they assume a continuous dependent variable measured at the interval scale. For this reason, a Poisson regression was conducted instead. In this form of regression, the dependent variable is instead a count variable, that is to say a non-negative integer, which should be theoretically unconstrained (Coxe et al., 2009). Since it would have been possible to measure a theoretically unbounded number of actions taken to protect privacy, the number of privacy related actions can be considered such a variable. Further assumptions are that the observations are independent of each other, and that the distribution of counts for the dependent variable follows a Poisson distribution. Independence is ensured because each observation stems from a different participant, observations were all made at the beginning of the study, so participants were not yet grouped by condition, and the participants did not have the chance to communicate with each-other. The distributional assumption was tested using the poissonness plot introduced by Hoaglin (1980). In this plot the quantity $log(x_k) + log(k!)$ is graphed against the category k, where x_k is the observed frequency for a category k. It is interpreted similarly to a normal Q-Q plot, so that when the points follow a straight line, the sample distribution is considered to follow a Poisson distribution (Hoaglin, 1980). As can be seen in Figure 4.7, this is the case for the dependent variable in this study. A final assumption for Poisson regression is that the conditional

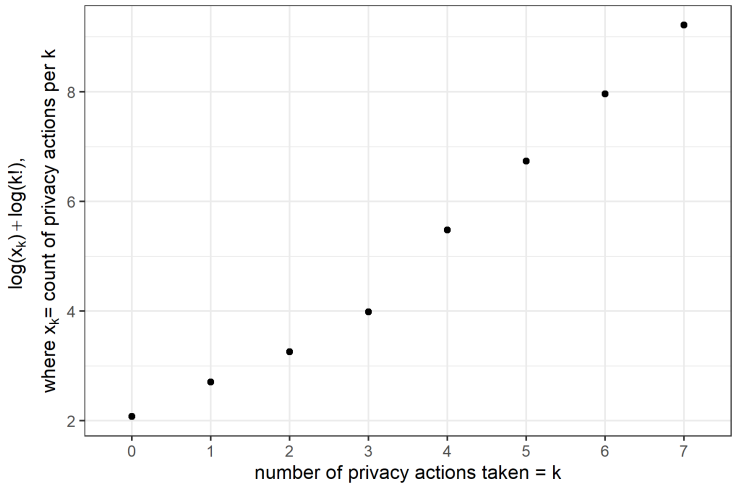


Figure 4.7.: Visual examination of Poisson distribution according to Hoaglin (1980)

variance is equal to the conditional mean. If this is not the case, and the variance is larger than the mean, this is called overdispersion. The opposite case, when variance is smaller than the mean, is called underdispersion. While in the total data, the mean number of protective actions is 2.66 and the variance is 3.63, which is not quite equal, the assumption refers to conditional variance and conditional mean. Consequently, tests of over and underdispersion were conducted with the fitted models to assess this assumption.

The starting point was a model without any predictors, which served as the baseline. Parameter estimates for this model are in Table 4.1. Privacy knowledge was added as the first predictor. This significantly improved the fit of the model, $\chi^2(1) = 4.12, p = .04$. Next, privacy concern was added to the model as a predictor, which also significantly improved the fit, $\chi^2(1) = 6.68, p = .010$. Finally, the interaction between privacy knowledge and privacy concern was added to the model. Again, the model fit was significantly improved, $\chi^2(1) = 9.44, p = .002$. The parameter estimates for all these models are in Table 4.1.

The assumption of equidispersion was tested for the final model using the *dispersiontest*-function from the *AER* package (Kleiber & Zeileis, 2008). The alternative hypothesis that dispersion is not equal to 1 could be rejected, $z = 0.48, p = .63$. The dispersion in the sample was estimated to be at 1.08, which is reasonably close to 1.

model	predictor	estimate	SE estimate	robust SE	robust CI		z	p	robust p
					lower	upper			
baseline	intercept	0.98	0.07	0.09	0.81	1.15	13.17	<.001	<.001
privacy knowledge	intercept	0.11	0.44	0.47	-0.82	1.04	0.24	.81	.82
	OPLIS	0.06	0.03	0.03	-0.004	0.12	2.02	.04	.07
with privacy knowledge and privacy concern	intercept	-1.39	0.75	0.89	-3.14	0.35	-1.87	.06	.12
	OPLIS	0.05	0.03	0.03	-0.01	0.11	1.77	.08	.11
	IUIPC	0.28	0.11	0.13	0.02	0.54	2.55	.01	.03
with interaction between privacy knowledge and privacy concern	intercept	8.66	3.27	3.39	2.03	15.3	2.65	.008	.01
	OPLIS	-0.62	0.22	0.24	-1.08	-0.16	-2.85	.004	.009
	IUIPC	-1.49	0.58	0.58	-2.62	-0.35	-2.57	.01	.01
	OLIS:IUIPC	0.12	0.04	0.04	0.04	0.20	3.10	.002	.003

Table 4.1.: Parameter estimates with robust 95% confidence intervals for main Poisson regression models

To not only assess the comparative model fit, but how well the model fits the data, a Chi-square goodness of fit test was performed using the residual deviance and the corresponding degrees of freedom, which for the final model is 84.79 on 64 degrees of freedom. The p-value for this test is $.04$, which would lead to the conclusion that the final model does not fit the data well. When interpreting this result, it has to be taken into account that the sample size in this study is relatively small, which may lead to an incorreect type 1 error rate (Bartlett, 2014). When means are small, the deviance goodness of fit test, as conducted above, may not be appropriate (Pawitan, 2001), as cited by Bartlett (2014). The mean of our dependent variable here is only 2.66 on a scale from 0 to infinity, which is relatively small, so the test may not be appropriate. Instead case-wise diagnostics were obtained. In the data, there are 8 cases with standardized residuals larger than |2|, which is above the threshold of 5%. In this sample, with N=68 , 5% are 3.4 cases, which is rounded to 3 cases, and

1% is equivalent to 0.68 cases, rounded to 1 case. These cases were then inspected in more detail.

There were no cases where the standardized residuals were larger than |2.5|, so this is below the theshold of 1%. Cook's distance measures the influence of a single case on the model, and values above 1 are a reason for concern (Cook & Weisberg, 1982). It was calculated using the *cooks.distance*-function from base R, and there were no cases where Cook's distance exceeded 1, and the largest value, 0.22, was still well below this threshold. Leverage, or hat values, also measure influence and were obtained using the *hatvalues*-function in base R. A single hat value should not be more than 3 times the average leverage (Stevens, 2002). The average leverage is calculated as $\frac{k+1}{n}$ with k as the number of predictors and n as the number of observations. For the model which is being investigated, k is three, since both privacy knowledge, privacy concern, and their interaction is included as a predictor, and n, as stated above is 68. This means, that the average leverage in this case is about 0.059. There were no cases of hatvalues larger than 3 times this value in the examined cases. In general, the case-wise diagnostics show that the model is fairly reliable.

While the significant main effects cannot be interpreted in the presence of an interaction (Field et al., 2012), the interpretation of the signification interaction between PRIVACY KNOWLEDGE and PRIVACY CONCERNS is facilitated by looking at Figure 4.8, where the interaction between PRIVACY KNOWLEDGE and PRIVACY CONCERN was plotted using the *interact.plot*-function from the *interactions* package (Long, 2019). It shows that when PRIVACY CONCERN is above average (+1 SD), the number of privacy protective actions becomes larger with growing PRIVACY KNOWLEDGE. With average PRIVACY CONCERN, this trend is not as obvious, and the slope of the regression line is much less steep, but the predicted number of protective actions still becomes larger with growing PRIVACY KNOWLEDGE. On the contrary, when PRIVACY CONCERN is below average (-1 SD), the number of privacy protective actions does not become larger with growing PRIVACY KNOWLEDGE, but even decreases, as is shown by the negative slope of the regression line.

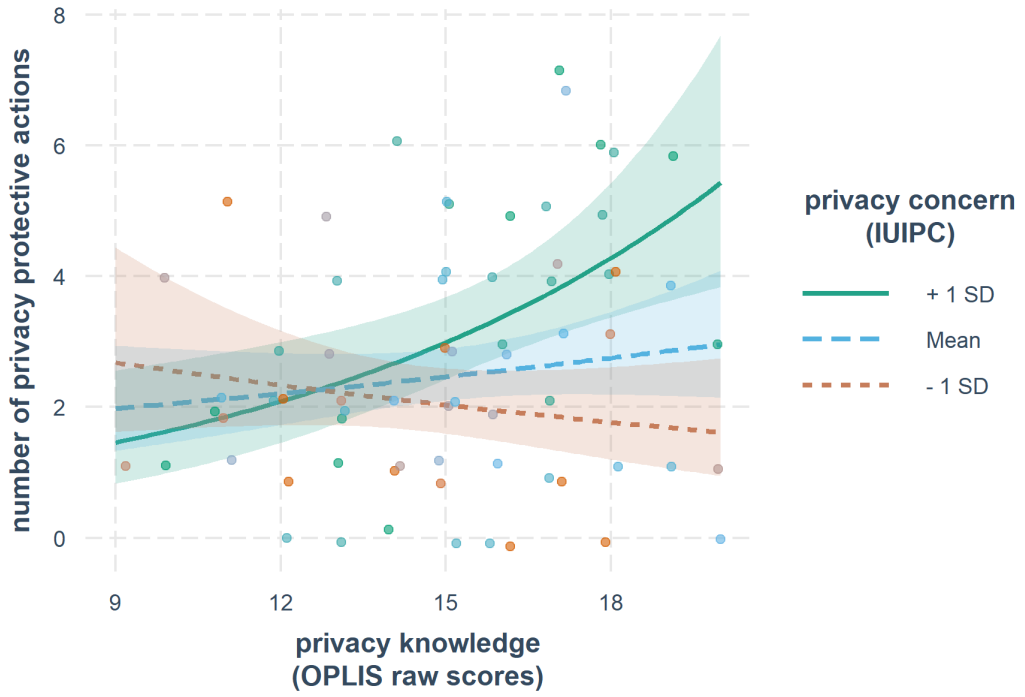


Figure 4.8.: Interaction of final model, regression lines for mean and mean +- 1 SD, with 95% CI, color depicts value of IUIPC, points slightly jittered to avoid overplotting

In an exploratory procedure, further predictors were added to the model, to investigate their influence on the dependent variable. Since only one predictor was added to the model at a time, these models are not nested, and thus are compared to the model using PRIVACY KNOWLEDGE, PRIVACY CONCERN and their interaction as predictors in the following. These were TRUST BELIEFS and RISK BELIEFS which have been used as predictors in the original work on privacy concerns (Malhotra et al., 2004), and AFFINITY FOR TECHNOLOGY, since it was hypothesized, that people with a higher affinity for technology would be more likely to translate concerns into action. Neither TRUST BELIEFS, $\chi^2(1) = 0.28, p = .60$, nor RISK BELIEFS, $\chi^2(1) = 0.20, p = 65$, nor AFFINITY FOR TECHNOLOGY $\chi^2(1) = 3.26, p = .07$, significantly improved the fit over the previous model, although including AFFINITY FOR TECHNOLOGY did provide a slight improvement. In this case, the Akaike information criterion (AIC) for the final model without additional predictors was 263 and with AFFINITY FOR TECHNOLOGY added, this improved to 262. Nevertheless the parameter estimates for these models are in Table 4.2.

model	predictor	estimate	SE estimate	robust SE	robust CI		z	p	robust p
					lower	upper			
with risk beliefs	Constant	8.75	3.27	3.39	2.12	15.39	2.68	.007	.010
	OPLIS	-0.62	0.22	0.24	-1.08	-0.15	-2.84	.005	.009
	IUIPC	-1.45	0.59	0.58	-2.58	-0.31	-2.47	.01	.01
	OPLIS:IUIPC	0.12	0.04	0.04	0.04	0.19	3.06	.002	.004
	risk beliefs	-0.06	0.14	0.15	-0.36	0.23	-0.45	.65	.68
with trusting beliefs	Constant	8.48	3.26	3.28	2.06	14.9	2.60	.009	.010
	OPLIS	-0.60	0.22	0.24	-1.06	-0.13	-2.70	.007	.01
	IUIPC	-1.43	0.59	0.57	-2.54	-0.31	-2.44	.015	.01
	OPLIS:IUIPC	0.11	0.04	0.04	0.03	0.19	2.91	.004	.005
	Trusting beliefs	-0.04	0.08	0.09	-0.22	0.14	-0.53	.60	.66
with affinity for technology	Constant	7.60	3.28	3.22	1.29	13.9	2.31	.021	.018
	OPLIS	-0.59	0.22	0.23	-1.04	-0.14	-2.74	.006	.010
	IUIPC	-1.37	0.58	0.56	-2.46	-0.28	-2.38	.017	.014
	OPLIS:IUIPC	0.11	0.04	0.04	0.03	0.19	2.93	.003	.005
	ATI	0.13	0.07	0.07	-0.01	0.28	1.78	.075	.070

Table 4.2.: Parameter estimates with robust 95% confidence intervals for additional exploratory Poisson regression models

These models above take into account self-reported privacy behavior, but this differs from actual behavior (Kokolakis, 2017). The study conducted for this thesis also observed participants' actual behavior. As such it is useful to consider the privacy paradox with actual behavior. Proxies for browsing privacy can be considered somewhat related to privacy behavior, even if they are not quite the same. It is assumed that they reflect the outcomes of privacy behavior in that if a participant behaves in a way to achieve high browsing privacy, then the proxies for browsing privacy, as utilized in this study, will be low.

Only the models with the same predictors as the final Poisson regression model reported above, are reported here, since their main purpose is to provide a comparison between using self-reported privacy behavior and the derivatives of actual behavior. For each of the two proxies for browsing privacy and each participant, an average value was calculated including all the website visits of the participants during the first week. The first week served as a control, to establish a baseline of behavior for all participants when none of them had been exposed to any intervention yet, so all the participants were in the same situation during this week. One participant was excluded from this analysis, because they visited only three websites on the last day of the first week.

Privacy concern and privacy knowledge did not predict a significant amount of the variance in the average number of third party requests during the first week of the study, $F(3, 61) = 0.81, p = .81$, $R^2 = 0.02$, $R^2_{Adjusted} = -0.03$. Likewise, they also did not predict a significant amount of the variance in the average number of changes in cookies during that time period, $F(3, 61) = 0.24, p = .87$, $R^2 = 0.01$, $R^2_{Adjusted} = -0.04$. The parameter estimates of these models are in Table 4.3.

Case-wise diagnostics for the two models are in Table 4.4. Cases with standardized residuals larger than |2| were considered to be potential outliers and, in a normal sample of the size of the one in this study, three such cases would be expected (Field et al., 2012). For both models, this number is surpassed. Further case-wise diagnostics were conducted both for these outliers, and for the total sample. If case-wise diagnostics for these potential outliers are different than what would be ex-

model	parameter	b	SE b	95% confidence interval		p
				lower	upper	
number of third party requests	intercept	69.1	143	-217	355	.63
	OPLIS	-0.68	9.51	-19.7	18.3	.94
	IUIPC	-5.06	25.6	-56.2	46.0	.84
	OPLIS:IUIPC	0.046	1.69	-3.33	3.42	.98
number of changes in cookies	intercept	-12.0	58.8	-130	106	.84
	OPLIS	1.25	3.91	-6.57	9.08	.75
	IUIPC	3.89	10.5	-17.1	24.9	.71
	OPLIS:IUIPC	-0.29	0.70	-1.68	1.10	.68

Table 4.3.: Parameter estimates for multiple regressions models predicting browsing privacy

model DV	sample for number of cases	number of cases				
		standardized residuals outside |2|	standardized residuals outside |2.5|	Cook's D >1	leverage >3 × average leverage	outside covariance ratio boundaries
average number of third party requests	total potential outliers	6 all	0 0	0 0	3 0	13 5
average change in number of cookies	total potential outliers	5 all	5 5	0 0	3 0	16 5

Table 4.4.: Summary of case-wise diagnostics for multiple regression models using proxies for browsing privacy as the DV and privacy concern and privacy knowledge as the IVs

pected, this is a larger source for concern, than for other values, with smaller standardized residuals. In general both models do not seem to be a good fit for the data, as multiple cases cause concern, both among potential outliers and the total sample. This is not surprising given the small amount of variance explained by the models.

It is debatable whether it makes sense to check assumptions for these models, which were mainly fitted to compare the outcome variables derived from actual behavior to the outcome variable used before, which was reported privacy behavior, since these models are not expected to generalize to the population. Nevertheless, assumptions were checked both visually, and using tests, e.g. the Durbin-Watson Test to check the independence of residuals. The assumption of independence of residuals can be assumed both for the model predicting the average number of third party requests (D-W-statistic$= 2.3$, $p = .13$). Multicollinearity among predictors was assessed using the variance inflation factor (VIF), which, on average, should not be much more than 1. For this model, the average VIF was 57.2, and the largest VIF was 89.4, for the interaction between privacy knowledge and privacy concern. Since the largest VIF should be larger than 10, this means that there is a lot of multicollinearity among the predictors. This is not surprising, since one of them is the interaction between the two others, which naturally correlates with both. Multicollinearity causes less precise coefficients and can reduce the power to find significant predictors, but it does not influence the general model fit or the predictions obtained from it (Kutner et al., 2004). Since R^2 for the model is low anyway, and the model was fitted more as a comparison than for the sake of finding significant effects, this is not considered a problem in this analysis. The remaining assumptions were assessed graphically, using a normal Q-Q plot depicted in Figure 4.9 and a plot of residuals against fitted values, shown in Figure 4.10. The values in the latter seem to be distributed in a relatively random fashion, which only indicates slight heteroscedasticity and does not indicate non-linearity. However, the normal Q-Q plot shows evidence of non-normality of residuals, suggests right skew, and perhaps the number of values at one end of the distribution is too high.

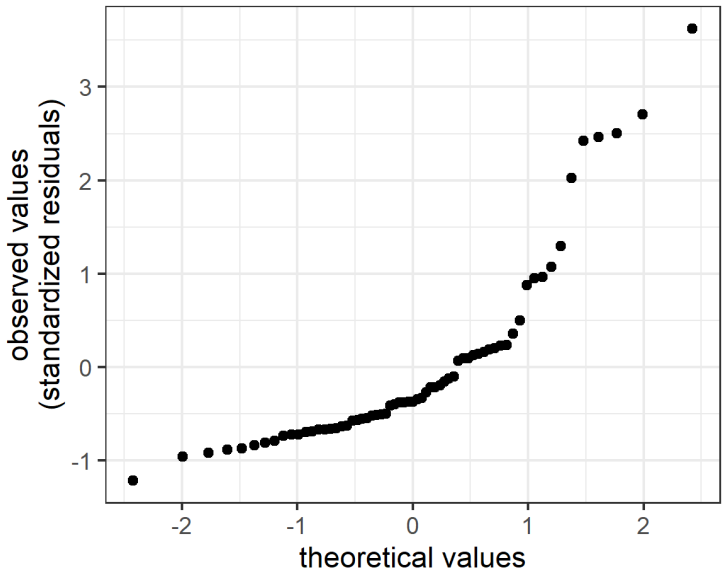


Figure 4.9.: Normal Q-Q plot used to check assumption of normality of residuals for multiple regression model with average number of third party requests as the dependent variable

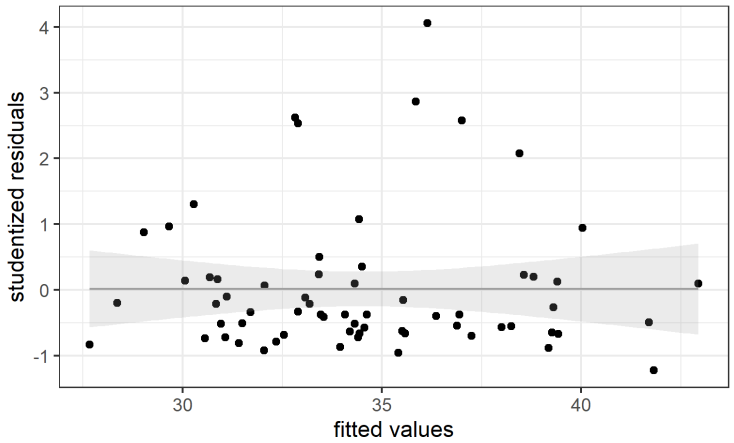


Figure 4.10.: Diagnostic plot for multiple regression model with average number of third party requests as the dependent variable

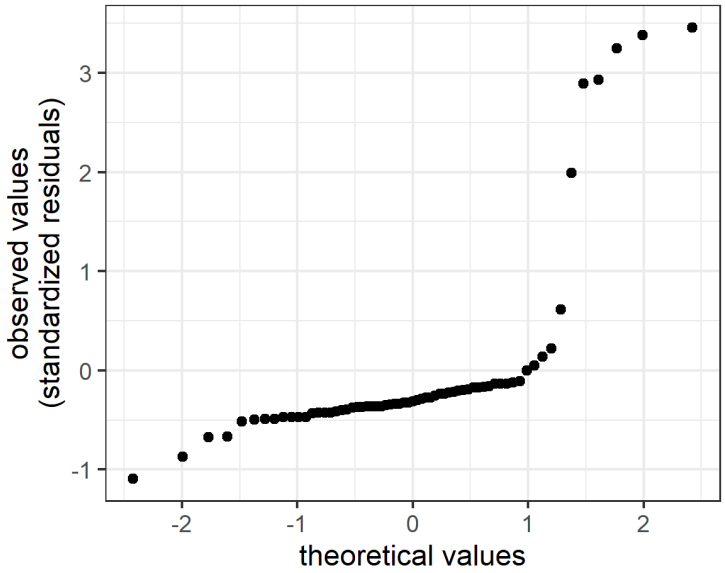


Figure 4.11.: Normal Q-Q plot used to check assumption of normality of residuals for multiple regression model with average change in number of cookies as the dependent variable

Residuals were independent for the model predicting the average number of cookie changes ($D - W - statistic = 2.31$, $p = .12$). Like for the previous model, there was multicollinearity among the predictors, with the average VIF being 57.2, and the largest VIF, the one for the interaction, being 89.4. The purpose of this model is similar to the one with the average number of third party requests as the dependent variable and, since the exact parameter estimate, and the significance are not the main area of interest here, this is not a large problem. Again, further assumptions were evaluated graphically. A normal Q-Q plot and a plot of residuals against fitted values are shown in Figure 4.11 and Figure 4.12. The residuals for this model are even less normal: The plot reveals right skew and a heavy tail on one side of the distribution. The points in Figure 4.12 are not distributed randomly either. In general, there is little variation, and most studentized residuals are very close to zero. Between fitted values of two and six, however, there are several studentized residuals which are larger than two. This may be an indicator of heteroscedasticity.

In general, the two models probably do not generalize very well to the population. This is in stark contrast to the fairly reliable Poisson regression model and seems

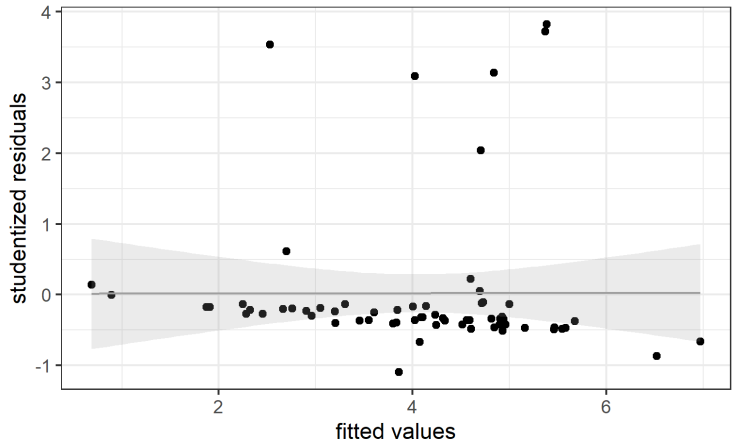


Figure 4.12.: Diagnostic plot for multiple regression model with average change in number of cookies as the dependent variable

to suggest that for actual behavior, rather than self-reported behavior, the privacy paradox is valid, meaning that privacy concern does not influence privacy behavior.

4.6. Actions mentioned in boosts

While the social nudges used in the study did not propagate a certain behavior, but only displayed users' rank with respect to certain measures, the boosts informed about threats to privacy (e.g. entertainment-boost) or ways to protect privacy (e.g. adblocking-boost). Following is an analysis whether behavioral changes are related directly to these boosts occurred.

Measuring setting changes proved somewhat difficult since there was not a specific API to track setting changes. Instead, the states of several privacy related settings were tracked with each visited website, as described in Subsection 3.4.4, this was the Do not track setting, the cookie blocking policy, the WebRTC IP handling policy, and whether the website was visited in private browsing mode or not. To assess how often these settings were changed, the website visits were grouped by participant and ordered chronologically. When any of these settings were not the same in a website visit, as in the one before it, this was counted as a change to settings. There were no changes to the WebRTC IP handling policy, which is not surprising, since this setting is not available through the graphical user interface, but rather only through about:config. However, there were changes to all other settings.

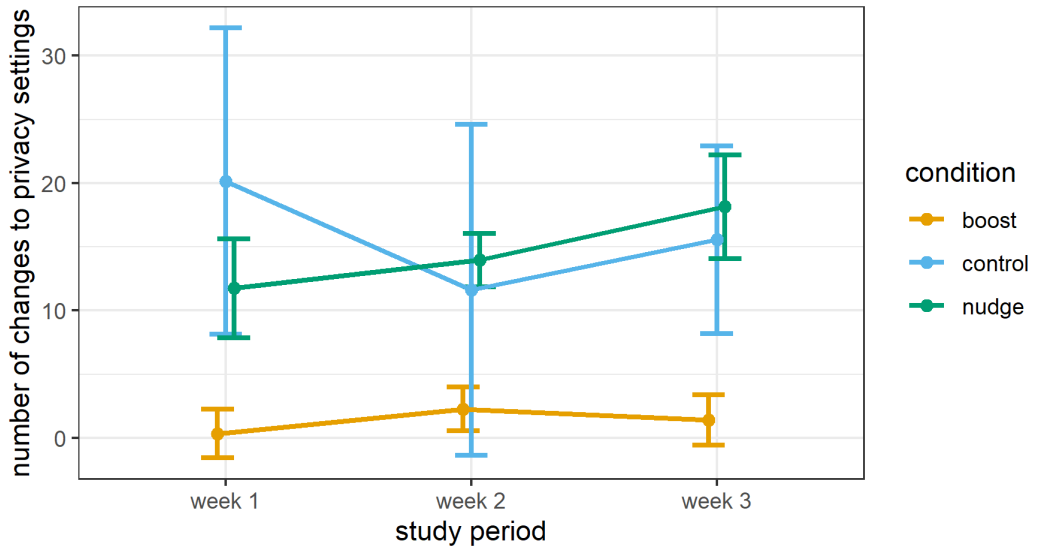


Figure 4.13.: Average number of changes to privacy related settings per week during the study phase, with 95% confidence intervals

Figure 4.13 depicts the mean numbers of setting changes, with all kinds of setting changes included. In general, the patterns between the three groups differ. The participants in the control group start with the highest number of setting changes in the first phase of the study, then change their settings less during the intervention period, and finally their number of changes rises again in the third week of the study. The confidence intervals for the means in this group are the widest, showing that there was more variation among this group than for the other two. Participants in the nudge group started with a lower number of setting changes, which became larger over each week of the study. Finally, participants in the boost group had the lowest average number of setting changes in all three weeks of the study. Their lowest average number of setting changes was during the first week. For the second week, this number was slightly larger, before falling again in the third week, although not to the level of the first week.

Since boosts in this study do not include general statements about setting changes, but refer to two specific settings, these are examined more closely. Figure 4.14 shows the number of changes to cookie settings for each of the three groups of participants during the three weeks of the study. Throughout all three weeks, participants in the control group changed their cookie settings the least. Participants in the boost

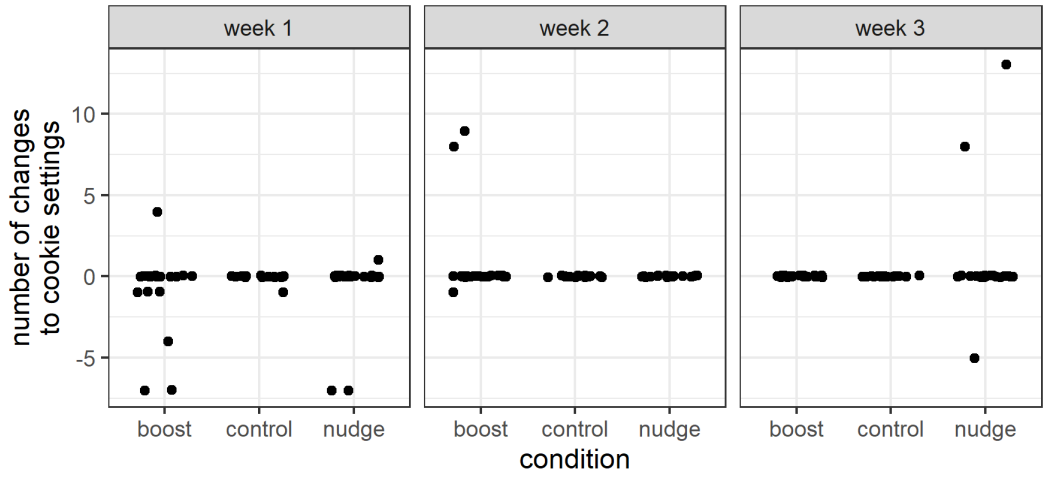


Figure 4.14.: Number of changes to cookie settings. Positive values show changes towards more privacy, negative values show changes towards less privacy. There may be two data points per participant in the graphic, if a participant implemented changes in both directions

group changed them more, especially during the first and second week. However, many of these changes were in a negative direction. In the nudge group, changes occurred only during the first and third week, and in both directions. To conclude, the participants in the boost group did not exhibit a higher number of changes of the cookies settings towards more privacy during and after the intervention period than the other two groups.

The changes to and from the private browsing mode are plotted in a similar way and are presented in Figure 4.15. Some participants exhibited a very high number of switches to and from private browsing mode, although most did not. To be able to visualize both, the y-axis was transformed using the log modulus transformation (John & Draper, 1980), whereby $L(y) = sign(y) \times \log(|y| + 1)$. Sign in this context means that the sign of y before the transformation is extracted and added again, to preserve it. It was adjusted slightly for this visualization in that pre-transformation values of 1 and -1 were adjusted to 0.3 and -0.3. Otherwise, the logarithmic part of the transformation would have transformed both these values to 0, and they would have become indistinguishable. |0.3| was chosen because it was smaller than $\log(2)$. While this adjustment is not mathematically justified, it helps to visualize the situation adequately.

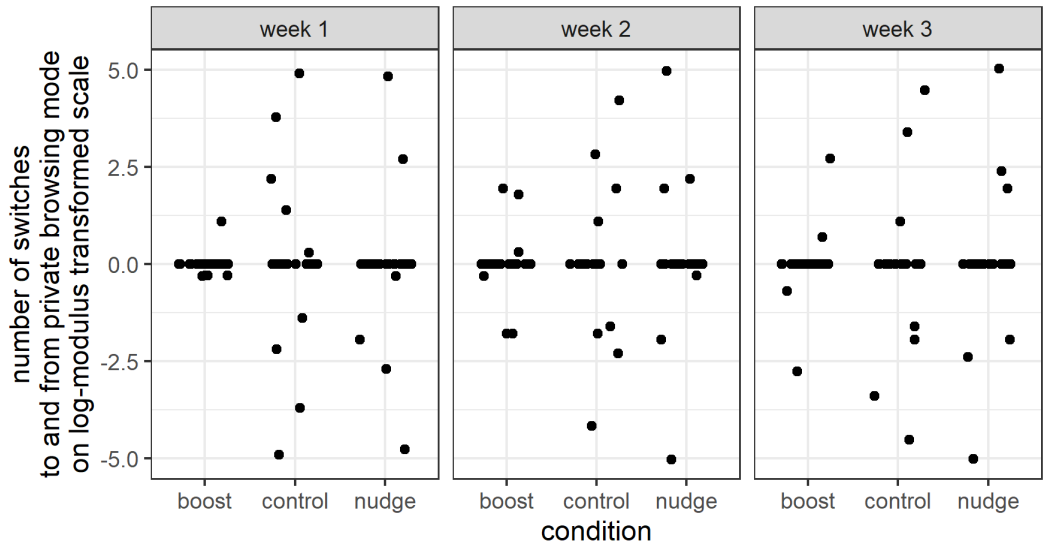


Figure 4.15.: Number of switches to and from the private browsing mode, positive values count the changes to private browsing mode, negative values count the changes away from private browsing mode. Multiple points in the graphic may be from the same participant

Again, an effect of the boosts is not evident from this visualization. The number of changes to private browsing mode only marginally increased during the second and third week of the study, and there were more participants changing to and from private browsing mode in the other two conditions.

It was not possible to directly monitor the use of adblockers during the study, but the participants were questioned on their use of adblockers during the final survey. In the boost condition, three people (14 %) installed an adblocker during the study, in the control condition, two people (10 %) did and in the nudge condition, one person (5 %) did. All three of those in the boost condition actually saw the boost concerning adblockers, for a total duration of 12 seconds, 69 seconds, and 11 seconds each. These new users of adblockers were included in the total number of participants using adblockers, which were distributed as follows. In the boost condition, eight people (38 %) used adblockers, four people (20%) in the control condition did, and there were five adblock users (24 %) in the nudge condition. So, proportionally, more participants in the boost condition installed and used adblockers than those in the other conditions. However, since it is not possible to know when the addblocker was installed, this cannot clearly be attributed to the effect of the boost.

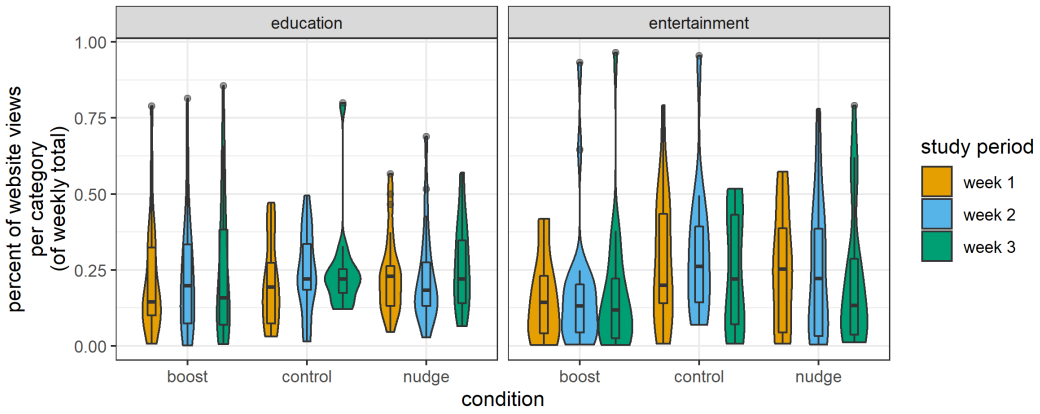


Figure 4.16.: Percentage of website visits for categories mentioned in boosts, by study phase and condition

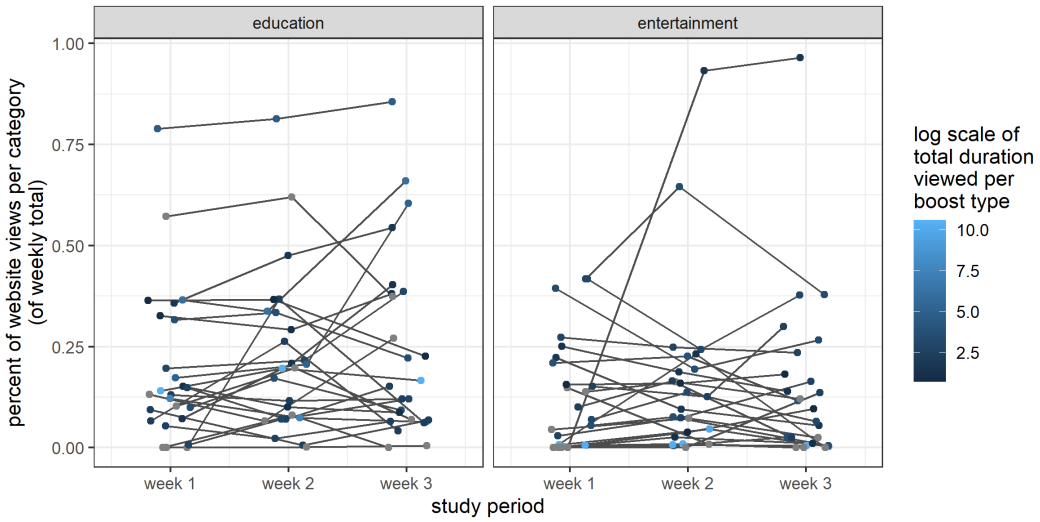


Figure 4.17.: Evolution of percentage of website visits for categories mentioned in boosts per participant, by study phase and condition

The final two boosts contained information about the privacy invasiveness of certain types of websites. More specifically, entertainment websites were revealed to be more privacy invasive than other websites, and education websites less. The distributions of the percentages of visits to these types of websites in the three phases of the study are displayed in Figure 4.16. These same distributions are also plotted for the other two conditions and these serve as a sort of control. However, there does not seem to be a difference between the conditions and between the weeks. This graphic does not visualize within-participant changes.

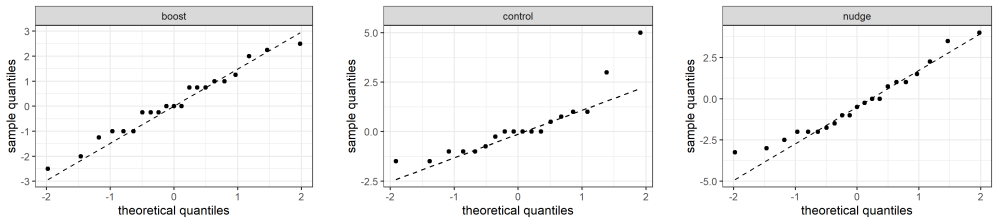
An attempt to include this aspect can be seen in Figure 4.17. The data points be-

longing to a single participant are connected by lines and filled according to the duration that the boost corresponding to each website type was seen by this participant. When a participant did not see the relevant boost at all, their points are colored gray. Because the distribution of this duration had a few outliers, it was plotted at a log scale, so colors for lower durations would also be distinguishable. If participants used the information from the boosts to make their behavior more private, one would have expected a rise in visits to education sites, and a decline in visits to entertainment websites during the second week. Since theoretically, the effect of boost is supposed to linger even after the boost itself is not present anymore, it would be expected for the percentage to stay higher and lower respectively during the third week. Although the graphic is quite cluttered, for some participants, this seems to be the case. For others however, patterns are different, so the boosts did not seem to have a pronounced influence on the the kinds of website visited.

4.7. Change in Privacy Knowledge

Another question is whether the boosts had an influence on the knowledge that was meant to be conveyed by them. In the pre-study described in Subsection 3.2.4, there had been significant differences in knowledge between those participants who were shown boosts and those who were not. However, this was tested right after they were shown the information, so it could be only a short-term effect. In the following analysis, participants from the boost condition, who were exposed to boost information during the second week of the study, are compared to participants from the other two conditions, who were not. It has to be noted, that one participant in the boost condition forgot to fill out the questionnaire on boost related knowledge at the beginning of the study, probably, since it was the last one. Unfortunately, this was only noticed after the end of the study.

A one-way ANOVA was conducted using the difference between pre-study and post-study boost knowledge as the dependent variable and CONDITION as the independent variable. Boost knowledge was calculated as the sum of the correct questions, where there was only yes or no as an answer possibility. In the case of ques-



(a) Boost condition (b) Control condition (c) Nudge condition

Figure 4.18.: Normal Q-Q plots for the difference between pre- and post-study boost knowledge for the different conditions

tions asking for how many third parties were thought to be active on a website under the conditions described, points were allocated depending on how far participants were from the correct answer. A difference of 1 meant they were awarded 0.5 points, a difference of 2 from the correct answer was rewarded with 0.25 points, and anything else with 0 points. The difference between boost knowledge measured by subtracting the pre-study boost knowledge from the boost knowledge measured post-study.

ANOVA assumes homogeneity of variance, normality of data within groups, and independence of observations. By using the difference between pre- and post-study boost knowledge as the dependent variable, each observation comes from a different participant, and thus independence is the case. The assumption of homogeneity of variance was tested using a Levene's test, which showed that there was a nonsignificant difference in group variances $F(57, 2) = 1.96$, $p = .15$, so this assumption is considered valid. Normality was tested using Shapiro Wilkes tests for each of the three conditions. According to these tests, data within the boost ($W = 0.98, p = .84$) and nudge ($W = 0.95, p = .39$) groups can be considered to be normally distributed, but data from the control group cannot ($W = 0.82, p = .003$). Additionally, normal Q-Q plots were examined. They are presented in Figure 4.18. This confirms the inferential statistics obtained from the Shapiro Wilkes tests in that, for the boost and nudge conditions, the points follow a straight line and so they can be considered normal. The graph for the control condition seems reasonably normal, but it also shows points at the upper end deviating from the line. This could be an indicator for a heavy right tail. ANOVA is robust in the face of non-normality

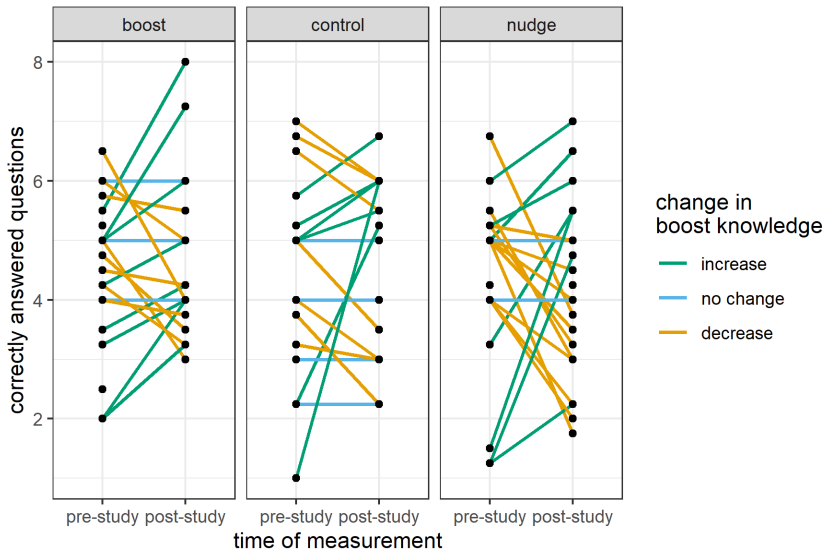


Figure 4.19.: Within participant change in boost knowledge, by condition

when group sizes are equal (Field et al., 2012), which is not the case here. Thus F could be biased for these data, and the following report should be taken with a grain of salt.

The effect of CONDITION on the difference between pre-study and post-study boost knowledge was not significant, F(2, 57)=3.53, p=.54. The η^2 effect size, which is often reported with ANOVA is considered to be slightly biased, since it is only based on sums of squares from the sample (Field et al., 2012); so another effect size was utilized instead. ω^2 is calculated as $\omega^2 = \frac{SS_M - df_M \times MS_R}{SS_T + MS_R}$, where SS_M are the model sums of squares, SS_T are the total sums of squares, MS_R is the residual mean squared error, and df_M are the model degrees of freedom. For this analysis, ω^2 was -0.01, which can be interpreted as being zero.

To account for the slightly non-normal data in the control group, an additional robust ANOVA was conducted to confirm this finding. The *t1waybt*-function from the *WRS* package (Wilcox & Schönbrodt, 2014) was used to conduct a robust ANOVA with 20% trimmed means and 2000 bootstrapped samples. This analysis confirmed that there was no significant difference between conditions concerning boost knowledge difference, $F = 0.75$, $p = .47$.

This analysis took into account by how much the boost knowledge changed, but to answer the research question, the main interest is whether there was an increase

knowledge change	variable	condition		
		boost	nudge	control
increase	actual number	9	7	6
	expected number	7.971	7.971	6.058
no change	actual number	7	6	6
	expected number	6.884	6.884	5.232
decrease	actual number	9	12	7
	expected number	10.145	10.145	7.710

Table 4.5.: Actual and expected values for χ^2-test with condition and knowledge change

or decrease in boost knowledge in the boost condition. The dependent variable was recoded to a nominal scale with differences in knowledge larger than 0 coded as "increase", values equal to zero coded as "no change" and values smaller than zero coded as "decrease". These data are visualized in Figure 4.19. There does not seem to be a difference in number of participants whose knowledge increased between the boost condition and the other conditions. Inferential statistics also showed that there was no significant difference between the conditions regarding the change in boost knowledge $\chi^2(4) = 1.01$, $p = .91$. Table 4.5 shows the actual and expected values for each of the conditions.

None of the expected values are below 5, so the χ^2-test was appropriate. This table also shows that even though the difference between the conditions was not significant, in the boost condition, the actual number of participants whose knowledge increased was above the expected values, and the actual number of participants whose knowledge decreased was below the expected values. For the nudge condition, it was the other way around and, for the control condition, increase and decrease of knowledge was closer to the expected values than for the other two conditions. To summarize, boost-related privacy knowledge increased slightly more often among participants in the boost condition than among those in the other conditions, but not significantly so.

4.8. The Effect of Condition and study phase on Browsing Privacy

To answer the question whether boosts and nudges change participants' behavior, the effect of CONDITION (nudge, boost or control) and STUDY PHASE (pre-intervention, intervention, post-intervention) on browsing privacy were examined. As discussed above (see Section 3.1 and Subsection 3.4.4), two variables were used as proxies for browsing privacy: the number of third party requests and the change in cookies. In the most extreme case, a participant who does not use the internet at all will achieve maximum browsing privacy, since that person will not encounter any third party requests or cookies, while even a privacy conscious participant will encounter some third party requests or cookies and thus achieve lower browsing privacy, simply because they visit more websites. To counter this effect and to be able to compare participants, the variables representing browsing privacy were averaged over the number of website visits for each day of the study separately.

4.8.1. The Multilevel Linear Modeling Approach

A linear mixed-effects model maximizing the log-likelihood was used to assess the effect of CONDITION and STUDY PHASE on the average amount of third party requests encountered on a website per day. While a repeated-measures two-way ANOVA is also a possible method of analysis for two categorical independent variables, it is not suitable for unbalanced designs (Field et al., 2012). Since the number of participants is not the same for all three conditions, and participants used the internet for different numbers of days during each study phase, the data can be considered unbalanced.

To assess the need for a multilevel model, a baseline model with only a fixed intercept was compared to a model with intercepts varying across participants, with values for week nested within participants. This represents the repeated measures nature of the data. There was significant variance in intercepts across participants, $\chi^2(2) = 61.42$, $p < .0001$. Parameter estimates for this model, which includes only the random intercept, are in Table 4.6. The standard deviation of the random intercept for participants was 26.6, 95% CI:(17.6, 40.1), while for weeks nested in partici-

model	predictor	b	SE b	df	CI		t	p
					lower	upper		
random intercept	intercept	25.5	4.8	1007	16.1	34.9	5.31	<.001
random intercept	intercept	29.7	12.9	1007	4.50	54.8	2.30	.02
and fixed effects	boost (compared to control)	11.8	17.2	66	-22.4	46.1	0.69	.49
	nudge (compared to control)	-5.31	17.0	66	-39.1	28.4	-0.31	.75
	intervention (compared to pre)	-8.72	16.0	117	-40.4	22.9	-0.54	.59
	post (compared to pre)	-3.42	16.2	117	-35.3	28.5	-0.21	.83
	boost : intervention	-18.7	21.0	117	-60.2	22.8	-0.89	.38
	nudge : intervention	12.3	21.8	117	-30.7	55.3	0.56	.57
	boost : post	-24.5	21.1	117	-66.1	17.1	-1.16	.25
	nudge : post	13.2	21.1	117	-28.4	54.8	0.62	.53

Table 4.6.: Baseline and final model parameter with average number of third party requests per website per day as dependant variable

pants, the standard deviation of the intercept was 30.9, 95% CI:(22.9, 41.8), and the residual standard deviation was 87.7, 95% CI:(84.0, 91.7).

Condition, study phase and their interaction were added subsequently as fixed parameters to the model in this order. Contrasts were specified for condition to have separate comparisons for both boost and nudge to control as the baseline category. Similarly, for study phase, both the intervention period and the post-intervention period were compared to the pre-intervention period, which served as a control. The fixed parameters did not significantly improve the model, $\chi^2(8) = 6.47$, $p = .59$. Parameter information for this model is also in Table 4.6. Concerning the random effects, the standard deviation for participants in this model was 27.3, 95% CI:(18.5, 40.1). For weeks nested in participants, the standard deviation was 28.9, 95% CI:(20.8, 40.1), and the residual standard deviation was 87.8, 95% CI:(84.0, 91.7).

Random slopes were evaluated for condition and study phase separately. The slopes varied significantly across conditions $\chi^2(10) = 30.67$, $p = .0007$, however, confidence intervals could not be calculated for the random effects, because the approximate covariance matrix for the maximum likelihood estimations could not be obtained. A possible reason for this is overparametrization in the model.

There was also significant variation of slopes across different phases of the study $\chi^2(10) = 40.90$, $p < .0001$. Again, confidence intervals for random effects could not be calculated for the same reason as when varying slopes across condition.

The AIC and Bayesian information criterion (BIC) were both lower when slopes varied across STUDY PHASE (AIC = 14277, BIC=14389) than when slopes varied across CONDITION (AIC=14287, BIC=14399). However, since overparametrization seemed to be a problem for both models, the model with random intercept and CONDITION, STUDY PHASE, and their interaction as fixed effects was accepted as the final model.

For sake of completeness, two more models were also fitted with an additional fixed effect each. The first one included the TOTAL DURATION OF DISPLAYED INTERVENTION MODALS, but there was not significant variation along this new variable $\chi^2(1) = 0.49$, $p = .49$. The second one included the TOTAL DURATION SPENT ON THE USER INTERFACE OF THE EXTENSION DURING THE INTERVENTION WEEK. However, this variable did not vary significantly either, $\chi^2(1) = 0.14$, $p = .71$.

To check for outliers and cases with too much influence on the parameters of the final model, case-wise diagnostics were conducted. Cases with standardized residuals larger than |2| are expected to make up 5% or less of cases in an ordinary sample (Field et al., 2012). In this sample, where n=1199 , 5% are around 60 cases. There are 25 cases where the standardized residuals are larger than |2|, which is below the threshold of 5%, but for 18 of those cases, the standardized residuals are larger than |2.5|. This is above the threshold of 1%, which would have been around 12 cases. Of those 18 cases, the maximum absolute value of the standardized residuals is 22.2 and the median absolute value is 3.4, which is quite concerning.

Looking at the participant ids for the 25 cases, some appear multiple times. In most cases, participant ids are not in this subset more than twice, but there are five cases from the participant with the id 133, and seven cases from the participant with the id 40. Consequently, their demographical data was examined to see whether any unusual traits justified removing those two participants from the analysis. Both these participants were students using Chrome on a Windows machine, who spoke German as a native language. They both reported problems during the study, but

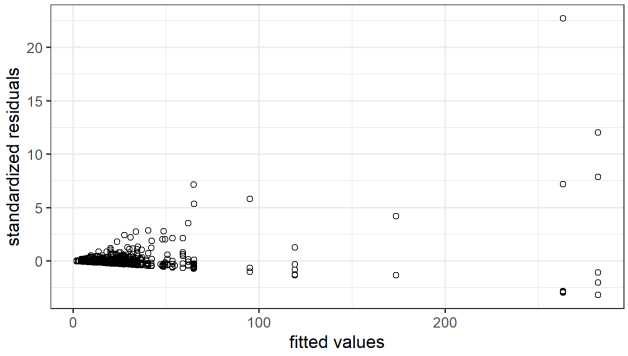


Figure 4.20.: Plot of standardized residuals against outcome (number of third party requests), used to check for linearity

only that their browser was slower and that ventilation was louder. This is not a sufficient reason to exclude their data from the analysis, since these problems were repeatedly reported, especially by Chrome users.

Cook's distance was calculated using the *influence* and *cooks.distance*-functions from the *car* package. For the data in question, no cases yield a Cook's distance larger than one. Other case-wise diagnostics exist, for example hat values or covariance ratios (Field et al., 2012). However, this functionality was not implemented in either the *car* package and the *HLMdiag* package, which was additionally used, for models with more than one level of nesting, such as the one described here. From the case-wise diagnostics which were obtained, one gave rise to concern, and the other did not.

To judge whether the model can be generalized to the population, assumptions were checked. The assumptions of linearity and normality of residuals were evaluated graphically. To inspect linearity, the residuals were plotted against the outcome variable. Figure 4.20 shows a non-random pattern, which means that linearity cannot be assumed. Normality of residuals was examined using a normal Q-Q plot, depicted in Figure 4.21. Since the points strongly diverge from the diagonal, normality of residuals is rejected. Instead, there seems to be a peak of values very close to 0, reflected in the horizontal succession of points at this value of the y-axis. Finally, homogeneity of variance of residuals was examined using a Levene Test.

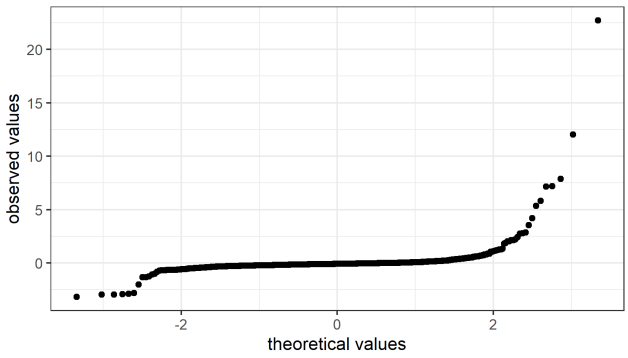


Figure 4.21.: Normal Q-Q plot used to assess normality of residuals for model predicting average number of third party requests

This assumption holds, $F = (1, 1197) = 0.008$, $p = .93$. Due to the violation of assumptions, the model can likely not be generalized to a larger population.

The second variable used as a proxy for browsing privacy in this study was the change in number of cookies per website. As with the number of third party requests, a linear mixed-effects model maximizing the log-likelihood was used to examine the effect of CONDITION and STUDY PHASE on the outcome variable. Since the intercepts varied significantly across participants and weeks nested within participants, $\chi^2(2) = 895$, $p < .0001$, a multilevel model was deemed sensible. The parameters of this baseline model are in Table 4.7. For the first-level random effects, the participants, the standard deviation was 0.03, while for weeks nested in participants, the standard deviation of the intercept was 72.2 and the residual standard deviation was 16.4. It was not possible to obtain confidence intervals for the random effects of this model due to non-positive definite approximate variance-covariance.

Again, CONDITION, STUDY PHASE, and their interaction were entered into the model subsequently, but none of these fixed effects significantly improved the model, $\chi^2(8) = 10.4, p = .24$. The parameter estimates for the model with both of the predictors and their interaction are also in Table 4.7. It shows that there was only one significant effect: There was a significantly higher number of cookie changes in the nudge condition than in the control condition. For the concrete test statistics, consider Table 4.7. While p is below the criterion for significance in this case, it is not by much. Similarly, the 95% confidence intervals for this variable do not cross zero, but they are

model	predictor	b	SE b	df	CI		t	p
					lower	upper		
random intercept	intercept	6.20	5.24	1007	-4.08	16.2	1.18	.24
random intercept	intercept	0.69	16.3	1007	-31.2	32.6	0.04	.97
and fixed effects	boost (compared to control)	3.29	22.0	66	-40.5	47.1	0.15	.88
	nudge (compared to control)	48.2	21.6	66	5.2	91.2	2.23	.03
	intervention (compared to pre)	-1.64	24.1	117	-49.2	45.9	-0.07	.95
	post (compared to pre)	-1.79	24.1	117	-49.3	45.7	-0.07	.94
	boost : intervention	-2.80	31.6	117	-65.3	59.6	-0.09	.93
	nudge : intervention	-48.0	31.6	117	-110.4	14.4	-1.52	.13
	boost : post	-2.67	31.7	117	-65.3	60.0	-0.08	.83
	nudge : post	-50.2	31.9	117	-113.2	12.8	-1.57	.12

Table 4.7.: Baseline and final model parameters with average number of cookie changes per website per day as the dependant variable

nevertheless relatively wide and multiple testing has to be kept in mind. During this analysis, multiple models were fitted with different dependent and independent variables. This can be considered multiple testing and thus, the probability for type I error may be inflated. As such, this significant result should be taken with a grain of salt.

The standard deviation of the random intercept for participants was 0.07, while for weeks nested in participants, the standard deviation of the intercept was 70.2, and the residual standard deviation was 16.4. Note that due to the reason given above, confidence intervals were not available for random effects.

An attempt was made to add random slopes to this model to see if this could improve the fit, but regardless of whether slopes were allowed to vary over CONDITION, or over STUDY PHASE, the models did not converge and errors were issued, which warned of a singular precision matrix. This could again be a problem concerning overparametrization, meaning that the model is overly complex and there is not enough data in each cell to compute random slopes.

As a consequence, the model with random intercepts and fixed predictors was considered the final model. Again, two further models were fitted for sake of completeness, which examined the influence of encountering interventions during the study. The variables used were the same as for the model with the average number of third party requests described earlier. Neither the TOTAL DURATION OF INTERVEN-

TIONS SEEN ($\chi^2(1) = 0.11$, $p = .74$), nor the TOTAL DURATION OF INTERACTING WITH THE USER INTERFACE OF THE EXTENSION ($\chi^2(1) = 0.001$, $p = .97$) are significant.

Case-wise diagnostics were calculated for this model to check for outliers or cases with too much influence on the model. Since the sample size was the same for this model, not more than 60 cases were expected to have standardized residuals larger than |2|, and there were only 28 cases of this in the data. However, there were 17 cases of standardized residuals above |2.5|, which is more than 1% of total cases. Of those, the maximum absolute value of the standardized residuals was 17.0 and the median absolute value was 4.4, which gives rise to concern. Again, there were two participants (P 26 and P 127) who appear more than two times among those cases where the standardized residual is larger than |2|. Their demographical traits were examined for unusual occurrences. Both were students who spoke German as a native language and used a Windows machine during the study, one with Firefox and one with Chrome as a browser. Neither encountered any problems during the study. Consequently, they cannot be excluded from the analysis on these grounds.

Cook's distance was calculated like above and there were no cases with values larger than one. However, for 696 cases, which is around 58% of the cases, Cook's distance could not be calculated. For the cases with standardized residuals larger than |2| this percentage is even higher, at 74%. While it is not clear what the cause of this failure is, since no error messages were present, it nevertheless does not bode well for the fit of the model.

This model also failed to generalize to the population, because the assumptions of normality and homogeneity of residuals and linearity did not hold. Figure 4.22 depicts a distinctly non-random pattern for the residuals, where most of the data is concentrated around 0, but the larger the absolute fitted values become, the more variance there is. Figure 4.23 shows a non-normal pattern, whereby the points strongly diverge from the diagonal. Again there is a peak of observed values close to zero. Finally, homogeneity of variance of residuals was tested using a Levene's Test and has to be rejected $F(1, 1197) = 5.26$, $p = .02$.

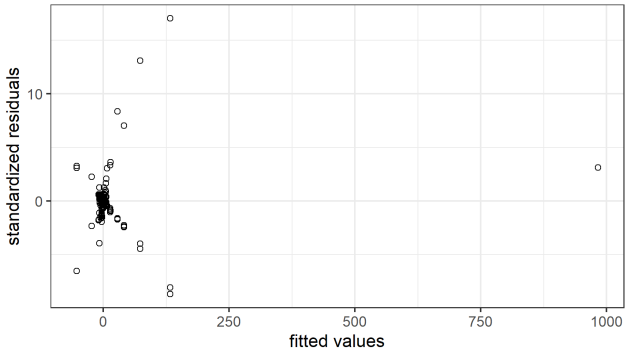


Figure 4.22.: Plot of standardized residuals against average cookie change, used to check for linearity

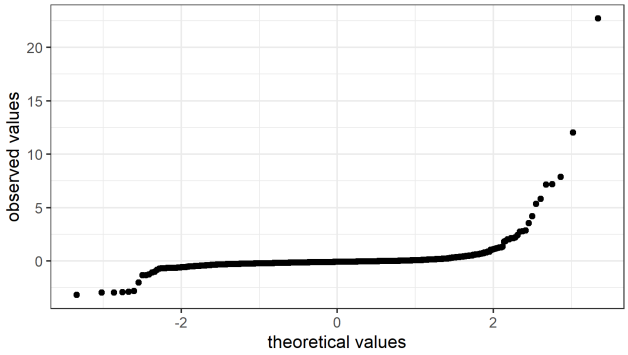


Figure 4.23.: Normal Q-Q plot used to assess normality of residuals for model predicting average cookie change

Dependent variable	Effect	F	df	p
average amount of third party requests	condition	0.18	2	.83
	study phase	12.77	2	<.001
	condition:study phase	0.38	4	.82
average cookie change	condition	45.5	2	<.001
	study phase	228.8	2	<.001
	condition:study phase	16.2	4	<.001

Table 4.8.: Results of the factorial ART for average amount of third party requests and average cookie change

4.8.2. The Non-Parametric Approach

Since the distributional assumptions did not hold for the multilevel modeling approach described above, and since the two predictor variables of the most interest, CONDITION, and STUDY PHASE, were categorical variables, an ART procedure was conducted as an alternative. It can be used as a non-parametric alternative to repeated-measures ANOVA. This procedure is implemented in the *ARTool* package (Wobbrock et al., 2011). The results of this analysis for both the dependent variables investigated above are in Table 4.8.

Post-hoc tests using the Tukey method to adjust p-values for multiple testing were conducted to further analyze the highly significant main effects of CONDITION and STUDY PHASE on the average amount of third party requests. For the main effect of STUDY PHASE, results were averaged over the levels of CONDITION. According to these, the average amount of third party requests in the first week differed significantly from those in the second week $t(120) = 3.74$, $p = .001$ and those in the third week $t(104) = 4.807$, $p < .001$. However, there was no significant difference between the second and the third week with respect to the average amount of third party requests, $t(106) = 0.94$, $p = .62$. To visualize these results, the means per CONDITION and STUDY PHASE were graphed along with error bars reflecting 95% confidence intervals, in Figure 4.24.

While both significant main effects and interaction were reported for the analysis using the average change in the number of cookies as a dependent variable, it should nevertheless be noted that there were several accompanying warning mes-

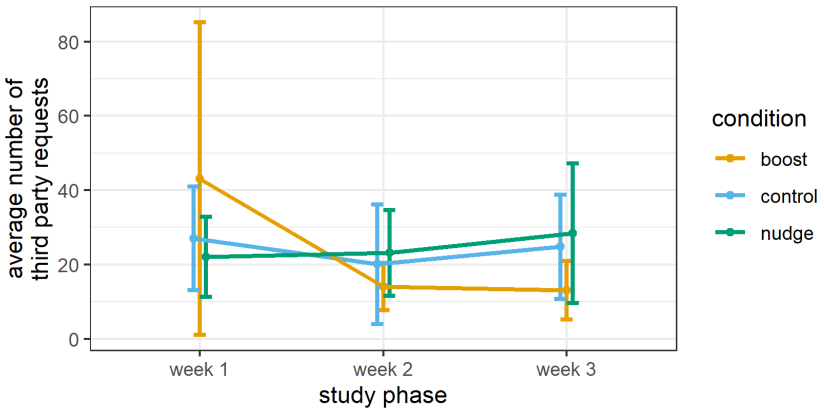


Figure 4.24.: Effect of condition and study phase for average number of third party requests, unnormed means with 95% confidence intervals

contrast	df	χ^2	p
control-boost : week 1-week 2	1	2.9	0.36
control-nudge : week 1-week 2	1	41.1	<.001
boost-nudge : week 1-week 2	1	26.4	<.001
control-boost : week 1-week 3	1	4.1867	0.20
control-nudge : week 1-week 3	1	45.0	<.001
boost-nudge : week 1-week 3	1	25.9	<.001
control-boost : week 2-week 3	1	0.1102	1.0
control-nudge : week 2-week 3	1	0.0092	1.0
boost-nudge : week 2-week 3	1	0.2103	1.0

Table 4.9.: Results of interaction contrasts for the value of average change in cookies

sages. The model did not converge, even though output was reported, and there were warnings that the model was a singular fit. This may be the cause of a random effect variance close to zero, which could be the case given that the variance in this data was generally low. Consequently, any further investigation should take this into consideration. Significant main effects cannot be interpreted in the presence of a significant interaction such as in this case, and so the *testInteractions*-function from the *phia* package (De Rosario-Martinez, 2015) was used to perform interaction contrasts, which analyze differences of differences. These are reported in Table 4.9. Any p-values reported are corrected for multiple testing using the Holm method. Again, it should be noted that there was a warning that the model failed to converge. The difference between the control and nudge condition was significantly different in week 1 and week 2 of the study, and also in week 1 and week 3 of the study. Like-

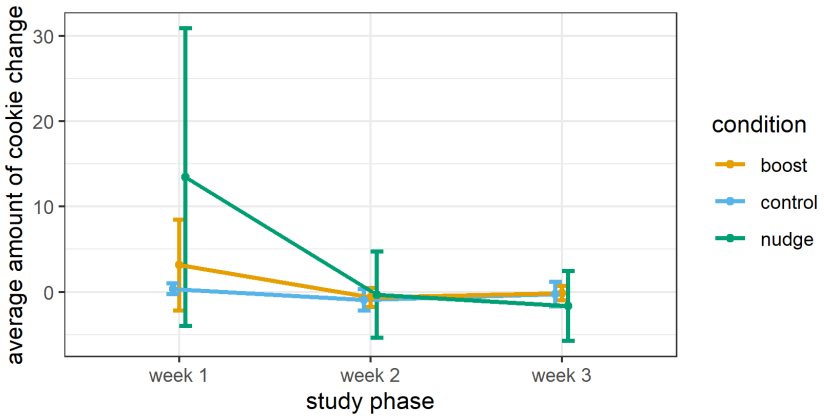


Figure 4.25.: Effect of condition and study phase for average change in cookies, unnormed means with 95% confidence intervals

wise, there was a significant difference between the differences between the boost and nudge conditions in week 1 and week 2 of the study, as well as in week 1 and week 3 of the study. Since interpretation of these differences of differences is not very straightforward, the interaction was additionally visualized in Figure 4.25.

5. Results

Some preliminary investigations were conducted, which do not directly relate to the two main research questions sought to be answered by this thesis, but which are nevertheless interesting. During the first week of the study, there was a ruling which prohibited online businesses from pre-selecting privacy invasive choices on cookie notices. A visual examination of the proxies for browsing privacy before and after this ruling did not show conclusive evidence that this improved privacy for the participants in this study, although perhaps there was just too little data before the ruling to judge that. Due to this reason, the ruling was not taken into consideration for further analyses.

Boosts in this study contained information from the literature, for example on the privacy invasiveness of certain types of websites. The data from this study supports the claims from the literature that news websites and entertainment websites have more requests to third parties and more cookies than education websites. While the exact rank concerning privacy invasiveness was not reproduced; a tendency similar to that in previous work is visible in the data. It has to be considered, that the categorization procedure and the smaller sample size in this study may be responsible for the difference.

Finally, using self-reported measures for privacy concern and privacy knowledge, as well as actions taken by participants to protect privacy, a Poisson regression approach was taken to study whether the privacy paradox is apparent in this study. Privacy concern and privacy knowledge both, as well as their interaction had a significant effect on the number of privacy protective actions undertaken by participants. This does not support the notion of a privacy paradox. Much research in the domain of privacy examines behavioral intentions instead of actual behavior

(Lowry et al., 2017), and, especially concerning the privacy paradox, this approach may not be valid (Kokolakis, 2017).

In the analysis above, actual behaviors were analyzed, but these behaviors were self-reported by the participants. So two multiple regression models were fitted using the same predictors as before, but browsing privacy was used as the dependent variable instead. Values were calculated for each participant to be used in the analysis, and were averaged over the first week, which served as a control phase in the study. In these two models, neither PRIVACY CONCERN, nor PRIVACY KNOWLEDGE, or their interaction were significant predictors, which would lead to a conclusion juxtaposing the one made on the basis of the self-reported privacy behavior.

However, it can be argued, that browsing privacy is not equivalent to privacy behavior. The latter relates to users actively making decisions and taking actions to protect their privacy, while the former is not dependent on actions taken to protect privacy, but also on other factors. Among these are whether the actions of a participant are effective to protect browsing privacy, especially with respect to the two metrics used to gauge browsing privacy, or how and where a participant spends their time on the internet. The exposure to features of websites, which are detrimental to privacy, such as cookies, or requests to third parties, can vary widely, for example depending on the type of websites visited (Englehardt & Narayanan, 2016; Urban et al., 2019). So instead, it can be concluded, that participants' knowledge about privacy, and their privacy concern were not able to significantly predict their achieved browsing privacy.

The first main research question of this thesis was whether boosts and nudges changed users' behavior to preserve privacy more. This was evaluated in a two-step process. First, an examination was conducted as to whether behaviors changed which were related directly to the content of boosts. This was not possible for nudges, since these did not offer specific suggestions for behavior change, but rather compared participants to their peers in the study. There was no obvious answer to the question whether users adopted the behavior which was named as beneficial in boosts. The cookie permission was not restricted more often than by participants

in other conditions, and similarly, private browsing was not switched on more often than in other conditions. The average amount of website visits to sites belonging to the categories mentioned in the boosts also did not change much, and when looking at within-participant change, there was no obvious pattern either. Some participants had higher percentages of website visits on education sites later in the study, and lower percentages on entertainment sites, but the opposite exists, too; so this is likely due to random variation. There was a slightly higher percentage of participants in the boost condition, who installed and who used adblockers during the study. However, it is not clear whether this was due to the boost, because it was not possible to track when participants installed such an extension, so this could also have happened before they were exposed to the boost. There were also some installments of adblocking extensions in the nudge and control groups. Just thinking about privacy may elicit privacy concerns and lead to participants taking actions (Marreiros et al., 2017). This was provoked through the informed consent stating study aims and through the questionnaires related to behaviors and knowledge surrounding privacy, to which the participants were exposed at the beginning of the study.

The first research question was not only whether participants' behavior changed under the influence of boosts and nudges, but also whether their behavior changed to preserve their privacy more during browsing. Two proxies were used as a measure for browsing privacy: The average amount of changes of cookies per website and the average number of requests to third parties per website. For both these dependent variables, a lower value means a higher level of privacy. Separate multilevel linear models were fit using these dependent variables, both including random intercepts for participants and weeks nested in participants, to account for the repeated-measures nature of the data. CONDITION, STUDY PHASE, and their interaction were included as independent variables in these models. Both models were a significant improvement over baseline models without predictors or random effects. For the number of third party requests, none of the predictors were significant. However, for the amount of changes in number of cookies, the nudge condition differed

significantly from the control condition. All other predictors in this model were not significant. In general, the introduction of random effects lead to the main improvement of the model over a baseline model. It can be concluded that the main variation in the data occurred within participants and the weeks nested within participants. The models' fit to the data was not clear. While some case-wise diagnostics were reassuring, others caused concern. Especially worrying was that there was a large number of cases in the model predicting the number of cookie changes, where Cook's distance could not be calculated. Since the assumptions for both models did not hold, both these models fail to generalize to a larger population.

ART, a non-parametric alternative to ANOVA was used as a different approach to circumvent the distributional assumptions in the multilevel modeling approach. Again, this analysis was repeated once for each of the dependent variables representing browsing privacy. There was a significant main effect for STUDY PHASE on the number of third party requests, which was followed up by post-hoc tests. They revealed a difference between the first and second, and first and third week of the study, but not between the second and third week. Graphing this relationship suggests that the average number of third party requests was higher in the first week of the study than in the second and third week. This plot also indicates, that for the boost condition, browsing privacy, as reflected by the number of third party requests was worse than for the other two conditions during the first week, and then improved in the second week, and stayed that way in the third week. However, there was no significant main effect of CONDITION. The development for the nudge and control conditions was similar and the browsing privacy in these conditions remained more or less constant.

For the change in number of cookies, both main effects, as well as the interaction between CONDITION and STUDY PHASE were significant. The interaction was investigated further with an analysis of differences of differences, which revealed that the differences between control and nudge conditions differed between the first and second, as well as the first and third week of the study. Additionally, the boost and nudge conditions differed between the first and second, and the first and third

week of the study. This pattern was somewhat similar to the pattern for the effect of study phase in the ART with the number of third party requests as the dependent variable, only that in the case of the number of changes to cookies, the participants in the nudge condition achieved worse browsing privacy in the first week, which then improved in the second week, and stayed this way in the third week.

The cause of this low level of browsing privacy for one of the conditions during the first week of the study is unclear. Browsing behavior, regarding the number of active internet users, as well as their average number of daily website visits, and the types of websites visited, was relatively similar across all three weeks of the study. However, for both these main analyses, confidence intervals were very wide, so there is a lot of uncertainty in the data. During the process of fitting models, both for multilevel linear models, and during the ART procedure, there were multiple instances, when models did not converge, or other issues, such as overparametrization, or singular fit, arose. This could be due to too little data, since not all the participants used their device to browse the internet on all days during the study. Additionally, only 68 participants started the study, and only 60 were still partaking at the end of the three weeks, which lead to further missing data.

As a conclusion, in the current study, there was only very weak evidence of behavior change because of nudges or boosts, and this behavior change did not lead to a significant improvement in browsing privacy. Interaction plots suggested that in the nudge group the average number of cookie changes decreased more from the first week to the second week, during the study period, than for the other conditions. In the boost condition, the same was true for the number of third party requests. These differences were not significant however. A high level of browsing privacy in this study was represented by a low number of third party requests and a low or negative number of added cookies.

The second research question in this thesis was whether being exposed to boosts changed users' knowledge about privacy. Since the boosts in this study contained very specific information, this information was tested, both at the beginning of the study, before any participants were exposed to the information, and afterwards,

when participants in the boost condition had seen the information, and participants in the other conditions had not. A one-way ANOVA with CONDITION as the independent variable, and the difference between pre-study and post-study knowledge as the dependent variable was conducted, but did not reveal a significant difference between the conditions. A Chi-Square-Test with a rescaled dependent variable did not report significant either, although there were slightly more people than expected in the boost condition whose knowledge had improved, and in the other two conditions this was not the case. This means that a week after the intervention and exposure to the boosts, knowledge was not significantly improved for the participants in the boost condition. During a pre-study, in a between-groups design, there was a significant difference in knowledge for most of the boosts used in the final study. An interpretation for this finding would be that contrary to what is claimed about boosts, their effect does not last after the intervention, or rather, since there was no significant change in behavior identified in this study, the knowledge conveyed in the boost is not retained. However, it is also possible, that this knowledge was simply never present in the participants. When using boosts in a naturalistic environment, it is not possible to control how participants interact with a boost, so it is entirely possible, that participants did not see all the boosts, or that they did not pay attention to them.

6. Discussion

Results from this study suggest, that the effect of boosts and nudges on browsing privacy is at most weak, but in general the results are not conclusive. Similarly, boosts did not significantly change participants' knowledge. However, the study was subject to some limitations. One is the sample of participants. Snowball sampling was used to try to acquire a more diverse sample, and succeeded partially, in that there were some older and non-student participants, but most of the participants were still students. This is likely because most of the acquaintances of the author were students, who forwarded the recruitment text more to friends than relatives, and these friends were also students. Another reason why non-students did not participate was that participation required them to install software, namely the study extension on a laptop device which they frequently use. Some adults who were contacted by the author during the recruitment phase declined participation because they were not allowed to install external software on their work laptop. Others mainly use mobile devices to access the internet, and were thus not eligible for participation. It can be concluded that the high amount of students in the sample is a result of their better fit to the criteria for participation than members of the working force. Earlier work comparing different populations' responses to SSL warnings suggested that students' responses may not necessarily differ from responses from a more diverse sample (Sotirakopoulos et al., 2011), however, they also reported that their broader sample was technically sophisticated, which may not be the case for a general population. Similarly, in this study, participants showed above average privacy knowledge, when compared to the overall German population, and also a tendency towards affinity for technology. As a consequence, conclusions drawn from this study may not generalize well to internet users with less privacy knowledge and less technical affinity.

Another limitation was partially caused by the remote execution of this study, but it is also common in longitudinal work. To achieve comparable results, all participants were expected to install the extension on the same day, and were reminded of this day per e-mail two days before. However, this did not work out as expected, and a sizable number of participants installed the extension some days later. A similar situation arose at the end of the study, when a number of participants did not complete their final questionnaires or uninstall the extension, until prompted multiple times. There were also some unexplained drop-outs, which is not unexpected for studies running over multiple weeks. One or two common installation sessions and uninstalling sessions at a set time and date might have resulted in higher compliance levels for the set start and end dates. On the other hand, some participants from different cities were able to take part due to the remote nature of the study.

Conducting a naturalistic study comes with advantages and disadvantages. Such work is valuable, because it has a high ecological validity. Participants were using their own devices, and, in most cases, the browser they were used to, over a period of multiple weeks. Laboratory studies have shown that boosts and nudges can influence behavior (e.g. Zimmerman et al., 2019b; van Bavel et al., 2019). In this study, the naturalistic nature of the data meant that a lot of noise was introduced because there was little control. There was no control over how long or how focused participants interacted with the boost and nudge interventions, there was no control over when, how often and how participants used their devices to browse the internet. While the participants were advised to use the device and browser with the study extension for their internet usage during the study, there were multiple questions concerning this during recruitment. Some potential participants had multiple devices, and asked if they could still participate. Others usually used a different browser than the two which were compatible with the study extension, but were willing to use a different browser than their normal one for the duration of the study. Recruiting a large enough number of participants for a longitudinal study was not easy, so these people were advised to try and use the device with the study extension as much as possible, and the final survey included questions on device

and browser usage. This showed that multiple device and browser usage was fairly common among participants.

As described above, there was also little control over when participants filled out the surveys at the beginning or end of the study. Since participants were manually recruited for this study, and required to install additional software themselves, the sample size was another limitation. Web browsing data was available for 69 participants, but not for all of the days in the study, and some participants only used the study extension for very few days. This lead to problems in the analyses, since for example, ANOVA is not robust in the face of missing data. Similarly in regression analyses, the estimate of R automatically becomes larger with the amount of predictors. Using condition, study phase, and their interaction as predictors means having 8 predictors, since categorical variables amount to multiple predictors. Often, warnings were issued that models were failing to converge, which could also be a result of two little data for each value of the variables used as predictors. So, as a consequence, the trends identified in laboratory studies are not visible in this work.

There were also issues concerning data collection. The first was that there were several problems which arose in conjunction with the study extension. The most severe of these was that in some cases, the extension uninstalled itself, for unknown reasons. This meant that when participants reinstalled the extension, they received a new participant id, and were possibly assigned to a new condition. It also meant that any behavioral summaries they received at the end of the study only included data collected with their newest id, and any data collected before an uninstall was excluded. Such summaries were not complete, and may have led to biased summaries for other participants. For example, if the extension of a participant with very low values in browsing privacy, as measured by privacy points, and the two proxies for browsing privacy, is repeatedly uninstalled, then there are multiple instances of these low values in the data. In comparison, other participants who achieve more browsing privacy than this person appear to be better than a larger percentage of participants than they actually are, since data for some of the participants actually

stem from the same person, but they are counted as different people. Of course, this could also occur the other way around. In any case, the summaries displayed in nudges and at the end of the study may have been biased. This problem could have been alleviated by allowing participants to enter their previous id, if they have any, on reinstalling the software, so that they become identified by their old id once more. The data saved in their local storage becomes lost on uninstalling of the extension, but this problem could be solved by storing that information on the backend, in the database, but encrypted, and only accessible for the participant, so that it, too, could be retrieved. Since the problem only became apparent after the start of the study, it was not possible to implement such a solution on short notice.

Another methodological issue concerned the questionnaire on knowledge conveyed by boosts. It consisted of those questions asked in the control condition of the crowd-sourcing pre study to explore boost comprehension, and was already adjusted according to suggestions from participants of this pre study. During the main study further comments were received, stating that the given options of "yes" and "no" were a problem for uncertain participants, who would have liked an option like "I don't know", to account for this situation. Since the same questionnaire was used at the beginning and the end of the study, it was not possible to change this questionnaire mid-study, so as to ensure consistency. However, this means that there is noise in the data, since participants who were unsure, guessed at the correct answer. This noise may have obscured the effect boosts had on privacy knowledge as conveyed by boosts.

In this study, browsing privacy was represented by two proxies, which have also been used in previous work (Mazel et al., 2019). There were some limitations concerning these proxies. A request was considered a third party request, if the hostname property of the target URL was not the same hostname property of the current tab. This heuristic may not lead to correct classification of requests as third or first party requests in all cases. Correctly identifying third parties is difficult, and it is hard to judge whether an URL still belongs together with the URL of the current open tab. More resources are necessary to accurately classify domains, but to

run such a classification process either on the participants' computers, or to send frequent requests to an API, would have likely lead to a further deterioration of performance. After considering these trade-offs, the heuristic described above was deemed the best solution.

There were also issues with the second proxy used, the change in cookies per website. First, contrary to what one might expect, this change was sometimes negative, which means that cookies were deleted. More insight into this deletion of cookies is needed, to find out whether cookies were deleted automatically on a site by site basis, or whether this was due to participants' browser settings, or their manual deletion of cookies. Another possible explanation for these deletions of cookies could be that cookies were first set, and then later retracted according to a participant's response to a cookie notice. Second, the change in cookies was calculated by comparing the current amount of cookies to the amount of cookies on the previous website visited by the user. For the very first website a participant visited during the study, there was no previous website, so all the cookies collected on the participant's device previous to participation in the study were assigned to this first website visited. This is not an accurate representation of cookies on this website, and overestimates the amount of cookies added on this website. Assigning this website zero cookies would just propagate the problem, as the amount of current cookies for the next website visited would then be compared to the zero for the first website, and thus the overestimation would happen for the second website. A solution for future work could be to ask participants to delete all their cookies prior to participation, although this may be cumbersome, since settings and other useful things may also be saved with cookies. Excluding the first instance of data from the analysis would remove the burden from the participant, but since aggregated data was used in the analyses in this study, and not all participants started the study on the same day, this tedious process was outside the scope of this thesis. Another problem with the proxy in its current form is that it obscures exactly how many cookies were added and deleted on a given website visits, since only the total number of cookies on that website is available. However a total change in cookies of two could mean that two

cookies were added and none were deleted, or it could mean that five cookies were deleted and seven were added. Generally, investigating the processes used by open source privacy protection or evaluation tools could be useful to get insight into how cookies are classified and identified.

A more general limitation of the proxies used, is that they are separate individual measures related to browsing privacy, and may only reflect certain aspects of this variable. Many different metrics have been used to approximate browsing privacy (Mazel et al., 2019), but since obtaining more metrics also costs more computational resources during the study, and an implementation workload outside the scope of this thesis, two widely used proxies were chosen for the study presented here. There has been work on using multiple automated procedures to calculate a privacy score for websites, which is based on four modules: tracking and privacy checks, website encryption, web security checks, and mail encryption (Maass et al., 2017). This tool has been used in analyses, for example of German university websites (Mueller et al., 2018). The project has a website, where scans can be initiated for websites to obtain a privacy evaluation, and they plan to provide machine-readable access to the data through an open API, however, the project is still in beta, and this is not possible yet (Maass et al., 2017). A privacy score could be used to triangulate other proxies for browsing privacy. Nevertheless, since it doesn't calculate the metrics using a specific user's browsing settings, it cannot be used to measure the browsing privacy of an individual user.

Furthermore, the dependent variables in the analyses in this thesis were aggregated averages per website visit per day of the study. Information is automatically lost when aggregating. For the amount of change in cookies, this could mean that when cookies were added and deleted on the same day, the value of the corresponding dependent variable could be zero on average, even though there were cookies present. Additionally, the mean is susceptible to outliers, so dependent variables could be biased. Analyzing the progression of individual website visits could identify possible changes in individual behavior directly after being exposed to a boost or a nudge. By first analyzing the relationship between study phase, condition and

browsing privacy using aggregated measures, this thesis lays the foundations for conducting more sophisticated and in-depth analyses on these data.

Like much research utilizing boosts and nudges in the domain of privacy, the manipulation in this study attempted to get participants to change their behavior so they encounter fewer third party requests and less cookies. This is based on the assumption that participants actually want to have more privacy, which may not necessarily be the case. In a study where participants were informed about privacy practices on popular websites, and interviews were conducted about their experience afterwards, some of the participants stated that they did not care about the privacy of their data or that functionality was simply more important for them (Ortloff et al., 2020). This trade-off, where participants may value privacy to some extent, but other features, like functionality or ease of use, more, is called privacy calculus (Kokolakis, 2017). Measuring attitude towards privacy, or privacy concern like in this study, makes it possible to gauge whether people care at all about privacy, and thus whether it even makes sense to try to change their behavior. Changing behavior is only necessary if users are unhappy with the current situation.

Depending on the method of analysis, the effect of nudges and boosts on browsing privacy in this study was marginal. There was either no significant effect, in the case of the multilevel modeling approach, or the effect was not quite trustworthy, in the case of the ART using change in cookies as the dependent variable. Notwithstanding the above, the decision whether to implement nudges or boosts should not only take into account the effectiveness of such an intervention. In fact there has been much criticism of nudges especially because of their lack of transparency regarding the intention behind a certain intervention (Renaud & Zimmermann, 2018). In this study, participants gave informed consent, and were debriefed about both nudges and boosts after the study, but in everyday browsing, this is not possible. It is important to keep such concerns in mind when designing nudges or boosts to deploy publicly, and in some cases, it could be advisable to use boosts instead of nudges (Hertwig, 2017). While there are generally less ethical concerns

about boosts, since these require users to purposefully act on presented knowledge, nudges can also be implemented in an ethical way (Renaud & Zimmermann, 2018).

7. Conclusion

Naturalistic, longitudinal research in the domain of privacy is rare, but nevertheless very important (Lowry et al., 2017). This thesis described the design, implementation and evaluation of a three-week naturalistic experiment, whereby the effect of nudge and boost interventions during the second week of the study was investigated. It contributes to research in the domain of online privacy by providing an extension which can be used to explore users' normal browsing habits, in addition to presenting nudges and boosts to users in a real browsing environment. The comparison of nudges and boosts, as two paradigms to induce behavioral change, in the scope of behavioral log data is another contribution of this thesis. Privacy is generally an interesting domain to compare nudges and boosts, since it spans the whole web and is not only focused on certain sites or applications, such as health search. As such, acquiring larger amounts of naturalistic data and getting a broader picture in the domain of privacy, is easier than in more narrowly focused topics. Additionally, online privacy is a topic that is not limited to online interactions, but which becomes relevant in the real world, for example in the case of identity theft.

Effects of these boost and nudge interventions on browsing privacy, as represented by the number of third party requests, and the change in cookies were not conclusive. Depending on the method of analysis, nudges seemed to have a small positive effect on browsing privacy, as represented by the change in cookies. Boosts did not have a significant effect on browsing privacy, but there was a tendency towards more private browsing behavior, represented by a decreasing number of requests to third parties. Behaviors mentioned in boosts did not change differently in the boost condition, from others, except that more adblockers were installed in this condition than in the others. Problems during the process of fitting models could possibly be resolved by reproducing this research with a larger sample.

Future work could involve trying to resolve the problems which occurred with the study extension. This could incorporate testing how the extension handles browser updates, investigating why it spontaneously uninstalled itself and trying to reduce the computational resources necessary to run it, to improve performance. If the extension worked well enough and provided sufficient added value to users, it might be possible to deploy it through official channels and conduct research with a larger number of participants, but possibly in a less controlled way.

Additionally, the comprehension of boosts was evaluated in a pre-study, but participants' perception of the nudges employed in this study are not yet well understood. Especially the nudge utilizing privacy points should be reexamined, since the number of privacy points assigned for each feature of a website visit, was somewhat arbitrary. Research into the actual effect of these features on users' privacy could help base such a measure on a stronger foundation. Exploring the thoughts of the participants on being presented with this kind of intervention could also be insightful.

Furthermore, in this study, the proxies for browsing privacy were aggregated per day, but the available data enables distinguishing website visits. It could be interesting to retrace the progression of website visits, including possible changes to settings, over time, to identify possible reasons for such changes. Aside from boost or nudge interventions, changes to behavior could, for example, be made depending on website type, e.g. if a participant has the need to feel more secure on a certain type of website. Another reason could be changing settings if a website breaks because of measures intended to protect privacy.

Finally, the use of mobile devices is on the rise, and an increasing amount of people mainly access the internet through their mobile phone (comscore, 2015). However, mobile devices are more vulnerable to privacy breaches, and do not offer as many privacy protective measures (Mylonas et al., 2013). Most research related to privacy on mobile devices focuses on apps (e.g. Almuhimedi et al., 2015; Alohaly & Takabi, 2016), even though browsing and searching the internet is the activity on which the second largest amount of time is spent on mobile devices, after commu-

nication with friends using social media apps (Carrascal & Church, 2015). Thus, this work could and should be extended beyond desktop and laptop computers. Multiple device usage was common in this study, and it would be interesting to investigate how users manage their privacy on multiple devices. Studying such effects further can ultimately help align people's privacy related behavior to the level of privacy they wish to achieve, on any device.

Acknowledgements

At this point, I need to thank the people who helped me on my journey to complete this thesis.

First of all, I am most grateful to my advisers. PD Dr. David Elsweiler often spontaneously found time for me and my questions, even though he had a lot on his own plate otherwise. His constant advice and feedback throughout the whole process kept me on track. Prof. Dr. Niels Henze's critique and comments often provided food for thought for me. Both were willing to work with me on a topic which was neither of their main area of interest.

Steven Zimmerman was introduced to me by Dr. Elsweiler and while he was not an official adviser on my thesis, he felt almost like one at times, providing me with topical expertise in the domain of privacy, boosting and nudging, as well methodological help. I also have him to thank for code samples to work with Selenium, a pointer to the whotracksme-dataset, an example of a Qualtrics survey which I used as a basis for my investigation of boost adequacy and a lot of time spent going over the Qualtrics survey, which he also hosted on his account, due to mine not having necessary privileges.

To Prof. Dr. Udo Kruschwitz I am grateful for agreeing to fund this work, which allowed me to pretest boosts using crowdsourcing, to use an API to categorize websites and to provide vouchers as an incentive to my participants in the main study.

Of course, without these participants, this thesis would have been impossible. I thank both those who were willing to participate in a 3-week study and those who tested the surveys or my extension. Questionnaires and my browser extension were pretested by different people. For their willingness to engage with my thesis in this way, I especially thank Maximiliane Windl and Markus Bink, who tested everything there was to test. Thank you to Robert Jackermeier for deploying my backend to the

information science server multiple times, even on the week-end before the start of my study.

The English in this thesis was very much improved through the help of Karen Schillinger as a proof-reader. Any remaining mistakes are mine!

A big heart-felt thank you also goes to those people who kept me sane during the past months: My family, Christine, Stefan and Anton Ortloff, who nagged me to stop working and got me to take much needed breaks, while still encouraging me on my journey in many different ways. Especially my mom made getting a breather together in fresh air an altogether pleasant affair. Likewise, my flatmate Franziska knocked on my door many times to get me out of my room for dinner or a walk or some ice-cream, and later "forced" me to take an afternoon off per week to visit her.

And finally, I owe so much to my boyfriend Andreas, who bore the brunt of my frustration over the many problems that occur on the way to a Master's degree. Even though he probably did not understand everything I talked about, he always listened to me patiently, both when I was venting and when I was excited and bubbling about a result. He must have felt like I was glued to my laptop more often than not, and nevertheless he replied with nothing but love and reassuring words.

Bibliography

About dbeaver. (2020). https://github.com/dbeaver/dbeaver/wiki. (Last accessed: 2020-07-13, archived at https://archive.st/rg3c)

Acquisti, A., Adjerid, I., Balebako, R., Brandimarte, L., Cranor, L. F., Komanduri, S., ... et al. (2017, 08). Nudges for privacy and security: Understanding and assisting users' choices online. *ACM Computing Surveys (CSUR)*, *50*(3). Retrieved from https://doi.org/10.1145/3054926 doi: 10.1145/3054926

Albrecht, M. (2010, Oct 01). Color blindness. *Nature Methods*, *7*(10), 775-775. Retrieved from https://doi.org/10.1038/nmeth1010-775a doi: 10.1038/nmeth1010-775a

Alexa. (2020). *Top sites in germany.* https://www.alexa.com/topsites/countries/DE. (Last accessed: 2020-02-28)

Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, *95*(9), 1082 - 1095. Retrieved from http://www.sciencedirect.com/science/article/pii/S0047272711000478 (Special Issue: The Role of Firms in Tax Systems) doi: https://doi.org/10.1016/j.jpubeco.2011.03.003

Almuhimedi, H., Schaub, F., Sadeh, N., Adjerid, I., Acquisti, A., Gluck, J., ... Agarwal, Y. (2015). Your location has been shared 5,398 times! a field study on mobile app privacy nudging. In *Proceedings of the 33rd annual acm conference on human factors in computing systems* (p. 787–796). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2702123.2702210 doi: 10.1145/2702123.2702210

Alohaly, M., & Takabi, H. (2016, 06). Better privacy indicators: A new approach to quantification of privacy policies. In *Twelfth symposium on us-*

able privacy and security (SOUPS 2016). Denver, CO: USENIX Association. Retrieved from https://www.usenix.org/conference/soups2016/workshop-program/wpi/presentation/alohaly

Altman, I. (1975). *The environment and social behavior: Privacy, personal space, territory, and crowding*. Monterey: Brooks/Cole.

Altman, I. (1977). Privacy regulation: Culturally universal or culturally specific? *Journal of Social Issues*, *33*(3), 66–84. doi: 10.1111/j.1540-4560.1977.tb01883.x

Baek, Y. M. (2014). Solving the privacy paradox: A counter-argument experimental approach. *Computers in Human Behavior*, *38*, 33 - 42. Retrieved from http://www.sciencedirect.com/science/article/pii/S0747563214002842 doi: https://doi.org/10.1016/j.chb.2014.05.006

Barker, K., Askari, M., Banerjee, M., Ghazinour, K., Mackas, B., Majedi, M., ... Williams, A. (2009). A data privacy taxonomy. In A. P. Sexton (Ed.), *Dataspace: The final frontier* (pp. 42–54). Berlin, Heidelberg: Springer Berlin Heidelberg.

Barocas, S., & Nissenbaum, H. (2009). On notice: The trouble with notice and consent. In *Proceedings of the engaging data forum: The first international forum on the application and management of personal electronic information.* Retrieved from https://ssrn.com/abstract=2567409

Bartlett, J. (2014). *Deviance goodness of fit test for poisson regression.* https://thestatsgeek.com/2014/04/26/deviance-goodness-of-fit-test-for-poisson-regression/. (Last accessed: 2020-07-05)

Baruh, L., Secinti, E., & Cemalcilar, Z. (2017, 01). Online privacy concerns and privacy management: A meta-analytical review. *Journal of Communication*, *67*(1), 26-53. Retrieved from https://doi.org/10.1111/jcom.12276 doi: 10.1111/jcom.12276

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Brandimarte, L., Acquisti, A., & Loewenstein, G. (2013). Misplaced confidences: Privacy and the control paradox. *Social Psychological and Personality Science*, *4*(3), 340-347. Retrieved from https://doi.org/10.1177/1948550612455931 doi: 10.1177/1948550612455931

Brandon, A., Ferraro, P. J., List, J. A., Metcalfe, R. D., Price, M. K., & Rundhammer, F. (2017, March). *Do the effects of social nudges persist? theory and evidence from 38 natural field experiments* (Working Paper No. 23277). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w23277 doi: 10.3386/w23277

Bujlow, T., Carela-Español, V., Solé-Pareta, J., & Barlet-Ros, P. (2017, Aug). A survey on web tracking: Mechanisms, implications, and defenses. *Proceedings of the IEEE*, *105*(8), 1476-1510. doi: 10.1109/JPROC.2016.2637878

Calo, R. (2014). Code, nudge, or notice. *Iowa Law Review*, *99*, 773.

Carrascal, J. P., & Church, K. (2015). An in-situ study of mobile app & mobile search interactions. In *Proceedings of the 33rd annual acm conference on human factors in computing systems* (p. 2739–2748). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2702123.2702486 doi: 10.1145/2702123.2702486

Carrascal, J. P., Riederer, C., Erramilli, V., Cherubini, M., & de Oliveira, R. (2013). Your browsing behavior for a big mac: Economics of personal information online. In *Proceedings of the 22nd international conference on world wide web* (p. 189–200). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2488388.2488406 doi: 10.1145/2488388.2488406

Choe, E. K., Jung, J., Lee, B., & Fisher, K. (2013). Nudging people away from privacy-invasive mobile apps through visual framing. In P. Kotzé, G. Marsden, G. Lindgaard, J. Wesson, & M. Winckler (Eds.), *Human-computer interaction – interact 2013* (pp. 74–91). Berlin, Heidelberg: Springer Berlin Heidelberg.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, N.J.: L. Erlbaum Associates.

comscore. (2015). *Number of mobile-only internet users now exceeds desktop-only in the u.s.* https://www.comscore.com/Insights/Blog/Number-of-Mobile-Only-Internet-Users-Now-Exceeds-Desktop-Only-in-the-U.S. (Last accessed: 2020-07-13, archived at https://archive.st/elq5)

Conway, J., Eddelbuettel, D., Nishiyama, T., Prayaga, S. K., & Tiffin, N. (2017). Rpostgresql: R interface to the 'postgresql' database system [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=RPostgreSQL (R package version 0.6-2)

Cook, R. D., & Weisberg, S. (1982). *Residuals and influence in regression*. New York: Chapman and Hall.

Coventry, L. M., Jeske, D., Blythe, J. M., Turland, J., & Briggs, P. (2016). Personality and social framing in privacy decision-making: A study on cookie acceptance. *Frontiers in Psychology*, *7*, 1341. doi: 10.3389/fpsyg.2016.01341

Coxe, S., West, S. G., & Aiken, L. S. (2009). The analysis of count data: A gentle introduction to poisson regression and its alternatives. *Journal of Personality Assessment*, *91*(2), 121-136. Retrieved from https://doi.org/10.1080/00223890802634175 doi: 10.1080/00223890802634175

Cranor, L. F. (2012). Necessary but not sufficient: Standardized mechanisms for privacy notice and choice. *J. on Telecomm. & High Tech. L.*, *10*, 273.

Cranor, L. F., Dobbs, B., Egelman, S., Hogben, G., Humphrey, J., Langheinrich, M., ... Wenning, R. (2006). *The platform for privacy preferences 1.1 (p3p1.1) specification. w3c working group note 13 november 2006.* https://www.w3.org/TR/2018/NOTE-P3P11-20180830/. (Retired 30 August 2018. Last accessed: 2020-06-11)

Cranor, L. F., Egelman, S., Sheng, S., McDonald, A. M., & Chowdhury, A. (2008). P3p deployment on websites. *Electronic Commerce Research and Applications*, *7*(3),

274 - 293. Retrieved from http://www.sciencedirect.com/science/article/pii/S1567422308000185 (Special Section: New Research from the 2006 International Conference on Electronic Commerce) doi: https://doi.org/10.1016/j.elerap.2008.04.003

Cranor, L. F., & Garfinkel, S. (2004, 09). Guest editors' introduction: Secure or usable? *IEEE Security Privacy*, *2*(5), 16-18. doi: 10.1109/MSP.2004.69

Cranor, L. F., Guduru, P., & Arjula, M. (2006, June). User interfaces for privacy agents. *ACM Transactions on Computer-Human Interaction*, *13*(2), 135–178. Retrieved from https://doi.org/10.1145/1165734.1165735 doi: 10.1145/1165734.1165735

Datta, A., Tschantz, M. C., & Datta, A. (2015). Automated experiments on ad privacy settings. *Proceedings on privacy enhancing technologies*, *2015*(1), 92–112.

Der Altstadt-Zehner – ein Gutschein, 1.000 Möglichkeiten. (2020). https://www.faszination-altstadt.de/gutschein/altstadt-zehner/. (Last accessed: 2020-07-13, archived at https://archive.st/3fwq)

De Rosario-Martinez, H. (2015). phia: Post-hoc interaction analysis [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=phia (R package version 0.2-1)

deutschland.de. (2020). *Die bundesregierung informiert über die corona-krise.* https://www.deutschland.de/de/news/bundesregierung-und-corona-krise. (Last accessed: 2020-07-13, archived at https://archive.st/iegl)

de Visser-Amundson, A., & Kleijnen, M. (2020). Nudging in food waste management: Where sustainability meets cost-effectiveness. In E. Närvänen, N. Mesiranta, M. Mattila, & A. Heikkinen (Eds.), *Food waste management: Solving the wicked problem* (pp. 57–87). Cham: Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-030-20561-4_3 doi: 10.1007/978-3-030-20561-4_3

Egelman, S., Tsai, J., Cranor, L. F., & Acquisti, A. (2009). Timing is everything? the effects of timing and placement of online privacy indicators. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 319–328). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/1518701.1518752 doi: 10.1145/1518701.1518752

Englehardt, S., & Narayanan, A. (2016). Online tracking: A 1-million-site measurement and analysis. In *Proceedings of the 2016 acm sigsac conference on computer and communications security* (p. 1388–1401). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2976749.2978313 doi: 10.1145/2976749.2978313

Ermakova, T., Fabian, B., Bender, B., & Klimek, K. (2018). Web tracking - a literature review on the state of research. In *Proceedings of the 51st hawaii international conference on system sciences* (pp. 4732–4741). Retrieved from http://hdl.handle.net/10125/50485

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using r*. Sage publications.

Flavián, C., & Guinalíu, M. (2006, Jan 01). Consumer trust, perceived security and privacy policy: Three basic elements of loyalty to a web site. *Industrial Management & Data Systems*, *106*(5), 601-620. Retrieved from https://doi.org/10.1108/02635570610666403 doi: 10.1108/02635570610666403

Flaxman, S., Goel, S., & Rao, J. M. (2016, 03). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, *80*(S1), 298-320. Retrieved from https://doi.org/10.1093/poq/nfw006 doi: 10.1093/poq/nfw006

Fouad, I., Bielova, N., Legout, A., & Sarafijanovic-Djukic, N. (2020). Missed by filter lists: Detecting unknown third-party trackers with invisible pixels. *Proceedings on Privacy Enhancing Technologies*, *2020*(2), 499 - 518. Retrieved from https://content.sciendo.com/view/journals/popets/2020/2/article-p499.xml

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (Third ed.). Thousand Oaks CA: Sage. Retrieved from https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Franke, T., Attig, C., & Wessel, D. (2019). A personal resource for technology interaction: Development and validation of the affinity for technology interaction (ati) scale. *International Journal of Human–Computer Interaction*, *35*(6), 456-467. Retrieved from https://doi.org/10.1080/10447318.2018.1456150 doi: 10.1080/10447318.2018.1456150

Fruchter, N., Miao, H., Stevenson, S., & Balebako, R. (2015). Variations in tracking in relation to geographic location. In *Proceedings of the 9th workshop on web 2.0 security and privacy (w2sp) 2015.*

Gao, X., Yang, Y., Fu, H., Lindqvist, J., & Wang, Y. (2014). Private browsing: An inquiry on usability and privacy protection. In *Proceedings of the 13th workshop on privacy in the electronic society* (p. 97–106). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2665943.2665953 doi: 10.1145/2665943.2665953

Geronimo, M. (2017). Online browsing: Can, should, and may companies combine online and offline data to learn about you. *Hastings Science and Technology Law Journal*, *9*(2).

Gervais, A., Filios, A., Lenders, V., & Capkun, S. (2017). Quantifying web adblocker privacy. In S. N. Foley, D. Gollmann, & E. Snekkenes (Eds.), *Computer security – esorics 2017* (pp. 21–42). Cham: Springer International Publishing.

Gigerenzer, G. (2006). Heuristics. In G. Gigerenzer & C. Engel (Eds.), *Heuristics and the law* (pp. 17–44). Cambridge, MA: MIT Press.

Gómez-Boix, A., Laperdrix, P., & Baudry, B. (2018). Hiding in the crowd: An analysis of the effectiveness of browser fingerprinting at large scale. In *Proceedings of the 2018 world wide web conference* (p. 309–318). Republic and Canton

of Geneva, CHE: International World Wide Web Conferences Steering Committee. Retrieved from https://doi.org/10.1145/3178876.3186097 doi: 10.1145/3178876.3186097

Grüne-Yanoff, T., & Hertwig, R. (2016, Mar 01). Nudge versus boost: How coherent are policy and theory? *Minds and Machines*, *26*(1), 149–183. Retrieved from https://doi.org/10.1007/s11023-015-9367-9 doi: 10.1007/s11023-015-9367-9

Hallsworth, M., List, J. A., Metcalfe, R. D., & Vlaev, I. (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics*, *148*, 14 - 31. Retrieved from http://www.sciencedirect.com/science/article/pii/S0047272717300166 doi: https://doi.org/10.1016/j.jpubeco.2017.02.003

Hannak, A., Soeller, G., Lazer, D., Mislove, A., & Wilson, C. (2014). Measuring price discrimination and steering on e-commerce web sites. In *Proceedings of the 2014 conference on internet measurement conference* (p. 305–318). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2663716.2663744 doi: 10.1145/2663716.2663744

Harborth, D., & Pape, S. (2019). How privacy concerns and trust and risk beliefs influence users' intentions to use privacy-enhancing technologies - the case of tor. In *Proceedings of the 52nd hawaii international conference on system sciences.* Retrieved from http://hdl.handle.net/10125/59923 doi: 10.24251/HICSS.2019.585

Hertwig, R. (2017). When to consider boosting: some rules for policy-makers. *Behavioural Public Policy*, *1*(2), 143–161. doi: 10.1017/bpp.2016.14

Hertwig, R., & Grüne-Yanoff, T. (2017). Nudging and boosting: Steering or empowering good decisions. *Perspectives on Psychological Science*, *12*(6), 973-986. Retrieved from https://doi.org/10.1177/1745691617702496 doi: 10.1177/1745691617702496

Hoaglin, D. C. (1980). A poissonness plot. *The American Statistician*, *34*(3), 146–149. Retrieved from http://www.jstor.org/stable/2683871

Isaak, J., & Hanna, M. J. (2018, August). User data privacy: Facebook, cambridge analytica, and privacy protection. *Computer*, *51*(8), 56-59. doi: 10.1109/MC.2018.3191268

Jiang, Z. J., Heng, C. S., & Choi, B. C. F. (2013). Research note—privacy concerns and privacy-protective behavior in synchronous online social interactions. *Information Systems Research*, *24*(3), 579-595. Retrieved from https://pubsonline.informs.org/doi/abs/10.1287/isre.1120.0441 doi: 10.1287/isre.1120.0441

Jibril, A. B., Kwarteng, M. A., Nwaiwu, F., Appiah-Nimo, C., Pilik, M., & Chovancova, M. (2020). Online identity theft on consumer purchase intention: A mediating role of online security and privacy concern. In M. Hattingh, M. Matthee, H. Smuts, I. Pappas, Y. K. Dwivedi, & M. Mäntymäki (Eds.), *Responsible design, implementation and use of information and communication technology* (pp. 147–158). Cham: Springer International Publishing.

John, J. A., & Draper, N. R. (1980). An alternative family of transformations. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *29*(2), 190–197. Retrieved from http://www.jstor.org/stable/2986305

Joni, & Neiman, C. (2019). *Lightbeam extension for firefox is no longer supported.* https://support.mozilla.org/en-US/kb/lightbeam-extension-firefox-no-longer-supported. (Last accessed: 2020-06-09)

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American economic review*, *93*(5), 1449–1475.

Karaj, A., Macbeth, S., Berson, R., & Pujol, J. M. (2018). *Whotracks.me: Shedding light on the opaque world of online tracking.*

Kelley, P. G., Cranor, L. F., & Sadeh, N. (2013). Privacy as part of the app decision-making process. In *Proceedings of the sigchi conference on human factors in computing*

systems (p. 3393–3402). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2470654.2466466 doi: 10.1145/2470654.2466466

Kemp, S. (2019). *Digital trends 2019: Every single stat you need to know about the internet.* https://thenextweb.com/contributors/2019/01/30/digital-trends-2019-every-single-stat-you-need-to-know-about-the-internet/. (Last accessed: 2020-07-13, archived at https://archive.st/hp33)

Kleiber, C., & Zeileis, A. (2008). *Applied econometrics with R*. New York: Springer-Verlag. Retrieved from https://CRAN.R-project.org/package=AER (ISBN 978-0-387-77316-2)

Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security*, *64*, 122 - 134. Retrieved from http://www.sciencedirect.com/science/article/pii/S0167404815001017 doi: https://doi.org/10.1016/j.cose.2015.07.002

Kontaxis, G., & Chew, M. (2015). Tracking protection in firefox for privacy and performance. *Computing Research Repository* (*CoRR*), *abs/1506.04104*. Retrieved from http://arxiv.org/abs/1506.04104

Krishnamurthy, B., Malandrino, D., & Wills, C. E. (2007). Measuring privacy loss and the impact of privacy protection in web browsing. In *Proceedings of the 3rd symposium on usable privacy and security* (p. 52–63). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/1280680.1280688 doi: 10.1145/1280680.1280688

Krishnamurthy, B., Naryshkin, K., & Wills, C. E. (2011). Privacy leakage vs. protection measures: the growing disconnect. In *Web 2.0 workshop on security and privacy* (pp. 2–11).

Kutner, M. H., Nachtsheim, C. J., & Neter, J. (2004). *Applied linear regression models*. New York, NY: McGraw-Hill/Irwin.

Landis, R. J., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. Retrieved from http://www.jstor.org/stable/2529310

Laperdrix, P., Rudametkin, W., & Baudry, B. (2016). Beauty and the beast: Diverting modern web browsers to build unique browser fingerprints. In *2016 ieee symposium on security and privacy (sp)* (pp. 878–894).

Lin, J., Liu, B., Sadeh, N., & Hong, J. I. (2014, 07). Modeling users' mobile app privacy preferences: Restoring usability in a sea of permission settings. In *10th symposium on usable privacy and security (SOUPS 2014)* (pp. 199–212). Menlo Park, CA: USENIX Association. Retrieved from https://www.usenix.org/conference/soups2014/proceedings/presentation/lin

Liu, B., Andersen, M. S., Schaub, F., Almuhimedi, H., Zhang, S. A., Sadeh, N., ... Acquisti, A. (2016, 06). Follow my recommendations: A personalized privacy assistant for mobile app permissions. In *Twelfth symposium on usable privacy and security (SOUPS 2016)* (pp. 27–41). Denver, CO: USENIX Association. Retrieved from https://www.usenix.org/conference/soups2016/technical-sessions/presentation/liu

Long, J. A. (2019). interactions: Comprehensive, user-friendly toolkit for probing interactions [Computer software manual]. Retrieved from https://cran.r-project.org/package=interactions (R package version 1.1.0)

Lowry, P. B., Dinev, T., & Willison, R. (2017). Why security and privacy research lies at the centre of the information systems (is) artefact: proposing a bold research agenda. *European Journal of Information Systems*, *26*(6), 546-563. Retrieved from https://doi.org/10.1057/s41303-017-0066-x doi: 10.1057/s41303-017-0066-x

Lutz, C., & Strathoff, P. (2014). *Privacy concerns and online behavior – not so paradoxical after all? viewing the privacy paradox through different theoretical lenses.* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2425132. SSRN. (Last accessed: 2020-06-09)

Maass, M., Wichmann, P., Pridöhl, H., & Herrmann, D. (2017). Privacyscore: Improving privacy and security via crowd-sourced benchmarks of websites. In E. Schweighofer, H. Leitold, A. Mitrakas, & K. Rannenberg (Eds.), *Privacy technologies and policy* (pp. 178–191). Cham: Springer International Publishing.

Machuletz, D., & Böhme, R. (2019). Multiple purposes, multiple problems: A user study of consent dialogs after GDPR. *arXiv preprint arXiv:1908.10048*.

Malhotra, N. K., Kim, S. S., & Agarwal, J. (2004). Internet users' information privacy concerns (iuipc): The construct, the scale, and a causal model. *Information Systems Research*, *15*(4), 336-355. Retrieved from https://pubsonline.informs.org/doi/abs/10.1287/isre.1040.0032 doi: 10.1287/isre.1040.0032

Margulis, S. T. (1977). Conceptions of privacy: Current status and next steps. *Journal of Social Issues*, *33*(3), 5-21. Retrieved from https://spssi.onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-4560.1977.tb01879.x doi: 10.1111/j.1540-4560.1977.tb01879.x

Margulis, S. T. (2011). Three theories of privacy: An overview. In S. Trepte & L. Reinecke (Eds.), *Privacy online: Perspectives on privacy and self-disclosure in the social web* (pp. 9–17). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from https://doi.org/10.1007/978-3-642-21521-6_2 doi: 10.1007/978-3-642-21521-6_2

Marreiros, H., Tonin, M., Vlassopoulos, M., & Schraefel, M. (2017). "now that you mention it": A survey experiment on information, inattention and online privacy. *Journal of Economic Behavior & Organization*, *140*, 1 - 17. Retrieved from http://www.sciencedirect.com/science/article/pii/S0167268117300896 doi: https://doi.org/10.1016/j.jebo.2017.03.024

Masur, P. K., Teutsch, D., & Trepte, S. (2017). Entwicklung und validierung der online-privatheitskompetenzskala (oplis). *Diagnostica*, *63*(4), 256-268. Retrieved from https://doi.org/10.1026/0012-1924/a000179 doi: 10.1026/0012-1924/a000179

Maurer, M.-E. (2010). Bringing effective security warnings to mobile browsing. In *Proceedings of the 2nd international workshop on security and privacy in spontaneous interaction and mobile phone use.*

Mayer, J. R., & Mitchell, J. C. (2012). Third-party web tracking: Policy and technology. In *2012 ieee symposium on security and privacy* (p. 413-427). doi: 10.1109/SP.2012.47

Mazel, J., Garnier, R., & Fukuda, K. (2019). A comparison of web privacy protection techniques. *Computer Communications*, *144*, 162 - 174. Retrieved from http://www.sciencedirect.com/science/article/pii/S0140366418300604 doi: https://doi.org/10.1016/j.comcom.2019.04.005

MDN contributors. (2019). *Browser support for javascript apis.* https://developer.mozilla.org/en-US/docs/Mozilla/Add-ons/WebExtensions/Browser_support_for_JavaScript_APIs. (Last accessed: 2020-06-02)

MDN contributors. (2020a). *Browser extensions.* https://developer.mozilla.org/en-US/docs/Mozilla/Add-ons/WebExtensions. (Last accessed: 2020-06-02)

MDN contributors. (2020b). *Chrome incompatibilities.* https://developer.mozilla.org/de/docs/Mozilla/Add-ons/WebExtensions/Chrome_incompatibilities. (Last accessed: 2020-06-02)

MDN contributors. (2020c). *Match patterns in extension manifests.* https://developer.mozilla.org/en-US/docs/Mozilla/Add-ons/WebExtensions/Match_patterns#%3Call_urls%3E. (Last accessed: 2020-07-13, archived at https://archive.st/8eue)

Melicher, W., Sharif, M., Tan, J., Bauer, L., Christodorescu, M., & Leon, P. G. (2016). (do not) track me sometimes: Users' contextual preferences for web tracking. *Proceedings on Privacy Enhancing Technologies*, *2016*(2), 135 - 154. Retrieved from https://content.sciendo.com/view/journals/popets/2016/2/article-p135.xml

Microsoft. (2016). *P3P is no longer supported.* https://docs.microsoft.com/en-us/previous-versions/windows/internet-explorer/ie-developer/compatibility/mt146424(v=vs.85)?redirectedfrom=MSDN. (Last accessed: 2020-06-11)

Mikians, J., Gyarmati, L., Erramilli, V., & Laoutaris, N. (2012). Detecting price and search discrimination on the internet. In *Proceedings of the 11th acm workshop on hot topics in networks* (p. 79–84). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2390231.2390245 doi: 10.1145/2390231.2390245

Mikians, J., Gyarmati, L., Erramilli, V., & Laoutaris, N. (2013). Crowd-assisted search for price discrimination in e-commerce: First results. In *Proceedings of the ninth acm conference on emerging networking experiments and technologies* (p. 1–6). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2535372.2535415 doi: 10.1145/2535372.2535415

Mueller, T., Marx, M., Pridoehl, H., Wichmann, P., & Herrmann, D. (2018). Sicherheit und privatheit auf deutschen hochschulwebseiten: Eine analyse mit privacyscore. In *25. dfn-konferenz "sicherheit in vernetzten systemen"*.

Mylonas, A., Tsalis, N., & Gritzalis, D. (2013). Evaluating the manageability of web browsers controls. In R. Accorsi & S. Ranise (Eds.), *Security and trust management* (pp. 82–98). Berlin, Heidelberg: Springer Berlin Heidelberg.

Mysore Sathyendra, K., Wilson, S., Schaub, F., Zimmeck, S., & Sadeh, N. (2017, September). Identifying the provision of choices in privacy policy text. In *Pro-*

ceedings of the 2017 conference on empirical methods in natural language processing (pp. 2774–2779). Copenhagen, Denmark: Association for Computational Linguistics. Retrieved from https://www.aclweb.org/anthology/D17-1294 doi: 10.18653/v1/D17-1294

Nienaber, M. (2018). *Germany to raise minimum wage to 9.35 euros in 2020.* https://uk.reuters.com/article/us-germany-economy-wages-idUKKBN1JM1AS. (Last accessed: 2020-07-13, archived at https://archive.st/h7hs)

Norberg, P. A., Horne, D. R., & Horne, D. A. (2007). The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of Consumer Affairs*, *41*(1), 100-126. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1745-6606.2006.00070.x doi: 10.1111/j.1745-6606.2006.00070.x

Obar, J. A. (2016). The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services. *SSRN Electronic Journal*. doi: 10.2139/ssrn.2757465

Ortloff, A.-M., Güntner, L., Windl, M., Feth, D., & Polst, S. (2018). Evaluation kontextueller Datenschutzerklärungen. In R. Dachselt & G. Weber (Eds.), *Mensch und Computer 2018 - Workshopband*. Bonn: Gesellschaft für Informatik e.V.

Ortloff, A.-M., Windl, M., Schwind, V., & Henze, N. (2020). Implementation and in situ assessment of contextual privacy policies. In *Proceedings of the 2020 conference on designing interactive systems*. Association for Computing Machinery. doi: 10.1145/3357236.3395549

Papadopoulos, P., Kourtellis, N., & Markatos, E. (2019). Cookie synchronization: Everything you always wanted to know but were afraid to ask. In *The world wide web conference* (p. 1432–1442). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3308558.3313542 doi: 10.1145/3308558.3313542

Pawitan, Y. (2001). *In all likelihood: statistical modelling and inference using likelihood*. Oxford University Press.

Pötzsch, S. (2009). Privacy awareness: A means to solve the privacy paradox? In V. Matyáš, S. Fischer-Hübner, D. Cvrček, & P. Švenda (Eds.), *The future of identity in the information society* (pp. 226–236). Berlin, Heidelberg: Springer Berlin Heidelberg.

Prolific Team. (2018). *Using attention checks as a measure of data quality.* https://researcher-help.prolific.co/hc/en-gb/articles/360009223553-Using-attention-checks-as-a-measure-of-data-quality. (Last accessed: 2020-06-01)

Pugliese, G. (2015). Web tracking: Overview and applicability in digital investigations. *it - Information Technology*, *57*(6), 366 - 375. Retrieved from https://www.degruyter.com/view/journals/itit/57/6/article-p366.xml

R Core Team. (2019). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

Rebonato, R. (2012). *Taking liberties: A critical examination of libertarian paternalism*. Palgrave Macmillan.

Reidenberg, J. R., Breaux, T., Carnor, L. F., French, B., Grannis, A., Graves, J. T., ... Schaub, F. (2014). Disagreeable privacy policies: Mismatches between meaning and users understanding. *Berkeley Technology Law Journal*, *30*(1). doi: 10.15779/Z384K33

Reijula, S., Kuorikoski, J., Ehrig, T., Katsikopoulos, K., & Sunder, S. (2018). *Nudge, boost, or design? limitations of behaviorally informed policy under social interaction.* https://doi.org/10.31235/osf.io/zh3qw. (Last accessed: 2020-06-09)

Renaud, K., & Zimmermann, V. (2018). Ethical guidelines for nudging in information security & privacy. *International Journal of Human-Computer Studies*, *120*, 22 -

35. Retrieved from http://www.sciencedirect.com/science/article/pii/S1071581918302787 doi: https://doi.org/10.1016/j.ijhcs.2018.05.011

Riggs, W. (2017). Painting the fence: Social norms as economic incentives to non-automotive travel behavior. *Travel Behaviour and Society*, *7*, 26 - 33. Retrieved from http://www.sciencedirect.com/science/article/pii/S2214367X16300357 doi: https://doi.org/10.1016/j.tbs.2016.11.004

Sanchez-Rola, I., Dell'Amico, M., Kotzias, P., Balzarotti, D., Bilge, L., Vervier, P.-A., & Santos, I. (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* (p. 340–351). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3321705.3329806 doi: 10.1145/3321705.3329806

Sanchez-Rola, I., Ugarte-Pedrero, X., Santos, I., & Bringas, P. G. (2016, 08). The web is watching you: A comprehensive review of web-tracking techniques and countermeasures. *Logic Journal of the IGPL*, *25*(1), 18-29. Retrieved from https://doi.org/10.1093/jigpal/jzw041 doi: 10.1093/jigpal/jzw041

Schöbel, S., Barev, T., Janson, A., Hupfeld, F., & Leimeister, J. M. (2020). Understanding user preferences of digital privacy nudges–a best-worst scaling approach. In *Proceedings of the 53rd hawaii international conference on system sciences.* doi: 10.24251/HICSS.2020.479

Selenium. (2020). *The selenium browser automation project.* https://www.selenium.dev/documentation/en/. (Last accessed: 2020-07-13, archived at https://archive.st/fbxh)

Sequeira, P. M. (2019). *lightbeam-we.* https://github.com/princiya/lightbeam-we. (Last accessed: 2020-07-13, archived at https://archive.st/i2cc)

Sobolewski, M., Mazur, J., & Paliński, M. (2017, Jul 01). GDPR: A step towards a user-centric internet? *Intereconomics*, *52*(4), 207–213. Retrieved from https://doi.org/10.1007/s10272-017-0676-5 doi: 10.1007/s10272-017-0676-5

Sørensen, J., & Kosta, S. (2019). Before and after gdpr: The changes in third party presence at public and private european websites. In *The world wide web conference* (p. 1590–1600). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3308558.3313524 doi: 10.1145/3308558.3313524

Sotirakopoulos, A., Hawkey, K., & Beznosov, K. (2011). On the challenges in usable security lab studies: Lessons learned from replicating a study on ssl warnings. In *Proceedings of the seventh symposium on usable privacy and security.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2078827.2078831 doi: 10.1145/2078827.2078831

Statcounter - GlobalStats. (2020). *Desktop browser market share germany. april 2019 - may 2020.* https://gs.statcounter.com/browser-market-share/desktop/germany. (Last accessed: 2020-06-02)

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Hillsdale, NJ: Erlbaum.

Taddicken, M. (2014, 01). The 'privacy paradox' in the social web: The impact of privacy concerns, individual characteristics, and the perceived social relevance on different forms of self-disclosure. *Journal of Computer-Mediated Communication*, *19*(2), 248-273. Retrieved from https://doi.org/10.1111/jcc4.12052 doi: 10.1111/jcc4.12052

Tageschau. (2020). *Aktives Ja zu Cookies muss sein.* https://www.tagesschau.de/inland/cookies-bundesgerichtshof-101.html.

Takano, Y., Ohta, S., Takahashi, T., Ando, R., & Inoue, T. (2014, July). Mindyourprivacy: Design and implementation of a visualization system for third-party web

tracking. In *2014 twelfth annual international conference on privacy, security and trust* (p. 48-56). IEEE. doi: 10.1109/PST.2014.6890923

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven & London: Yale University Press.

Trepte, S., Teutsch, D., Masur, P. K., Eicher, C., Fischer, M., Hennhöfer, A., & Lind, F. (2015). Do people know about privacy and data protection strategies? towards the 'online privacy literacy scale' (OPLIS). In S. Gutwirth, R. Leenes, & P. de Hert (Eds.), *Reforming european data protection law* (pp. 333–365). Dordrecht: Springer Netherlands. Retrieved from https://doi.org/10.1007/978-94-017-9385-8_14 doi: 10.1007/978-94-017-9385-8_14

Tsai, J. Y., Egelman, S., Cranor, L., & Acquisti, A. (2011). The effect of online privacy information on purchasing behavior: An experimental study. *Information Systems Research*, *22*(2), 254-268. Retrieved from https://pubsonline.informs.org/doi/abs/10.1287/isre.1090.0260 doi: 10.1287/isre.1090.0260

Tsalis, N., Mylonas, A., Nisioti, A., Gritzalis, D., & Katos, V. (2017). Exploring the protection of private browsing in desktop browsers. *Computers & Security*, *67*, 181 - 197. Retrieved from http://www.sciencedirect.com/science/article/pii/S0167404817300597 doi: https://doi.org/10.1016/j.cose.2017.03.006

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.

Ur, B., Leon, P. G., Cranor, L. F., Shay, R., & Wang, Y. (2012). Smart, useful, scary, creepy: Perceptions of online behavioral advertising. In *Proceedings of the eighth symposium on usable privacy and security.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2335356.2335362 doi: 10.1145/2335356.2335362

Urban, T., Degeling, M., Holz, T., & Pohlmann, N. (2020). Beyond the front page:measuring third party dynamics in the field. In *Proceedings of the web con-*

ference 2020 (p. 1275–1286). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3366423.3380203 doi: 10.1145/3366423.3380203

Urban, T., Tatang, D., Degeling, M., Holz, T., & Pohlmann, N. (2019). A study on subject data access in online advertising after the gdpr. In C. Pérez-Solà, G. Navarro-Arribas, A. Biryukov, & J. Garcia-Alfaro (Eds.), *Data privacy management, cryptocurrencies and blockchain technology* (pp. 61–79). Cham: Springer International Publishing.

Utz, C., Degeling, M., Fahl, S., Schaub, F., & Holz, T. (2019). (Un)Informed consent: Studying GDPR consent notices in the field. In *Proceedings of the 2019 acm sigsac conference on computer and communications security* (p. 973–990). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3319535.3354212 doi: 10.1145/3319535.3354212

van Bavel, R., Rodríguez-Priego, N., Vila, J., & Briggs, P. (2019). Using protection motivation theory in the design of nudges to improve online security behavior. *International Journal of Human-Computer Studies*, *123*, 29 - 39. Retrieved from http://www.sciencedirect.com/science/article/pii/S1071581918306475 doi: https://doi.org/10.1016/j.ijhcs.2018.11.003

Wang, Y., Norcie, G., Komanduri, S., Acquisti, A., Leon, P. G., & Cranor, L. F. (2011). "i regretted the minute i pressed share": A qualitative study of regrets on facebook. In *Proceedings of the seventh symposium on usable privacy and security.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2078827.2078841 doi: 10.1145/2078827.2078841

Webshrinker. (n.d.). *Category taxonomies.* https://docs.webshrinker.com/v3/website-category-api.html#category-taxonomies. (Last accessed: 2020-06-02)

Westin, A. F. (1967). *Privacy and freedom*. New York: Atheneum.

Westin, A. F. (2003). Social and political dimensions of privacy. *Journal of Social Issues*, *59*(2), 431-453. Retrieved from https://spssi.onlinelibrary.wiley.com/doi/abs/10.1111/1540-4560.00072 doi: 10.1111/1540-4560.00072

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. doi: 10.21105/joss.01686

Wilcox, R. R., & Schönbrodt, F. D. (2014). The WRS package for robust statistics in R (version 0.24) [Computer software manual]. Retrieved from http://r-forge.r-project.org/projects/wrs/

Wills, C. E., & Tatar, C. (2012). Understanding what they do with what they know. In *Proceedings of the 2012 acm workshop on privacy in the electronic society* (p. 13–18). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2381966.2381969 doi: 10.1145/2381966.2381969

Wirtz, J., Lwin, M. O., & Williams, J. D. (2007, Jan 01). Causes and consequences of consumer online privacy concern. *International Journal of Service Industry Management*, *18*(4), 326-348. Retrieved from https://doi.org/10.1108/09564230710778128 doi: 10.1108/09564230710778128

Wobbrock, J. O., Findlater, L., Gergle, D., & Higgins, J. J. (2011). The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the acm conference on human factors in computing systems* (*chi '11*) (pp. 143–146). New York: ACM Press. Retrieved from http://depts.washington.edu/aimgroup/proj/art/

Ziegeldorf, J. H., Henze, M., Hummen, R., & Wehrle, K. (2016). Comparison-based privacy: Nudging privacy in social media (position paper). In J. Garcia-Alfaro, G. Navarro-Arribas, A. Aldini, F. Martinelli, & N. Suri (Eds.), *Data privacy management, and security assurance* (pp. 226–234). Cham: Springer International Publishing.

Zimmerman, S., Thorpe, A., Chamberlain, J., & Kruschwitz, U. (2020). Towards search strategies for better privacy and information. In *Proceedings of the 2020 conference on human information interaction and retrieval* (p. 124–134). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3343413.3377958 doi: 10.1145/3343413.3377958

Zimmerman, S., Thorpe, A., Fox, C., & Kruschwitz, U. (2019a). Investigating the interplay between searchers' privacy concerns and their search behavior. In *Proceedings of the 42nd international acm sigir conference on research and development in information retrieval* (p. 953–956). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3331184.3331280 doi: 10.1145/3331184.3331280

Zimmerman, S., Thorpe, A., Fox, C., & Kruschwitz, U. (2019b). Privacy nudging in search: Investigating potential impacts. In *Proceedings of the 2019 conference on human information interaction and retrieval* (p. 283–287). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3295750.3298952 doi: 10.1145/3295750.3298952

A. Boosts

boost	used to test pre-study	used in pre-study	used in final study
Nachrichten- und Medien-Webseiten haben mehr Cookies und Drittanbieter pro Seite, als andere Arten von Webseiten.	x	x	
Webseiten, die sich mit Unterhaltung beschäftigen haben mehr Cookies und Drittanbieter pro Seite, als die meisten anderen Arten von Webseiten.	x	x	x
Webseiten, die sich mit (Weiter-)Bildung beschäftigen haben weniger Cookies und Drittanbieter pro Seite, als die meisten anderen Arten von Webseiten.	x	x	x
Das Blockieren von Drittanbieter-Cookies (Third Party Cookies) in den Browsereinstellungen führt zu einer Verringerung der Anzahl Drittanbieter pro Seite von circa 30%.	x	x	x
Das Blockieren von Drittanbieter-Cookies (Third Party Cookies) in den Browsereinstellungen führt zu einer Verringerung von Cookies insgesamt.	x		
Durch das Nutzen des privaten Modus (Firefox) bzw. Inkognito Modus (Chrome) werden Cookies nach dem Schließen des Browsers automatisch gelöscht.	x	x	x
Durch das Nutzen eines Adblockers werden, bereits ohne die Blockereinstellungen zu ändern, die Anzahl der Drittanbieter auf einer Seite um 40% reduziert.	x	x	x

Table A.1.: Original German version of boosts used in different phases of study preparation and study

B. Webshrinker Categories

Abortion Sites which provide views either in favor or against abortion, provide details on procedures, offer help or discuss outcomes or consequences of abortion.

Adult Sites which may contain sexually explicit content, images, or that are portrayed through visually expressive language.

Advertising Sites or businesses which directly sell ads to consumer through various mediums, including Internet, TV, or radio.

Alcohol and Tobacco Sites that sell, discuss, or glorify the consumption of various alcoholic and tobacco products, including beer, wine, and liquor.

Blogs and Personal Sites Includes sites that make use of common blogging software including WordPress, Joomla!, and Drupal, amongst others, which generate dynamic content. Normally used as a secondary category to a more descriptive primary category.

Business Sites which exhibit business-like attributes such as selling of services, products, or consulting. Normally used as a secondary category to a more descriptive primary category.

Chat and Instant Messaging Sites which provide chat or text messaging services or such abilities through a download or application.

Content Server Sites whose main purpose is to serve static image, CSS, and JavaScript files.

Cryptomining Sites which serve files or host applications that force the web browser to mine cryptocurrency, often utilizing considerable system, network, and power resources.

Deceptive Sites that attempt to trick the user into believing they are on a different site in order to gather information or for other purposes. Also includes sites with deceptive advertising practices such as performing click redirections without the users consent.

Drugs Sites that contain content whose main focus is on controlled substances, including the sale, discussion, or glorification of such substances. Does not include alcohol and tobacco as that has its own category.

Economy and Finance Includes sites that are mainly focused on stocks and current market information or provide financial services such as banks or lenders.

Education & Self-Help Sites whose main purpose is to offer educational information, community information, or how-tos. Also includes educational facilities and related organizations.

Entertainment Sites that focus on art and entertainment, including topics like TV/Hollywood, tattoos, cartoons and anime.

Food and Recipes Sites that contain food related information or recipes, food preparation, or restaurant services.

Gambling Sites that allow a visitor to play games using wagers/placing bets, lottery pools, or provides information on such activities.

Games Sites that provide games, including online games or through an application.

Hacking & Cracking Sites that disseminate information, hold discussions, or provide a means to gain unauthorized or illegal access to computers and networks.

Health Sites whose content is focused on a person's well-being, including fitness or workout information, medical conditions, diagnoses, and medical services.

Humor Sites that contain content with a focus on jokes or comedy, including satire.

Illegal Content Sites that focus on providing links to pirated movies, commercial software, or providing application keys and cracks for commercial applications.

Information Technology Sites whose main focus is on computers or distributing computer related information, including computer networking, Internet telephony, operating systems, or programming.

Jobs & Careers Sites that focus on assisting visitors in finding employment, career guidance or improvement.

Malicious Sites that are infected with or distribute any kind of malware, spyware, or viruses. Also contains sites acting as a C&C for bots, worms, trojans, and other malware.

Media Sharing Sites that allow visitors to upload content and share media such as photos and videos.

Messageboards and Forums Sites which provide some type of a messaging or bulletin board system whose content is largely community generated.

News and Media Sites whose content is mostly focused on current events and topics. Includes various news outlets, radio, TV stations, and magazines.

Parked Sites & Domains Sites/domain names that are no longer controlled by the original owner or are being offered for sale. Content on these sites can often be misleading for non technical users.

Dating and Personals Sites whose main focus is on connecting individuals for the purposes of dating.

Proxy and Filter Avoidance Sites that provides information or a means to circumvent filtering proxies or detection systems, including VPN services and anonymous surfing.

Real Estate Sites that provide information or services for renting, selling, or buying property.

Religion Sites that provide information on one or more religious beliefs, practices, or are affiliated with a religious institution such as a church or synagogue.

Search Engines and Portals Sites that enable their visitors to search the Internet or whose main focus is to provide links to other Internet sites.

Shopping Sites that sell products or services, normally with an online purchasing interface.

Social Networking Sites that provide a community portal whereby members join and contribute posts or media and forge connections with other members.

Sports Sites that contain information about various sports and sporting activities, including sports scores or team information.

Streaming Media Sites that are mainly dedicated to the serving of video or audio streams and downloads.

Translation Sites Sites that perform translation from one language to another, usually performed by a computer. May also include personal translation services.

Travel Sites that focus on travel planning services, travel reservations, and tourist information.

Uncategorized Sites that are not currently classified as belonging to one of the other categories or are not yet classified.

Vehicles Sites that mainly hold discussions or share information about vehicles, including cars, trucks, boats, and aircraft.

Virtual Reality Sites that host files specific to virtual reality or run communities related to the technology.

Weapons Sites that primarily discuss, review, or sell items such as hunting knives, guns, rifles, or BB guns.

C. Questionnaires

This annex contains all questionnaires used in the main study conducted in this thesis. The German versions of questionnaires translated from English are in Section C.2, while questionnaires, which are not from previous work are in Section C.1. All other questionnaires are published in German versions. These are the ATI (Franke et al., 2019) and OPLIS (Masur et al., 2017) scales.

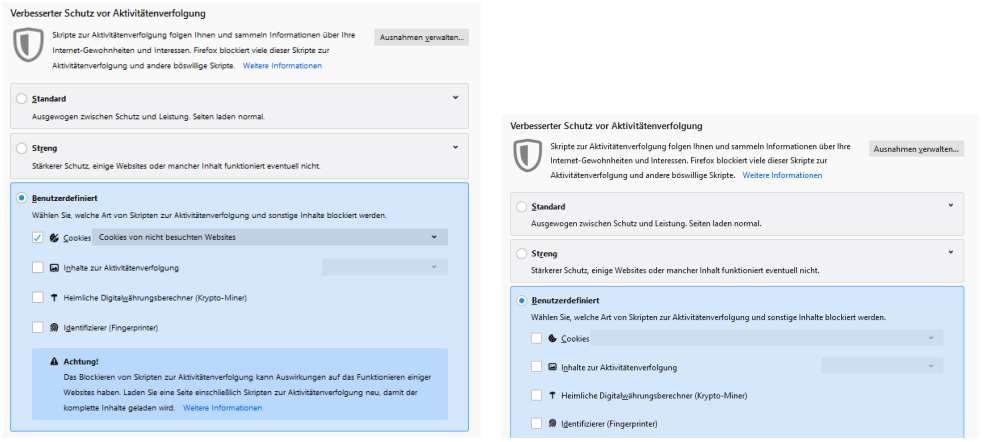
Any questions without further annotations required participants to enter text or numbers. Other questions include the choices available to participants. If participants were able to select multiple choices in a question, this is annotated as well.

C.1. Original Questionnaires

All questionnaires, including the ones in Annex C.2 were preceded by a question asking for the participant's participant id. Depending on when a questionnaire was administered, a second question referring to the participant's id, was included.

Asked at the beginning and the end of the study Bitte geben Sie hier Ihre aktuelle Teilnehmer-Id aus der Browser-Erweiterung an. Diese finden Sie, wenn Sie auf das Lupen-Icon oben rechts in Ihrem Browser klicken. Dann öffnet sich ein Fenster, in dem Sie unter anderem Ihre Teilnehmer-Id finden. Diese sollte eine 5 bis 7-stellige Zahl sein.

Asked only at the end of the study Falls Sie während der Studie mehrere Teilnehmer-Ids erhalten haben, z.B. weil Sie die Erweiterung neu installiert haben, geben Sie bitte alle hier an. Falls Sie sich nicht sicher sind, geben Sie alles an, an das Sie sich erinnern (Beginn oder Ende mit einer bestimmten Zahl, Anzahl Ziffern).



(a) Pre-study version (b) Main-study version

Figure C.1.: Images used in boost knowledge questionnaire, depicting Firefox settings of third party cookies not being blocked

C.1.1. Questionnaire on Boost-related Knowledge

There are two versions of this questionnaire. One is the questionnaire used in the pre-study. A Qualtrics file for this questionnaire is in the additional material. It can be viewed using a free Qualtrics account. A PDF-Version for this questionnaire, generated by Qualtrics, is also in the additional material, but this version may suffer from readability issues and is merely meant for convenience.

The second version is the questionnaire administered in the main study, both at the beginning and the end of the study. It encompasses only the questions from the control condition of the questionnaire above. One image was replaced in the main study version of the questionnaire. A comparison of both images is in Figure C.1.

C.1.2. Demographic End-of-Study Questionnaire

This questionnaire was administered at the end of the study. Since the problem of multiple participant ids was known at this time, there were two questions regarding the participant ids at the beginning of the study. It contained several conditional questions, which are marked with [**conditional**] here, and were only displayed to participants if they answered the question before the conditional questions affirmatively.

1. Welches Betriebssystem läuft auf dem Rechner, den Sie für die Studie genutzt haben?
   Choices were: “Windows”, “Mac Os”, “Linux”, “Weitere...”
2. Für welchen Browser haben Sie die Erweiterung ”Browsing Studie” installiert?
   Choices were: “Chrome”, “Firefox”
3. Waren Sie während der Studiendauer noch mit einem anderen Laptop- oder Desktop-Gerät im Internet, neben demjenigen, auf dem die Erweiterung installiert war?
   Choices were: “Ja”, “Nein”

4. Sie haben angegeben, dass Sie während der Teilnahme an der Studie noch ein weiteres Laptop- oder Desktopgerät genutzt haben, neben demjenigen, auf dem die Studienerweiterung installiert wurde. [**conditional**]
   1) Warum haben Sie noch ein weiteres Laptop- oder Desktopgerät zur Nutzung des Internets verwendet?
   2) Etwa viel Prozent der Zeit haben Sie dieses andere Laptop- oder Desktopgerät, im Vergleich zu demjenigen mit der Studienerweiterung, verwendet?
      (Geben Sie eine Zahl zwischen 0 und 100 an. Wenn Sie das andere Laptop- oder Desktopgerät häufiger als das Gerät mit Studienerweiterung genutzt haben, dann sollte Ihre Zahl größer als 50 sein, weil Sie das andere Laptop- oder Desktopgerät mehr als 50% der Zeit genutzt haben.)
   3) In welchem Verhältnis haben Sie dieses andere Laptop- oder Desktopgerät, im Vergleich zu demjenigen mit der Studienerweiterung verwendet?
      Choices were:
      - Ich habe in über 90% der Zeit das Gerät mit der Studienerweiterung verwendet
      - Ich habe das Gerät mit der Studienerweiterung in zwischen 90% und 60% der Zeit verwendet

- Ich habe das Gerät mit der Studienerweiterung etwa gleich häufig verwendet, wie das/die anderen ähnlichen Geräte (d.h. in 41% bis 59% der Zeit)
- Ich habe das Gerät mit der Studienerweiterung in zwischen 40% und 10% der Zeit verwendet
- Ich habe in weniger als 10% der Zeit das Gerät mit der Studienerweiterung verwendet

5. Nutzen Sie den Browser, den Sie in der Studie genutzt haben, auch normalerweise?
   Choices were: "Ja", "Nicht so oft wie während der Studie", "Nein"
6. Falls Sie in der letzten Frage nicht mit "Ja" geantwortet haben, welche Unterschiede haben Sie zwischen dem Browser, den Sie normalerweise benutzen, und dem Browser, den Sie in der Studie benutzt haben, bemerkt?
7. Haben Sie während Ihrer Studienteilnahme auf dem Gerät, mit dem Sie an der Studie teilgenommen haben, noch andere Browser verwendet, neben demjenigen, auf dem die Erweiterung installiert war?
   Choices were: "Ja", "Nein"

8. Sie haben angegeben, dass Sie während der Studie mindestens einen weiteren Browser zusätzlich zu dem, auf dem die Studienerweiterung installiert wurde, genutzt haben. [**conditional**]
   1) Welche(n) Browser haben Sie zusätzlich verwendet?
   2) Etwa viel Prozent der Zeit haben Sie diese/n anderen Browser, im Vergleich zu demjenigen mit der Studienerweiterung, verwendet?
      (Geben Sie eine Zahl zwischen 0 und 100 an. Wenn Sie den/die anderen Browser häufiger als denjenigen mit der Studienerweiterung genutzt haben, dann sollte Ihre Zahl größer als 50 sein, weil Sie den/die anderen Browser mehr als 50% der Zeit genutzt haben.)
   3) Warum haben Sie diese(n) Browser zusätzlich verwendet?

4) In welchem Verhältnis haben Sie diese/n anderen Browser im Vergleich zum Browser mit der Studienerweiterung verwendet?
   Choices were:
   - Ich habe in über 90% der Zeit den Browser mit der Studienerweiterung verwendet
   - Ich habe den Browser mit der Studienerweiterung in zwischen 90% und 60% der Zeit verwendet
   - Ich habe den Browser mit der Studienerweiterung etwa gleich häufig verwendet, wie den/die anderen Browser (d.h. in 41% bis 59% der Zeit)
   - Ich habe den Browser mit der Studienerweiterung in zwischen 40% und 10% der Zeit verwendet
   - Ich habe in weniger als 10% der Zeit den Browser mit der Studienerweiterung verwendet

9. Sind in Ihrem Browser zusätzliche Browser-Erweiterungen installiert?
   Choices were: "Ja", "Nein"

10. Sie haben angegeben, dass in Ihrem Browser zusätzliche Browsererweiterungen installiert sind, neben derjenigen, die für die Teilnahme an der Studie genutzt wurde. [**conditional**]
    1) Wie viele zusätzliche Browsererweiterungen sind in Ihrem Browser installiert?
    2) Haben Sie während der Dauer der Studie zusätzliche Browsererweiterungen installiert?
       Choices were "Ja", "Nein"
    3) Wenn ja, welche zusätzlichen Browsererweiterungen haben Sie während der Studie installiert?

4) Sind in Ihrem Browser Erweiterungen installiert, die Ihre Privatsphäre im Internet schützen sollen?
   Choices were: “Ja”, “Nein”
5) Wenn ja, welche?

11. Sind in Ihrem Browser oder mit der Studienerweiterung während der Studie Probleme aufgetreten?
    Choices were: “Ja”, “Nein”

12. Sie haben angegeben, dass in Ihrem Browser oder mit der Studienerweiterung während der Studie Probleme aufgetreten sind. [**conditional**]
    1) Welche der folgenden Probleme sind während der Studie bei Ihnen aufgetreten?
       Choices were:
       - Der Browser, den ich der Studie genutzt habe, war langsamer als sonst.
       - Die Lüftung meines Laptops war lauter als sonst
       - Ich musste die Erweiterung während der Studie neu installieren.
       - Weitere...
    2) Bitte geben Sie hier, falls erforderlich, eine genauere Beschreibung Ihres Problems an.
13. Als welches Geschlecht identifizieren Sie sich?
    Choices were: “männlich”, “weiblich”, “divers”, “ich möchte mein Geschlecht nicht angeben”
14. Wie alt sind Sie?
15. Bitte bewerten Sie Ihre Fähigkeit, Deutsch zu sprechen.
    Choices were:
    - muttersprachlich
    - nahezu muttersprachlich / fließend

- ausgezeichnete Beherrschung / hohe Kompetenz in gesprochenem und geschriebenem Deutsch
- sehr gute Beherrschung
- gute Beherrschung / gute praktische Kenntnisse
- grundlegende Kommunikationsfähigkeit / praktische Kenntnisse

16. Was ist der höchste Bildungsabschluss, den Sie erreicht haben? (Wenn Sie sich noch in der Weiter-/Ausbildung befinden, wählen Sie bitte den höchsten Abschluss, den Sie bereits erreicht haben)
    Choices were:
    - Hauptschulabschluss/Volksschulabschluss
    - Realschulabschluss oder gleichwertig
    - Fachhochschul- oder Hochschulreife
    - Abgeschlossene Berufsausbildung
    - Bachelorabschluss oder gleichwertig
    - Masterabschluss oder gleichwertig (z.B. Staatsexamen, Magister, Diplom)
    - abgeschlossene Promotion
17. Welche Aussage beschreibt Ihren derzeitigen Erwerbsstatus am besten?
    Choices were: "In Ausbildung/Studium", "Erwerbstätig", "Derzeit nicht erwerbstätig", "Im Ruhestand"
18. Bitte beschreiben Sie Ihre derzeitige Beschäftigung kurz etwas genauer, d.h. in welchem Studiengang studieren Sie bzw. als was arbeiten Sie.
19. Hier ist noch Platz für eventuelle Kommentare zur Studie:

C.2. Translated Questionnaires

C.2.1. IUIPC Questionnaire

Below is the German version of the IUIPC scale and related concepts, as received in personal communication from David Harborth on May 5th 2020. Collection, aware-

ness, and control are subscales of the IUIPC, while trusting beliefs and risk beliefs are related concepts (Malhotra et al., 2004). All items are used with 7-point likert scales, labeled with “lehne stark ab”, “lehne ab”, “lehne eher ab”, “teils/teils”, “stimme eher zu”, “stimme zu”, “stimme stark zu” in this order.

Trusting Beliefs:

1. Online-Unternehmen sind vertrauenswürdig bezüglich des Umgangs mit Informationen.
2. Online-Unternehmen sagen die Wahrheit und halten die Versprechen in Bezug auf die von mir bereitgestellten Informationen ein.
3. Ich vertraue darauf, dass Online-Unternehmen beim Umgang mit Informationen in meinem Interesse handeln würden.
4. Online-Unternehmen sind in Bezug auf die Nutzung der Informationen im Allgemeinen berechenbar und beständig.
5. Online-Unternehmen sind gegenüber Kunden immer ehrlich, wenn es um die Nutzung der Informationen geht, die ich bereitstellen würde.

Risk Beliefs:

1. Es wäre im Allgemeinen riskant, Online Unternehmen Informationen anzuvertrauen.
2. Es wäre mit einem großen Verlustrisiko verbunden, Online-Firmen Informationen anzuvertrauen.
3. Es wäre mit zu großer Unsicherheit verbunden, Online-Firmen Informationen anzuvertrauen.
4. Online-Firmen Informationen zur Verfügung zu stellen, würde viele unerwartete Probleme mit sich bringen.
5. Ich würde mich sicher fühlen, Online Unternehmen Informationen anzuvertrauen.

Collection:

1. Ich mache mir für gewöhnlich Gedanken darüber, wenn Unternehmen mich nach meinen persönlichen Informationen fragen.
2. Ich denke manchmal zweimal darüber nach, meine persönlichen Daten auszuhändigen, wenn Unternehmen mich danach fragen.
3. Es stört mich, meine persönlichen Informationen an so viele Unternehmen weiterzugeben.
4. Ich mache mir Sorgen, dass Unternehmen zu viele persönliche Daten von mir sammeln.

Awareness:

1. Unternehmen, die online Informationen sammeln, sollten bekanntgeben, wie die Daten gesammelt, verarbeitet und genutzt werden.
2. Eine gute Online-Verbraucherdatenschutzrichtlinie sollte über eine klare und deutliche Offenlegung verfügen.
3. Es ist mir sehr wichtig, gut darüber informiert zu sein, wie meine persönlichen Informationen genutzt werden.

Control:

1. Online-Verbraucherdatenschutz ist im Grunde eine Frage des Rechts der Verbraucher, Kontrolle und Autonomie über Entscheidungen darüber auszuüben, wie ihre Informationen gesammelt, genutzt und verbreitet werden.
2. Im Mittelpunkt des Verbraucherdatenschutzes steht die Kontrolle der Verbraucher über ihre persönlichen Informationen.
3. Ich glaube, dass die Online-Privatsphäre verletzt wird, wenn die Kontrolle verloren geht oder gegen den eigenen Willen infolge einer Geschäftsabwicklung verringert wird.

C.2.2. Questionnaire on Self-reported Privacy Behavior

This questionnaire was used by Zimmerman et al. (2019a), and then translated to German by the author. Since it did not measure a specific construct, and answers

were mostly evaluated in a descriptive way, it was not deemed necessary to follow the normally recommended approach of translation and retranslation (see e.g. Harborth & Pape, 2019, for this approach). An additional item was added, which asked for the frequency of use of the Ecosia search engine. It takes a middle ground with respect to privacy, and was used by multiple participants in the study, some of which had asked during the recruitment phase whether they would be able to use this search engine during the study. The German translation is below, while the original English questions can be found at `https://github.com/stevenzim/sigir-2019`. They are the questions AQ3-7 and HQ2-3.

1. Welche der folgende Dinge tun Sie? Bitte wählen Sie alle zutreffenden Möglichkeiten aus.

   Choices were (multiple choices possible):

   - Anonyme Kommunikationsnetzwerke verwenden (z.B. Tor)
   - Ende-zu-Ende-Verschlüsselung beim Nachrichtenaustausch verwenden (z.B. Signal)
   - Erweiterungen von Drittanbietern nutzen, um Tracking-Cookies oder Skripte zu blockieren (z.B. Ghostery, Privacy Badger, Disconnect)
   - Beim Betrachten von Informationen im Internet virtuelle private Netzwerke (VPNs) verwenden
   - Software verwenden, um Browser-Fingerprinting/Browser-Identifizierung zu verhindern (z.B. uBlock, Privacy Badger)
   - Ihre Browsing Cookies automatisch (durch Verwendung von Software) löschen
   - Software verwenden, um HTTPS Kommunikation mit Webseiten sicherzustellen item Javascript in Ihrem Browser deaktivieren
   - Cookies in Ihrem Browser deaktivieren
   - Anti-Virus-Software auf Ihren Geräten installiert haben (z.B. Norton, Sophos)
   - Keine der obengenannten

2. Wie häufig nutzen Sie die folgenden Web-Browser?

   Possible choices were “Nie”, “Monatlich”, “Wöchentlich”, “Täglich”

   - Chrome
   - Internet Explorer
   - Edge
   - Safari
   - Mozilla Firefox
   - Tor
   - Brave

3. Falls Sie in der letzten Frage für alle der gelisteten Browser ”nie” ausgewählt haben, welche(n) Web-Browser benutzen Sie dann?

4. Wie häufig nutzen Sie die folgenden Suchmaschinen?

   Possible choices were “Nie”, “Monatlich”, “Wöchentlich”, “Täglich”

   - Google
   - Bing
   - Yahoo
   - Baidu
   - Yandex
   - Duck Duck Go
   - Qwant
   - Ecosia

5. Falls Sie in der letzten Frage für alle der gelisteten Suchmaschinen ”nie” ausgewählt haben, welche(n) Suchmaschine(n) benutzen Sie dann?