The Dilemma of User Engagement in Privacy Notices: Effects of Interaction Modes and Habituation on User Attention

FARZANEH KAREGAR, JOHN SÖREN PETTERSSON, and SIMONE FISCHER-HÜBNER,
Karlstad University, Sweden

Privacy notices and consent forms are the means of conveying privacy policy information to users. In Europe, a valid consent needs to be confirmed by a clear affirmative action. Despite previous research, it is not yet clear whether user engagement with consent forms via different types of interactions for confirming consent may play a significant role in effectively drawing user attention to the content, even after repeated exposure. We investigate, in a laboratory study, how different types of interactions that engage users with consent forms differ in terms of their effectiveness, efficiency, and user satisfaction. In addition, we examine if and how habituation affects user attention and satisfaction, and the time they spend on giving their consent. We conducted a controlled experiment with 80 participants in four different groups where people either were engaged actively with the policy content via Drag and Drop (DAD), Swipe, or Checkboxes, or were not actively engaged with the content (as the control condition) in a first-exposure phase and in a habituation phase. We measured user attention to consent forms along multiple dimensions, including direct, objective measurements and indirect, self-reported measures. Our results show that the different types of interactions may affect user attention to certain parts of policy information. In particular, the DAD action results in significantly more user attention to the data items compared to other groups. However, with repeated exposure to consent forms, the difference disappears. We conclude that user engagement with policy content needs to be designed with care, so that attention to substantial policy information is increased and not negatively affected.

CCS Concepts: • **Security and privacy** → **Social aspects of security and privacy; Usability in security and privacy; Privacy protections**; • **Human-centered computing** → **User studies; Graphical user interfaces; Laboratory experiments**;

Additional Key Words and Phrases: Affirmative actions, attention to policy information, habituation, informed consent, privacy notices

ACM Reference format:

Farzaneh Karegar, John Sören Pettersson, and Simone Fischer-Hübner. 2020. The Dilemma of User Engagement in Privacy Notices: Effects of Interaction Modes and Habituation on User Attention. *ACM Trans. Priv. Secur.* 23, 1, Article 5 (February 2020), 38 pages.
<https://doi.org/10.1145/3372296>

The work is supported by European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 653454.

Authors' addresses: F. Karegar and S. Fischer-Hübner, Department of Computer Science, Karlstad University, Universitetsgatan 2, 65188 Karlstad, Sweden; emails: {farzaneh.karegar, simone.fischer-huebner}@kau.se; J. S. Pettersson, Information Systems, Karlstad University, Universitetsgatan 2, 65188 Karlstad, Sweden; email: john_soren.pettersson@kau.se.

This work is licensed under a Creative Commons Attribution International 4.0 License.

© 2020 Copyright held by the owner/author(s).

2471-2566/2020/02-ART5

<https://doi.org/10.1145/3372296>

1 INTRODUCTION

Given the number of online services users access through different channels, the numerous digital trails they leave, and the amount of data processed by service providers, it comes as no surprise that researchers are actively seeking to devise guidelines and solutions for protecting users' data privacy. The de facto way of informing users and providing the opportunity for them to keep control over their data is through privacy notices. Privacy notices, generally speaking, are the means of conveying privacy policy information to users.¹ These notices come in various forms, from general privacy policies and consent forms, to cookie consents, permission dialogues in the context of Identity Providers (IdPs) and mobile applications, and privacy setting notifications.

However, there are various issues surrounding privacy notices and consent forms, confirming their ineffectiveness to fulfill the intended goals, i.e., attracting user attention, informing them, helping them to comprehend data practices, and finally keeping control over their data. Privacy notices, especially privacy policy forms, include requirements from different stakeholders [48, 49]. Users expect to receive clear, simple information about data practices and privacy controls. However, service providers employ privacy policies to demonstrate compliance with legal requirements. Consequently, privacy notices usually provide users with a surfeit of information conveyed in long, jargon-filled texts that are difficult to comprehend and ignored by users [14, 36, 43]. Notice fatigue and habituation, a lack of choice, and decoupled notices are other reasons why users do not pay attention to privacy notices [48, 49] or do not comprehend them fully but nonetheless consent to use different services [9]. The problems with privacy notices, in addition, influence service providers that rely on informed consent as the legal basis for processing personal data, because they cannot fully depend on the consent they obtain [50].

One approach to solving the problems related to privacy notices and consent forms is to make incremental improvements to the User Interfaces (UIs) of notices to help users to pay attention, which may consequently help them to get informed and make better decisions. Many researchers have focused on the improvements that can be achieved by modifying the content of privacy notices, i.e., modifying the information provided to users and how to provide the information. Researchers have proposed privacy nutrition labels [33], multi-layered short notices summarising key data practices [5, 6, 22], personalized privacy notices [24, 59], privacy icons and images [16, 26], and comic-based interfaces to convey policy information [54]. However, the proposed methods all come with their own new challenges. For example, providing information in a compact form may impair transparency [37]. In addition, due to the abstract and complicated nature of the concept of privacy, the meaning of icons often needs to be learned, thus requiring user education [41].

To solve the problems related to privacy notices, however, some researchers took another approach and studied improvements to the ways in which users interact with the content of notices, for example, by actively involving them with the content using Drag and Drop (DAD) [32], Checkboxes [58], and Swiping [12]. Literature not focusing on privacy also supports the notion that different interaction techniques can influence the number of user actions and user memory of the content [53].

Nonetheless, none of the studies that suggested some solutions for the problems of privacy notices measured user attention by both direct and indirect measures, which complement each other. In addition, to the best of our knowledge, there is no study that investigates whether user engagement with the consent form content via different types of interactions may play a significant role in effectively and efficiently drawing user attention to the content compared to situations

¹The terms "privacy notices" and "privacy policies" are usually used interchangeably. In this article, *privacy policies* refer to the exact information that should be provided to users and *privacy notices* refer to different channels and means through which the information is communicated to users.

lacking active user engagement, even after repeated exposure. The enforcement of new privacy rules and regulations such as the EU General Data Protection Regulation (GDPR) [38] and their emphasis on informed consent, which should be given by clear affirmative actions, also augment the importance of investigating if and how different interaction design solutions, which can serve as affirmative actions in consent forms, are effective in helping users to pay attention when they provide their consent.

In this article, we therefore address the first research question whether different types of interactions that engage users with the consent forms differ in terms of their effectiveness in drawing user attention to the content, efficiency, and user satisfaction. In addition, we address the second research question if and how habituation, i.e., diminished attention or decreased response as a result of repeated exposure to the consent forms, affects user attention and satisfaction, and the time they need to spend on giving their consent.

To this end, we conducted a laboratory experiment with 80 participants in four different groups in two phases, namely a first-exposure phase and an overexposure phase (habituation phase), to study the potential effects on informed consent. Each phase was a between-subject study that included a control group in which users were not actively involved in the content of the consent forms. In other words, users in the control group just clicked “Accept” or “Reject” to proceed, which served as the only form of active interaction in the form. The other three groups included consent forms utilizing *checkboxes* as common practices on some services, *DAD* and *swiping actions*, which are derived, designed, and adapted based on a literature review and previous works. In the Checkbox, DAD, and Swipe groups, the interaction types involved active user engagement with the content to make choices (without such choices, it was impossible to proceed) and then clicking a button to proceed. We measured user attention to consent forms along multiple dimensions, including direct, objective measurements and indirect, self-reported measures, using eye-tracking methods and recall of information, respectively. Our results show that different types of interactions can create some biases in user attention to different parts of policy information. Therefore, if engagement in privacy notices is chosen for increasing user attention to the content, then this needs to be carefully designed to direct user attention to substantial policy information.

The article proceeds as follows. In Section 2, we provide some background information on the concepts used in the study. Section 3 presents a review of related literature. In Section 4, we explain the four groups and the designed interfaces in detail. We discuss our methodology and the experimental design in Section 5, while we report the results of the study in Section 6. Section 7 includes a discussion of the results and Section 8 concludes the article with design recommendations and future work.

2 BACKGROUND

In this section, we first exemplify the legal requirements of transparency and consent, by referring to the GDPR, which mandate designers to consider the Human-Computer Interaction (HCI) implications of them. Then, we explain habituation and its characteristics. The concepts introduced in this section have guided our research and the design of our study.

2.1 Consent and Ex-ante Transparency

Ex-ante transparency concerns the information that should be provided to users before collecting any of their personal data and it is a precondition for consent. Several Recitals and Articles of the GDPR (e.g., Article 13 and Recital 42) and Guidelines of Article 29 Working Party on Consent [4] shed light on the policy information, which should be provided to users before they give their consent. We refer readers to legal literature on privacy for in-depth discussions of the GDPR regulations [20, 27].

Pursuant to Article 6 (1) GDPR, the processing of personal data is lawful if one of the six legal grounds of data processing, including individual consent, applies. The individuals' intention of agreeing to the processing of their personal data, given by a statement or a clear affirmative action, is considered consent if it is unambiguous, specific, informed and freely given (Article 4 (11) GDPR). Ticking a box, choosing technical settings or any other statements that clearly show the individuals' acceptance of the processing of their personal data are valid affirmative actions, according to Recital 32 GDPR. Therefore, the GDPR does not recognize implicit or opt-out consent as valid. Specifically, silence, pre-selected checkboxes or inactivity are inadequate for inferring consent, according to Recital 32 GDPR.

Despite the fact that the GDPR mandates that policy information has to be provided in a "concise, transparent, intelligible, and easily accessible form, using clear and plain language" (Article 12 (1)), it does not elaborate on what constitutes an effective consent for users to make informed decisions. In other words, the GDPR takes a technology-neutral approach and only specifies the legal requirements of consent, which, if appropriately obtained, can help users to retain control of their data and protect their privacy. However, how the legal requirements should be mapped to usable user interfaces for achieving informed consent is not thoroughly addressed.

Similar to the arguments by Schaub et al. [49] regarding the role of privacy notices and choice, we believe that the main problem is not inherent in the requirements of control and transparency but in how we currently design notices and choice.

2.2 Habituation, Eye-tracking, and User Attention

Habituation is a "decreased response to repeated stimulation" [56]. In this article, *habituation* refers to the response decrements (e.g., diminished attention) to repeated consent forms. Habituation occurs if the stimulus and an existing mental model of the stimulus are similar. If the stimulus is new, then sensitization occurs, which increases one's response to the stimulus [23]. Stimulus specificity and response recovery (dishabituation) are two of the characteristics of habituation [44, 56]. Response recovery is the recovery of the habituated response either after a period of time or following the presentation of a novel stimulus. Stimulus specificity predicts responding to a novel stimulus while dishabituation predicts an increase in response to the original stimulus when the novel stimulus is removed [44].

In HCI, habituation occurs, for example, when people are repeatedly exposed to warning messages as they interact with various computer programs on different devices. It is difficult to observe habituation, a mental state, with conventional methods such as warning adherence [12]. Some research, therefore, directly measures habituation using functional magnetic resonance imaging (fMRI) [2] or eye-tracking [1]. Information derived from eye-tracking to measure human visual activities complements other sources of data that cannot be obtained through subjective, self-reported measurements [18]. Eye-tracking, in addition, helps researchers to understand to what parts of privacy and security notices participants pay attention and has been used in the literature to investigate, for example, user attention to permission dialogues [3, 21]. User attention is a prerequisite for the perception and understanding of notices.

One of the neurophysiological manifestations of habituation is an unconscious decrease in scrutinizing previously seen stimuli compared to novel stimuli, termed the Eye Movement-based Memory (EMM) effect [47]. The EMM effect is evident in fewer eye-gaze fixations and less visual sampling of the areas of interest within the visual stimulus [1]. For the first time, based on the EMM effect, Anderson et al. [1] used eye movements as the direct measure of habituation to security warnings. They managed to demonstrate the decrease in users' visual sampling with successive viewings of warnings.

As habituation develops over time, Vance et al. [57] conducted longitudinal experiments on the habituation to security warnings using fMRI and eye-tracking. Their findings indicate that cross-sectional habituation studies are valid proxies for longitudinal studies and eye-tracking is a valid measure of the mental process of habituation to warnings.

Consequently, in our study, which is a cross-sectional study conducted in the laboratory, we used eye-tracking to measure user attention and as a manifestation of habituation.

3 RELATED WORK

This section briefly discusses literature on active user involvement in privacy notices to improve user attention and habituation studies conducted for warning messages.

3.1 Active User Involvement in Privacy Notices

In the context of permission dialogues of IdPs, Wang et al. [58] were some of the pioneers suggesting giving users control over their data by deselecting what they do not want to share using checkboxes. Wang et al. [58] provided pre-selected checkboxes for requested personal data and the possibility to opt out of the optional ones in their proposed interfaces. Pre-selected checkboxes do not require a clear affirmative action and thus do not constitute a valid consent (see Recital 32 of the GDPR). Wang et al. observed that people who used the checkbox-enabled interfaces released significantly less information in total and made use of granular choices to opt out from certain data collection.

Javed et al. [29] integrated animation in the permission dialogues of Facebook and compared animated dialogues with dialogues using checkboxes and a control dialogue showing the content in plain text. They used the eye-tracking method as a direct, objective measure of attention. Javed et al. showed higher fixation counts and longer fixation duration on the animated dialogues. The participants authorised significantly lower number of permissions for animated dialogues compared to two other types of dialogues in Reference [29]. Javed et al. showed that the animated dialogue was able to switch and maintain user attention towards the content. However, the robustness of the proposed animated dialogues against repeated exposure remains an open issue.

Karegar et al. [32] studied users' recall of personal information shared using checkboxes in the permission dialogues of an IdP. However, their study does not show to what extent checkbox-enabled dialogues could be more effective in helping users to recall what they share and make informed consent compared to control dialogues, which convey information in plain text.

Karegar et al. [31], in another study, adapted DAD to the context of IdPs for selecting personal information. They showed that using DAD significantly helped users to recall the personal information they shared compared to the Facebook permission dialogues in which everything is pre-selected by default, with checkboxes enabling users to opt out of disclosing some of the requested personal data (if users click on a link in the dialogue). However, more investigation is needed for direct comparison of checkboxes and other methods like the DAD, because only a few users clicked on the link to edit the selected information in Reference [31]. The concept of DAD was first suggested by Reference [42] to address the problem of habituation to which Just-In-Time-Click-Through Agreements (JITCTAs) [40] are vulnerable. However, its effectiveness was not evaluated more than in a preference test where the DAD was used in a more metaphorical graphical user interface than ordinary pop-up dialogues [8].

Recently, Lindegren et al. [35] proposed and compared interfaces in which three different types of active selection of personal information, i.e., Swiping, Checkboxes, and DAD, are utilized in the context of permission dialogues of IdPs on mobile devices. No significant difference between the users' recall values of what personal information they shared were observed, although the recall rate for Swiping and DAD actions were slightly higher. Moreover, it is not clear if the slight

improvement in recall values for Swiping and DAD methods was the result of the novelty of methods for users, increased user attention (which could be measured by an eye-tracker device) due to the type of selection they experienced, or other possible personal differences between individuals related to privacy. The implemented Swiping and DAD actions in Lindegren's et al. study are different from how we utilized the DAD and Swipe actions in our interfaces (see Section 4).

Karegar et al. [31, 32] and Lindegren et al. [35] measured user attention using indirect, subjective measures of recalling personal information shared by participants without ascertain if and how participants paid attention to different items in the permission dialogues. In addition, the effectiveness of their proposed designs is possibly vulnerable to effects of repeated exposure.

In sum, contrary to the previous works, we investigate and compare the effectiveness of different design solutions that actively involve users in privacy notices by using both indirect and direct measures of user attention. We also investigate if and how the effectiveness of different design solutions changes under repeated exposure to consent forms while taking some steps in the direction of ecological validity in our study to avoid having the same limitations that can be found in the existing literature on habituation to notifications (see Section 3.2 for more details).

3.2 Habituation to Notices

Recently, Reeder et al. [46] conducted an experience sampling study of web browser security warnings with more than 6,000 users and reported that the majority of their participants showed strong evidence of not being habituated to warning messages. Nonetheless, quite an extensive body of research focuses on habituation as a frequently cited factor influencing the effectiveness of warnings and investigates how users habituate to security warnings and possible solutions [1, 2, 11, 12, 30, 57].

Bravo-Lillo et al. in two studies [11, 12] investigated habituation to security warnings that had visual attractors, including inhibitive attractors. Inhibitive attractors prevent a user from making a potentially dangerous choice until either some period of time has expired or the user performs a required action, i.e., gets involved in the dialogues. Bravo-Lillo et al. measured habituation indirectly in terms of the percentage of users who immediately recognized (clicked the "No" button to finish the study) that the salient field in the dialogue changed after a habituation period. They used attractors to highlight a field that was of no value during habituation trials but contained critical information after the habituation period.

In 2013, Bravo-Lillo et al. showed that all inhibitive attractors outperformed the control condition at a high habituation level, although they incurred delays to users' workflows [12]. In 2014, they demonstrated that some of the high-performing attractors, including swiping over the salient fields (Swipe) or typing the important information in a text field (Type), actually became less effective with additional exposure [11]. One of the limitations of the way in which they measured habituation is that they had no means to control if users who did not immediately click "No" to finish the study after the habituation trial did not really read the salient field.

Although studies on habituation in the context of security warnings abound, the study by Javed et al. [30] is, to the best of our knowledge, the only study of habituation to privacy notices for permission dialogues. Javed et al. built on the experiment by Bravo-Lillo et al. and leveraged an eye-tracker device to force users to pay attention to information requested in Facebook permission dialogues [30]. Participants who experienced the eye-tracker enabled permission dialogues paid significantly more attention to the permissions they gave, as measured by comparing the number of eye-gaze fixations between groups. Nonetheless, Javed et al. could not find any significant difference between groups based on the rate at which participants denied a dangerous permission. In addition, Javed et al. did not report on the efficiency of the eye-tracker enabled permissions or on user satisfaction [30] compared to that of the control group.

Fig. 1. DAD prototype.

Compared to studies conducted by Bravo-Lillo et al. [12] and Javed et al. [30] on habituation to notices, our study, despite also being cross-sectional, takes some further steps in the direction of ecological validity. First, we used both direct and indirect measures of user attention contrary to the Bravo-Lillo's et al. study. Further, we did not speculate without any strong clue that what we, as researchers, assume to be sensitive and critical information might also be regarded as sensitive by users who should, therefore, reject its sharing, something that served as a manifestation of robustness against habituation in the both studies. Additionally, in both studies users were continuously and uninterruptedly exposed to permission dialogues or security warnings without having a primary task as a motivation or a moment for response recovery (see Section 2.2), which is not consistent with what happens in real-world situations. Participants in our study had the primary task of rating photos. During our habituation trial, rating a photo after accepting a dialogue better simulates real situations. The details of our study design are described in Section 5.

4 USER INTERFACES IN OUR STUDY

This section first presents the common layout for all four groups of consent forms, i.e., permission dialogues, which we designed and tested for the purpose of our study. Then we describe each of the proposed types of selection: (i) DAD, see Figure 1, (ii) Swipe, shown in Figure 2, and (iii) Checkboxes, depicted in Figure 3. Figure 3 also shows the Area of Interests (AOIs) defined in the study (see Section 5.3) on checkbox consent forms as an example. The Control permission dialogue is the fourth group, presented in Figure 4.

4.1 General Layout

According to Schaub et al. [49], a privacy notice should only include the information and control options most relevant to the specific audience at the time of displaying the notice to avoid long notices. This can be achieved by using multi-layered privacy notices, as recommended by Article 29 Working Party [5, 6], and details-on-demand patterns [51]. The top layer may include initial notices pointing to further information in other layers. Nonetheless, it is crucial to decide wisely on the

Fig. 2. Swiping prototype.

Fig. 3. Checkbox prototype with defined AOIs.

inclusion of the information conveyed in the first layer, as a failure at this stage may hide essential information [37].

Therefore, in all four of our designs, the first layer includes the most relevant policy information, which needs to be presented based on the GDPR transparency requirements and which is relevant to the audience in the context of social networks acting as IdPs when they decide about sharing their data [31]. The policy information in the first layer includes four different AOIs consisting of the identity of the controller, the personal data requested, the data processing purposes, and some additional policy information about the conditions of consent (see Figure 3). Accordingly, each design is structured in three main sections (see Figures 1 to 4): a top section for conveying the identity of the data recipient, a middle section for showing what data are requested for which

The screenshot shows a web browser window with a "Log in with Google" button. Below it is the TinyPic logo and contact information: Website: <http://tinypic.com/> and Email Address: contact@tinypic.com. The main heading reads "requests to access the following information from your Google account:". Under "Mandatory Information", there are two sections: "1. Email address:" with a list of purposes (update policy, send best picture) and "2. Birthday:" with a list (verify age, send birthday wishes). At the bottom, a section titled "If you accept the requests:" lists terms of use. At the very bottom are "REJECT REQUEST" and "ACCEPT REQUEST" buttons.

Fig. 4. Control prototype.

processing purposes (with the opportunity to select and decide), and finally some additional policy information in the bottom part, above the Reject/Accept buttons. The content of the consent forms for all sections differs within the groups, except for additional policy information, but is the same between groups. Additional policy information is repeated identically in each consent form. The policy information² at the bottom should have pointed to further information, i.e., to the full policy information, but we removed the link due to the challenges a click on the link would pose to our eye-tracker data collection and our analysis. Moreover, our focus was not on if users click to see further information. Karegar et al. [32], in their study on authorisation dialogues of IdPs, reported that of 60 participants only one clicked on the full privacy policy link.

The middle section, which includes the AOIs for data requested and purposes of processing, is the part that differs in structure (not content) in our four designs. The middle sections differ in the way of how users can get involved in the dialogues to control their data and select what they want to share, depending on the types of interactions deployed. Users' willingness to grant access to their data depends on the purposes associated with the request for access (for example, shown in Reference [34]). In all of our designs except the Control group (see Figure 4), the requested information is divided into mandatory and optional information. In real scenarios, services may request some information required to provide the service (mandatory information) and other information that, for example, may be used for other purposes such as marketing (optional information). Even mandatory information could be used for other purposes than providing the necessary services, such as marketing or offering extra services to users. Therefore, our consent forms were designed to serve a broader range of service providers' requirements and to have opt-in choices for different purposes as required for a valid, specific consent.

4.2 The Three Types of Interactions Tested in this Study

DAD. The first interaction type is a new version of DAD. We adapted the suggestion by Pettersson et al. [42] and the DAD used in Reference [31]. In our design, personal data requested to be shared are provided as text in separate rectangles that users can drag and then drop to one or more special

²For convenience sake, from now on whenever we refer to the policy information, policy recalling, or policy questions we mean the *additional* policy information in the bottom part of consent forms.

places tagged with data processing purposes (see Figure 1). In other words, users select where each draggable item should be dropped among available options.

The DAD we utilized is different from that tested in the studies by Lindegren et al. [35] and Karegar et al. [31]. Lindegren et al. [35], who tested their designs on a touch screen device, assume that each requested piece of information, represented in text accompanied by a relevant icon, had to be dragged to the only dedicated place for that piece of information. In other words, each draggable item in Reference [35] had a single area to be dropped, providing no other option. Karegar et al. used DAD on desktop and their participants had to drag the items individually and drop them to a common, single destination place implemented for all data requested [31]. We avoided this type of DAD design, because users may ignore paying attention to what they drag and the meaning of the place where they drop it and instead just execute the action, i.e., dragging and dropping the item, especially when they are not provided with different options to select among.

Regardless of its accessibility limitation, it does not stray too far from reality by considering all users familiar with DAD actions. Both in desktop systems and mobile devices, users often do tasks requiring them to use DAD functionality.

Swipe. The second interaction type is the *Swipe* action. The *Swipe* action we integrated into our dialogues (see Figure 2) is different from the *Swipe* action Lindegren et al. [35] used on mobile devices; in their study, the *Swipe* action is specific for touchscreen devices, although also achievable by a mouse, and it does not necessarily involve users with the text. A quick left-to-right *Swipe* action in Reference [35], which starts from the requested data item moves that data item from the left part of its dedicated place to the right part to show that it has been selected to be shared. On touch devices, *swipe/slide* involves placing your finger on the screen and moving it in any direction. For example, we may need to slide or *swipe* a finger from left-to-right or right-to-left to unlock the phone.

As discussed in Section 3, Bravo-Lillo et al. showed that *Swipe* and *Type* are the best attractors [12]. Although in the context of security warnings the *Type* attractor may help, we excluded it as it would not fit in the context of permission dialogues of IdPs. In permission dialogues of IdPs, we have several salient fields and it is error-prone and time-consuming if we expect users to type every single piece of salient information. The *Swipe* action, as defined by Bravo-Lillo et al., involves moving the mouse to highlight text. In our user interfaces, we slightly adapted the *Swipe* in Reference [12] by adding an arrow and combining the *swipe* action with a slider. A slider is a graphical control element with which a user may set a value by moving an indicator, but our slider was without values. Our slider-like design recognizes only the horizontal component of users' movement and was chosen to give users more control over data selection compared to standard text highlighting with a mouse, and to make it easier in the future to utilize this interaction type on touch devices.

Checkbox. The third type of interaction is the *Checkbox*. It was included to be the representative of current common practices on some websites, and it is similar to the checkboxes used in References [29, 58], except that nothing is pre-selected by default in compliance with the GDPR (see Figure 3).

5 METHOD

5.1 Research Questions

The research questions of this article are as follows:

RQ1: Do different types of interactions actively involving users with consent forms differ in their effectiveness in terms of attracting user attention to the content, efficiency, and user satisfaction?

Our three null hypotheses related to RQ1 (H_{1_0}) are that user *attention* to the content ($H_{1_0} - \text{attention}$), their *satisfaction* ($H_{1_0} - \text{satisfaction}$), and the *time* it takes to handle consent forms ($H_{1_0} - \text{time}$) do not differ based on the different types of interactions in such consent forms.

RQ2: If and how does habituation, i.e., repeated exposure to consent forms with integrated interaction types, affect user attention, efficiency, and their satisfaction? Our three null hypotheses related to RQ2 (H_{2_0}) are that after repeated exposure to different types of interactions, user *attention* to the content ($H_{2_0} - \text{attention}$), their *satisfaction* ($H_{2_0} - \text{satisfaction}$), and the *time* it takes to handle consent forms ($H_{2_0} - \text{time}$) do not differ between the groups.

5.2 Study Design

To answer our research questions, we conducted a between-subject user study with four groups of permission dialogues: (1) a DAD group (see Figure 1), (2) a Swipe group (see Figure 2), (3) a Check-box group (see Figure 3), and (4) a Control group in which people were not actively involved with the content, as depicted in Figure 4. The functional user interfaces of the study run on Chrome browsers and were designed using Bootstrap and JQuery Javascript Library. We wanted our participants to act as they would act in real-world scenarios and feel connected with the information requested in the dialogues, to the greatest extent attainable. However, it would be hard to use participants' mobile phones/laptops to conduct the study, for example, in the context of permission dialogues of mobile apps or cookie banners, as we needed extra time and resources for setting-up the study and the eye tracker for each participant with her own device. Consequently, we chose the permission dialogues of social networks, just as an example, and we used the laboratory computer but asked users to sign into their own social network accounts (see Cover story). However, the generality of the results should not be affected as we discuss it later in Section 7.

Cover story. To answer our research questions, we first needed to measure user attention. However, we did not want to prime our participants for privacy and attention to the content of the consent forms we designed; i.e., we wanted to avoid non-realistic attention due to the study set-up. Privacy and handling privacy notices are users' secondary goals. Therefore, to conduct the study in a more ecologically valid way, we needed a story engaging users with a primary task and with which we could avoid the users' focusing on the primary research objective of this study until the debriefing session. The real purpose of our study was carefully obscured while meeting principles of ethical research (see Section 5.5) during the recruitment of participants and when obtaining their consent to participate in the study.

We introduced our study as an eye-tracking photo-rating study in which users had to rate some photos based on the hashtags assigned to them while their eye movements would be tracked. We selected a series of real photo websites³ and a series of photos.⁴ We told the participants that their task in the study was to rate as many photos as possible. Participants would rate one photo for each website if they signed up for the website. Nonetheless, we assured them that signing up for a website is their own choice and refusing to sign up would not affect our study. They had two sign-up options: (i) using their Facebook account, or (ii) using their Google account. The moderator showed a printed version of Facebook and Google login buttons and asked participants if they had seen them before and used them. If they had not used any social media login buttons before, then the moderator provided a brief description.

The permission dialogues they received when signing up for websites using their Facebook or Google accounts were the consent forms we designed. There was no communication behind the

³We did not use well-known websites, to avoid the effects of familiarity with the websites on user behaviour. However, we wanted to use real names so that we could claim that these were existing photo websites.

⁴Derived from <https://pixabay.com>.

Fig. 5. Study design.

scenes with Facebook or Google on sharing user data. Even though we did not directly inform users of the lack of connection between the dialogues and the social networks, we did not claim the opposite either. No party, including us, accessed users' personal data. We acknowledge that our participants might have assumed that there was a communication behind the scene with Facebook or Google. While we could have set up the study to involve data transfers with Google and Facebook, we decided to follow the privacy principle of data minimisation.

Figure 5 provides an overview of the study design and the collected data types. Our user study consisted of two phases: (i) Phase 1 designed to address RQ1, and (ii) Phase 2 conducted to contribute to RQ2. Each phase is accompanied by a related post-test questionnaire about what users experienced in that phase. In the following, we describe the different steps of the study in detail.

Welcome Session. Participants took part in the study individually and voluntarily. The moderator first introduced all steps of the study to participants and asked them to read the paper consent form. This paper consent form should not be confused with the consent forms we designed and tested in our study. The general objective of the study was stated in the consent form and explained orally as analysing the different user activities (including their eye movements) between different groups of participants. The paper consent forms were written in compliance with the GDPR and local privacy regulations. Users were informed about their rights, e.g., to revoke their consent.

Then, in the pre-test questionnaire, participants answered four demographic questions (age range, gender, field of study, and the highest educational degree) and one question on the kind of personal information they regard as sensitive and usually do not want to share with service providers, especially non-anonymously. We used the answer to this question to customize the information some websites requested in their consent forms in the study (see the description of Phase 1). After the pre-test questionnaire, participants were asked to sign into either their Facebook or Google accounts. Although they could have rejected to sign in or left the study whenever they wanted while still being compensated, it did not happen for any of the participants.

People may have several accounts for the same service provider for different purposes. For example, one main account to use on a daily basis (i.e., a real account) and one account for other less serious purposes or for receiving advertising emails or promotions (i.e., a fake account). All of our participants had a Facebook or Google account, because during the recruitment we mentioned it as an eligibility requirement. We did not oblige our participants to use their main accounts. However, during the closing session, we asked them if they used their main accounts (see Section 5.4 for exact numbers).

Eye-tracking set-up. After signing in, the moderator helped participants to calibrate the eye-tracker. A five-point calibration was performed and if there was an error, we recalibrated after repositioning the participant and checking for factors that could be interfering with pupil detection such as droopy eyelids or type of glasses. However, for some participants the eye-tracker could not be calibrated properly. We made a note and excluded them ($n = 3$) from the analysis afterwards.

Participants' eye movements were collected and analysed using a Tobii Pro X3-120 Eye-Tracker and Tobii Studio software. Tobii X3-120 is a remote eye-tracker with a sampling rate of 120 Hz, which we installed with adhesive brackets on a 24-inch monitor, providing unobtrusive eye-tracking. Eye movements stable for at least 100 ms were classified as fixations. All participants had normal or corrected-to-normal vision using contact lenses or glasses. After the data collection, we set a threshold of 20% as the acceptable proportion of missing data samples and excluded three further participants (i.e., six in total) to prevent missing data from negatively affecting [10] our study.

Phase 1. Participants in all four groups were shown three photo websites and they had to click on the logo of each to handle their permission dialogues, i.e., our designed consent forms (henceforth P1D1, P1D2, and P1D3). Each website requested access to different types of personal information for different purposes. We measured the user attention to dialogues using recall values, eye-tracking data, and whether they could identify dubious purposes or very sensitive information in the consent forms (more details in Section 5.3). For this last measurement, we needed to include specific and critical data requests in the dialogues to see if users would pay attention to them. Therefore, the mandatory information requested by the first website included access to data for dubious, privacy-invasive data processing purposes. For example, it requested access to users' email accounts, either to read the content of their emails or to send randomised photos to their friends from their email. The second website requested access to sensitive personal data derived from the results of the pre-test questionnaire. We listed eight sensitive data items, such as debit/credit card information, passport number, and medical issues. We also gave users the option to indicate that they regarded none of the data items on the list as sensitive. If nothing was selected as sensitive information from the list, then we used *medical issues* as the default value to be shown in corresponding consent forms. Finally, the third website neither requested sensitive information nor to process data for dubious purposes. If users clicked the "Accept" button in a dialogue, then they would see a photo and were able to rate it; otherwise, they continued to the next step.

In the DAD and Swipe groups, users first saw a tutorial when they clicked on the first website. DAD and Swiping are new concepts for users in this context, so we designed short, simple, animated tutorials illustrating the actions for selection to decrease the time participants might spend on figuring out how to select rather than what to select.

The post-test questionnaire of Phase 1 started with a question on what information users shared with each of the websites if they accepted their requests (P1Q1). Participants had a list of data items and for each, they had five options: not shared, probably not shared, not sure, probably shared, and shared. The questionnaire then continued with two questions on the accepted data processing purposes in the consent form of the last website in Phase 1, i.e., P1D3 (if users accepted its request). The third website requested the user's email address and birthday as mandatory information. In two multiple-choice questions, with 6 options (P1Q2 and P1Q3) each, users had to identify the purposes for which they had disclosed their data. We included three true/false questions and one multiple-choice policy question related to the information we provided at the bottom of each dialogue (P1Q4) to see if users had paid attention and understood. The questionnaire concluded with two 10-point Likert scale questions to subjectively measure how users feel about the time they spent handling and selecting information in each consent form (P1Q5) and its level of difficulty (P1Q6). Participants answered whether handling a dialogue and selecting information/purposes of data processing, overall, took them too much time (point one on the 10-point Likert scale) or too little time (point 10). Likewise, they answered if handling a dialogue and selecting information/purposes of data processing were very difficult (point one) or very easy (point 10).

Phase 2. After Phase 1, participants might have become biased towards paying attention to dialogues because of the questions they had received.⁵ Therefore, Phase 2 tests if and how repeated exposure to consent forms affects user attention in different groups when they intend paying attention from the start. Note that we tried to decrease the gap between our simulated scenario in the lab and real scenarios by having rest periods or a moment for response recovery. Rating photos was the primary task of the simulated scenario, which also served as the rest period after visiting each consent form (the stimulus).

We did not have a benchmark indicating how many consent forms users should see and handle to become habituated. In addition, no benchmark can realistically be set, due to the differences between people and between the consent forms in each group. Furthermore, showing users many consent forms continuously may exhaust or bore rather than habituate them, which could contribute significantly to their lack of attention. Andersson et al. [2] showed that habituation sets in after only a few exposures to a warning and progresses rapidly with further repetitions.

Consequently, in Phase 2, our participants had to handle the consent forms of 15 websites: 12 of these (P2D1–P2D12) belonging to the habituation trial and three (P2D13–P2D15) added for measuring attention after the trial (test session). Nothing prominent could help participants to distinguish between the trial and the rest sessions.

We ensured that the content of the consent forms during the habituation trial differed regarding the data sharing requests, the stated purposes of processing, and the identities of controllers, thus better resembling user experiences in real scenarios. For example, P2D1 requested one mandatory item while P2D2 and P2D3 requested two other mandatory items with one and two optional items, in that order. This pattern was repeated four times for every three consecutive consent forms in the habituation trial. The last three dialogues in Phase 2 (the test session of Phase 2) requested the same number of data items (two mandatory items plus one optional) but with different contents.

⁵This was confirmed in the debriefing session when we asked participants if they thought that they had paid more attention to dialogues in Phase 2 because of the questions in Phase 1.

The first consent form after the habituation trial included dubious data processing purposes. The website requested access to the full names of participants to either tag their names on offensive pictures or randomly upload offensive pictures under their names. The second website after the habituation trial included the same sensitive information as in Phase 1, and finally the last one neither had dubious purposes nor sensitive information.

The first question in the post-test questionnaire for Phase 2 (P2Q1) was similar to the first question in Phase 1 but was on the last website participants handled in this phase (P2D15). This was followed by a question on the data processing purposes accepted in P2D15 and including six true/false sub-questions (P2Q2). We then repeated the same policy questions (P2Q3) and self-measured questions on the time it took to handle a dialogue in Phase 2 (P2Q4) and the level of difficulty (P2Q5), similar to those asked after Phase 1. The Phase 2 questionnaire was followed by a SUS questionnaire [13]. The SUS questionnaire is a subjective measurement of usability providing ten five-point Likert scale questions. The ten questions lead to a SUS score between 0 to 100.

Closing session. At the end of the study session, participants were requested to sign out from their social network accounts and were debriefed on the actual primary purpose of the study. Participants stated that they were not at all bothered by the cover story when they realized what the primary purpose of the study had been. However, very few ($n = 5$) participants stated that they would have done better if they were explicitly told to pay attention to the content of dialogues. We clarified that we wanted to create a real-life situation, where dealing with privacy notices is usually a secondary task, and we were interested in their natural behaviour, which our test users then acknowledged. The majority of participants remarked that they learned a lot through the cover story and will be more aware of privacy and pay more attention to privacy notices in future, even if they are engaged in other activities and may not be explicitly asked to consider privacy.

During the study, the moderator took notes on user reactions to the consent forms in Phase 1 and Phase 2 (excluding the trial session). After the study, the moderator asked participants why they accepted or rejected the dialogues in Phase 1 and the last three dialogues in Phase 2. We wanted to make sure that rejection/acceptance was an informed decision and that they noticed critical fields in the dialogues. Participants were also asked to indicate if they had used their real or fake accounts. The moderator then rewarded participants (with a lunch coupon) and thanked them. To meet ethical rules, we made it clear to participants from the beginning that they would receive the lunch coupon even if they had withdrawn from the study during or after the test.

5.3 Measures

To answer our two research questions in Section 5.1, we used both indirect and direct measures for analysing user attention to dialogues. By indirect measures, we mean self-reported, subjective measures of information recall and the notice adherence⁶ as the manifestation of attention to the dialogues. However, recall scores may be affected by other factors such as memory capacity; recall of specific items visited for a short time is also vulnerable to guesswork. Therefore, in addition to indirect measures, we used eye-tracking to measure the behavioural manifestation of attention (i.e., eye movements) to dialogues to complement and validate our indirect measures. Table 1 shows an overview of the variables we defined in our study to measure user attention, efficiency, and satisfaction. In the following, we elaborate on how each variable is defined and calculated.

Information recall includes recalling what is shared (i.e., data recall), for which purposes (i.e., purpose recall), and under which conditions (i.e., policy recall) in both phases.

Data recall is calculated based on how users answered P1Q1 in Phase 1 and P2Q1 in Phase 2. From the data items listed in the questions, we included just the data items that were actually

⁶If users notice the sensitive information or dubious purposes of data processing in consent forms.

Table 1. Overview of the Variables Defined to Measure Effectiveness, Efficiency, and Satisfaction

Variable name	Definition
Effectiveness: indirect measure of user attention (information recall)	
Data recall	
DRP1/P2	The score calculated for data recall in P1D3/P2D15
DRP1/P2-uncertainty	The uncertainty score calculated for recalling personal data in P1D3/P2D15
Purpose recall	
P1/P2Purpose	The score calculated for recalling data processing purposes in P1D3/P2D15
Policy recall	
P1/P2oR	The score calculated for recalling policy information in Phase 1/Phase 2
Notice adherence	
P1D1/D2NoN	The binary value calculated to show if the dubious purposes are noticed in P1D1/P1D2
P2D13/D14NoN	The binary value calculated to show if the sensitive information is noticed in P2D13/P2D14
Effectiveness: direct measure of user attention (eye-tracking measures)	
Fixation counts	The numbers of fixation counts calculated for each AOI in every consent form
Average fixation duration	The numbers of average fixation duration calculated for each AOI in every consent form
Efficiency	
AverageTP1/P2	The average time spent for a consent form in Phase 1/Phase2
P1/P2PerceivedTimeSpent	Subjective opinion on the time spent in each consent form in Phase 1/Phase 2
Satisfaction	
SUS	The calculated score for the system usability scale (SUS)
P1/P2PerceivedEaseofUse	Subjective opinion on perceived ease of use for a consent form in Phase 1/Phase 2

requested in the dialogue (four data items for P1D3 and three for P2D15) to calculate participants' data recall scores. Then we defined four variables: DRP1, DRP1-uncertainty, DRP2, and DRP2-uncertainty. Data Recall for Phase 1 (DRP1), for example, shows the number of data items the participant correctly mentioned that were shared/not shared in the last dialogue of Phase 1. If a data item had been shared and the participant said it had probably been shared, then we considered it an uncertain answer, while we considered it a wrong answer if the participant said the same item had probably not been shared. The same holds for conditions where the data items were not shared. All answers selecting the mid-point scale, i.e., NOT SURE, were also deemed uncertain answers. Therefore, DRP1-uncertainty, for example, shows the total number of uncertain answers for the four data items requested in the dialogue.

Purpose recall in Phase 1 was calculated based on P1Q2 and P1Q3 (see the description of Phase 1 for more details). P1Purpose is the variable defined to show purpose recall in Phase 1, namely the average of two scores a participant received for P1Q2 and P1Q3. For each correctly selected/not selected option, of a total of six options each for P1Q2 and P1Q3, participants received one point. In Phase 2, P2Purpose was calculated based on the answers to P2Q2 (see the description of Phase 2 for more details) and shows purpose recall. Each correct answer to the six sub-questions of P2Q2 gave the participant one point for P2Q2.

Policy recall was in both phases calculated based on the same four policy questions repeated in the post-questionnaires (P1Q4 and P2Q3). Each correct answer to one of the four questions gave one point and in total a participant could receive a maximum score of four for policy recall in each phase (P1oR in Phase 1 and P2oR in Phase 2).

Notice adherence is the extent to which users make informed decisions, or give informed rejections or acceptances of the dialogue, based on its content.⁷ We defined specific binary variables (P1D1NoN, P1D2NoN, P2D13NoN, and P2D14NoN, where the “NoN” stands for “Noticed or Not noticed”) for two dialogues in Phase 1 (P1D1 and P1D2, see the description of Phase 1) and two dialogues (P2D13 and P2D14, see the description of Phase 2) in Phase 2, which included critical contents. The binary variables show if the participant had paid attention to the critical fields.

A manifestation of habituation is an unconscious decrease in scrutinizing previously seen stimuli compared to novel stimuli, evident in fewer eye-gaze fixations and less visual sampling of the areas of interest within the visual stimulus [47]. Thus, we used eye-gaze fixation counts to measure the decrease in attention. Besides the number of fixation counts, we also measured the average fixation duration. It is important to consider these two **eye-tracker measures** separately to get a complete picture of how a participant processes the stimulus [52, 55]. The correlation between fixation counts and fixation duration may not exist in particular studies. Literature shows that, based on the context, average fixation duration may be indicative of the complexity of information, user engagement, cognitive effort, and the complexity of the visual content of the performed task [45]. The number of fixations, however, is an appropriate indicator of the importance of stimuli [15, 25]. However, the context of use plays a significant role in drawing a conclusion on eye-tracker data. For example, longer tasks tend to require more fixations [28], or an AOI with a higher fixation count but a lower average fixation duration can indicate search behaviour and how little understanding participants have of the presented information [17].

For each dialogue, we defined some AOIs and measured the number of eye-gaze fixations and average eye-gaze duration for each AOI. Our AOIs in each dialogue (see Figure 3) include the AOI for: (i) the identity of a controller at the top of the dialogue (for simplicity called “Title” in Figure 6), (ii) the data processing purposes in the middle, (iii) the information requested in the middle, and iv) the additional policy information at the bottom of the dialogue. The AOIs for the data processing purposes and the information requested in the DAD, Swipe, and Checkbox groups only included mandatory data items and their corresponding processing purposes to enable fair comparison with the Control group, which lacks optional items.

Efficiency is measured by timing our participants. We defined the time it takes to handle a permission dialogue as the difference between when it is first seen and when the requested permission is given/denied by the participant. We calculated the average time it took to handle a consent form in each group in Phase 1 and Phase 2 as AverageTP1 and AverageTP2 variables, respectively. Additionally, participants’ subjective opinions on the time they spent handling a single dialogue in each phase (e.g., P1PerceivedTimeSpent variable derived from P1Q5, see the detailed description of Phase 1 and 2) complemented the objective time measurement.

Satisfaction is measured by collecting our participants’ subjective opinions on the perceived ease of use for a dialogue in each phase (e.g., P1PerceivedEaseofUse variable derived from P1Q6). In addition, the SUS variable shows the calculated SUS score for each participant.

To see the effects of interaction types, we compared the defined variables between different groups in both Phase 1 and Phase 2. We also analysed how some variables changed in each group before repeated exposure to dialogues (in Phase 1) and after (in Phase 2). The results are reported in Section 6.

⁷Informed decision means that the participant has paid attention to critical fields in the dialogue and understands them regardless of accepting or rejecting the dialogue.

5.4 Participants and Demographics

We recruited our participants by posting on Facebook pages related to our university, sending the advertisement to some mailing lists, and distributing paper flyers to all departments of our University except Computer Science and Information Systems to avoid biases. We did not limit our recruitment by only asking for student participants. The only eligibility requirements of the study were age (+18) and possession of Facebook or Google accounts. In addition, we asked our colleagues (who were not involved in the project) to distribute the advertisements to their families and friends. Participants were assigned to one of the four groups of the study, G1 to G4 in that order, and on a first-come, first-served basis.

Table 2 summarises the demographics. In total 86 people participated in our study. We excluded the data collected from six of the participants (see Section 5.2, Eye-tracking set-up, for more details), which gave us 20 participants in each group, among them, only five were from Computer Science-related fields such as Computer Science, Computer Engineering, or Information Systems. Most of our participants had backgrounds in the social sciences ($n = 25$) and applied sciences ($n = 21$). We had 53 female and 26 male participants while one preferred not to reveal their gender. Most of our participants were quite young, with 45 in the 18–25 age group and 20 between the ages of 26 and 35, which explains why most of them ($n = 54$) were bachelor students or had their bachelor degrees.

Most of the participants used their real accounts ($n = 76$) and had previously used social login buttons to sign up for websites ($n = 75$). On average, the study took almost 28 minutes, excluding the debriefing session, per participant. The total time reported in Table 2 is the time to complete the whole study including the time for answering the questionnaires and calibrating the eye tracker.

5.5 Ethics

According to the Swedish Ethical Review Act (SFS 2003:460) any research involving humans that includes sensitive personal information (as also defined by Article 9 GDPR), involving physical encroachment on a subject, or using a method that aims or risks to affect the subject physically or psychologically, and studies on biological material traceable to specific individuals, must be vetted by an ethical review board. Our study does not fall under the Act as we did not collect any sensitive personal data and did not use any methods to affect subjects psychologically. To respect users' privacy, we asked demographic questions on a high level of granularity (e.g., age ranges and fields of study at a group level) that was sufficient for our study rather than going into details (following the principle of data minimisation). We obtained informed consent in compliance with the GDPR, did not request users' actual or sensitive personal data, and stated the correct study objective in general terms while did not claim anything about the study's purpose or how it would be conducted that would have been incorrect.

As discussed in Section 5.2, participants might have assumed that there was a communication between their Google/Facebook accounts and the websites in the study, even though this was never stated. Therefore, it could be debated whether the study could be deemed as mildly deceptive. However, we strictly followed the "Ethical Principles of Psychologists and Code of Conduct" by APA [7], to which also the Swedish Research Council refers, that apply for deception in users studies: The study design could not cause severe emotional distress or physical pain. Moreover, our research questions could not have been adequately addressed by other means while at the same time respecting the legal data minimisation principle. In addition, during the closing session participants were fully debriefed about further details of the study objectives, the lack of data transfers between their social accounts and the websites and the reasons for it, as well as the right to withdraw their consent.

Table 2. Demographic Information

Demographic info	DAD (<i>n</i> = 20)	Swiping (<i>n</i> = 20)	Checkbox (<i>n</i> = 20)	Control (<i>n</i> = 20)	Total (<i>n</i> = 80)
Gender					
Female	14	12	14	13	53
Male	5	8	6	7	26
Preferred not to say	1	0	0	0	1
Age					
18–25	12	11	12	10	45
26–35	5	7	5	3	20
36–45	0	2	1	6	9
46–55	3	0	2	1	6
56+	0	0	0	0	0
Highest education level (currently pursuing or received the degree)					
High School	1	0	0	0	1
Bachelor Degree	10	8	13	13	54
Master Degree	6	7	4	5	22
Doctorate Degree	3	5	3	2	13
Field of study					
Formal Science	1	1	1	1	4
Computer-related	1	2	2	0	5
Social Science	4	6	7	8	25
Natural Science	5	4	3	2	14
Applied Science	5	4	5	7	21
Humanities	3	3	2	2	10
None	1	0	0	0	0
Total time to complete the study (in minutes)					
Min	19.00	14.00	19.00	18.00	14.00
Max	44.00	40.00	39.00	51.00	51.00
Mean	29.55	28.15	27.27	27.50	28.11
Standard Deviation	7.15	7.20	5.02	8.36	6.96
If used social login before					
No	1	0	2	2	5
Yes	19	20	18	18	75
Type of account					
Real, daily life	20	20	17	19	76
Fake	0	0	3	1	4
Social login selected					
Google	10	8	7	11	36
Facebook	10	12	13	9	44

6 RESULTS

In this section, we report the results of indirect measures of attention to consent forms in Phase 1 and Phase 2 and compare them with the results concerning direct measures of user attention. We also report the results of satisfaction and efficiency. Table 3 shows an overview of the results of indirect measures of attention and the results achieved in both phases for satisfaction and efficiency. For the summary of the variables defined in our study, we refer the reader to Table 1.

Table 3. Overview of the Results of Satisfaction, Efficiency, and Indirect Measures of Attention

RQ1 - Phase 1			RQ2 - Phase 2		
$H_{10} - attention$: Not rejected			$H_{20} - attention$: Not rejected		
Dependent Variable	Test	Significant result	Dependent Variable	Test	Significant result
Data recall					
DRP1	Kruskal-Wallis H test	YES. The DAD group recalled much better what they shared/not shared compared to the Control group.	DRP2	Kruskal-Wallis H test	NO
DRP1-uncertainty	Kruskal-Wallis H test	NO	DRP2-uncertainty	Kruskal-Wallis H test	NO
Purpose recall					
P1Purpose	Kruskal-Wallis H test	NO	P2Purpose	Kruskal-Wallis H test	NO
Policy recall					
Between-subject comparison					
P1oR	Kruskal-Wallis H test	NO	P2oR	Kruskal-Wallis H test	NO
Within-subject comparison					
Exact Sign Test: a statistically significant median increase in the policy recall score in Phase 2 compared to Phase 1.					
Notice adherence					
Between-subject comparison: dubious data processing purposes					
P1D1NoN	Chi Square Test of Independence	YES. The DAD group were more likely not to notice the dubious purposes while the Checkbox group were more likely to notice them.	P2D13NoN	Fisher's Exact Test of Independence	NO
Within-subject comparison: dubious data processing purposes					
Exact McNemar's test: No significant difference in the proportion of participants who paid attention to the dubious data processing purposes in the pre and post habituation trial.					
Between-subject comparison: sensitive information					
P1D2NoN	Fisher's Exact Test of Independence	NO	P2D14NoN	Fisher's Exact Test of Independence	NO
Within-subject comparison: sensitive information					
Exact McNemar's test: No significant difference in the proportion of participants who paid attention to the sensitive information in the pre and post habituation trial.					
$H_{10} - time$: Rejected			$H_{20} - time$: Not rejected		
P1Perceived-TimeSpent	Kruskal-Wallis H test	YES. The Control group perceived that they had spent much less time than the Checkbox group, who perceived to have spent a lot of time.	P2Perceived-TimeSpent	Kruskal-Wallis H test	YES. The perceived time spent in the Checkbox group is significantly higher compared to all other groups.
averageTP1	Kruskal-Wallis H test	YES. The average time of the Control group differed significantly from the average time spent by each of the other groups.	averageTP2	Kruskal-Wallis H test	NO
$H_{10} - satisfaction$: Not rejected			$H_{20} - satisfaction$: Not rejected		
P1Perceived-EaseofUse	Kruskal-Wallis H test	NO	P2Perceived-EaseofUse	Kruskal-Wallis H test	NO
Overall satisfaction					
Dependent Variable		Test		Significant result	
SUS		One way ANOVA		NO

Fig. 6. The average of fixation counts on different AOIs in each consent form in four groups.

6.1 Phase 1

On average, in all groups in Phase 1, the most fixation counts were on data processing purposes (see Figure 6), showing that data processing purposes were the most engaging part for users. Between-group comparison in Phase 1, however, shows that although similar in the DAD, Swipe, and Checkbox groups, fixation counts on the data processing purposes are noticeably lower in the Control group, as depicted in Figure 7. On the whole, participants found the policy information, data items, and the identity of the controller almost equally important and notably *less* important than the data processing purposes in Phase 1. Fixations counts on policy information, data items and the identity of the controller are almost equal for all consent forms, except DAD (see Figure 6). For DAD consent forms, there is a noticeable difference between the average fixation duration and fixation counts on AOIs for data items and for policy information and the identity of the controller.

Data recall. To test whether data recall, measured by DRP1 variable (range 0 to 4), differed based on the type of consent forms, we used non-parametric tests, since the assumptions of normality and homogeneity of variances were violated. A Kruskal-Wallis H test showed that there is a significant difference in data recall scores between the four groups studied, $\chi^2(3) = 8.447, p = 0.038$, with a mean rank of 48.90 for DAD, 41.93 for Swiping, 42.53 for Checkbox, and 28.65 for the Control group. The Dunn-Bonferroni post hoc comparison test, i.e., Dunn's pairwise multiple-comparison procedure [19] with adjustment using Bonferroni corrections, revealed that participants who experienced the DAD consent forms ($M = 2.35, SD = 1.27$) were significantly better at recalling what information they shared compared to the Control group that did not actively select the data ($M = 1.20, SD = 1.20$). There was no evidence of a difference between the other pairs.

(a) Fixation counts

(b) Average of fixation duration

Fig. 7. Fixation on AOIs for data processing purposes: The average of fixation duration and fixation counts in each consent form in each group.

(a) Fixation counts

(b) Average of fixation duration

Fig. 8. Fixation on AOIs for data requested: The average of fixation duration and fixation counts in each consent form in each group.

We also compared the uncertainty level in participants' data recall, measured by DRP1-uncertainty variable, between different groups. Although participants in the DAD ($M = 0.90$, $SD = 1.12$) group were more certain about what they had shared/not shared there was no evidence of a statistically significant difference ($\chi^2(3) = 3.776$, $p = 0.287$) between the DAD, Swipe ($M = 1.20$, $SD = 1.20$), Checkbox ($M = 1.30$, $SD = 1.22$), and Control groups ($M = 1.55$, $SD = 1.15$).

Figure 8 shows that in P1D3, the Control consent forms, compared to those of the other groups, did not succeed at attracting user attention to data items, as participants had fewer gaze fixations, each of which was much shorter on average than the average fixation duration in other groups. The average fixation duration in P1D3 is similar between the consent forms that involved users with dialogues; however, users fixated much more on data items in the DAD group compared to the three other groups, which helped them to recall significantly better what personal information they had shared with the service provider.

Purpose recall. To test whether purpose recall, measured by the P1purpose variable, differed based on the type of consent forms users experience, we conducted a Kruskal-Wallis H test. The result did not show any significant difference between groups. However, participants who experienced Swipe ($M = 4.92$, $SD = 1.50$) and Checkbox ($M = 4.75$, $SD = 1.53$) consent forms could better recall the purposes for which they shared their data with the last website in Phase 1 (P1D3) compared to those who handled DAD consent forms ($M = 4.45$, $SD = 1.82$) and the Control group ($M = 4.30$, $SD = 1.63$). Consequently, although DAD outweighed other interaction types and the

Fig. 9. Fixation on AOIs for additional policy information: The average of fixation duration and fixation counts in each consent form in each group.

Table 4. The Number of Participants in Each Group Who Noticed or Did Not Notice the Critical Contents and Who Accepted or Rejected the Requests

	Including dubious purpose								Including sensitive information							
	P1D1				P2D13				P1D2				P2D14			
	N		DN		N		DN		N		DN		N		DN	
	R	A	R	A	R	A	R	A	R	A	R	A	R	A	R	A
DAD	7	2	2	9	14	1	0	5	11	7	0	2	17	2	0	1
Swipe	6	11	0	3	17	1	0	2	5	13	0	2	12	8	0	0
Checkbox	13	7	0	0	15	2	2	1	11	8	0	1	16	2	0	2
Control	8	3	0	9	10	1	2	7	10	3	0	7	17	0	0	3

N: Noticed. DN: Did not notice. R: Rejected. A: Accepted.

Control group in terms of attracting user attention to the data items they shared, the DAD interaction was not more effective in terms of conveying data processing purposes.

Participants who experienced DAD, Swiping, and Checkbox consent forms had longer and more fixations on AOIs for data processing purposes in P1D3 compared to the Control group (see Figure 7). However, more and longer fixations did not lead to a significant difference between the groups regarding their purpose recall scores.

Policy recall. Although participants in our Control ($M = 1.65, SD = 1.10$) and Checkbox groups ($M = 1.50, SD = 1.15$) answered more policy questions correctly than those in the Swipe ($M = 1.4, SD = 1.30$) and DAD groups ($M = 1.3, SD = 0.92$), the difference was not statistically significant in a Kruskal-Wallis H test with P1oR defined as the dependent variable. Figure 9 shows that the DAD group had the least number of fixations on policy information in Phase 1. The average fixation count on policy information in Phase 1 for DAD was 4.2 for all the three dialogues, with 8.13 for Swipe, 10.43 for Checkbox, and 9.33 for Control, which is aligned with slightly better policy scores in Control, Checkbox, and Swipe compared to the DAD group.

Notice adherence. The variables P1D1NoN, P1D2NoN, P2D13NoN, and P2D14NoN (see Section 5.3) show if the participant paid attention to the critical fields in the consent forms. We wanted to analyse if attention to critical fields is dependent on the type of consent forms participants experienced. Table 4 shows the total number of participants in each group who noticed or did not notice the sensitive information or dubious purposes. Additionally, based on Table 4, we can see how many of the participants did not notice the critical fields and accepted the dialogues that

impose higher privacy risks compared to those who did not notice the critical content but rejected the dialogues.

The total number of participants who did not pay attention to the dubious purposes of data processing in Phase 1 is highest in the DAD group ($n = 11$), followed by the Control group ($n = 9$). A Chi Square Test of Independence showed significant evidence of association ($\chi^2(3) = 19.222, P < 0.001$) between whether users paid attention to dubious purposes and the type of the consent forms. A post hoc test, based on p values associated with adjusted standardized residuals and the Bonferroni corrected p value ($p = 0.05/8 = 0.00625$), showed that the number of participants who noticed or failed to notice dubious data processing purposes in Phase 1 differed significantly in the DAD ($p = 0.003^8$) and Checkbox groups ($p = 0.001^8$). Participants in the DAD group were more likely not to notice the dubious purposes and accept the dialogue while those in the Checkbox group were more likely to notice the dubious purposes.

As depicted in Figure 7, for Dialogue 1 in Phase 1 (P1D1), the eye-gaze fixation counts on the data processing purposes were much higher for the consent forms that actively involved users with content than the Control group. However, Figure 7 shows that the average eye-gaze fixation duration on the data processing purposes did not differ noticeably across all groups. Surprisingly, we concluded from indirect measurements and statistical tests that participants who experienced DAD consent forms, and not the Control group, were more likely to miss the dubious data processing purpose in Phase 1, while the Checkbox group were more likely to spot it. An equal number of participants failed to notice the critical fields of P1D1 but accepted the request in the DAD and Control groups. More fixation counts in the DAD group on the data processing purposes, each of which on average almost equalled the average duration of fixations in the Control group, could be representative of the more complex structure in DAD for extracting information on data processing purposes, thus causing users to focus less on reading the content.

In total, more participants paid attention to the sensitive information ($n = 68$) than to the dubious purposes ($n = 57$) in Phase 1. The result of a Fisher's Exact Test of Independence⁹ did not give us enough evidence that there is an association between the four groups and the attention to the sensitive information with the probability $p = 0.071$ for the distribution in Table 4. However, the total number of participants who did not pay attention to the sensitive information in Phase 1 (P1D2) is higher in the Control group ($n = 7$) than in the other groups, where only one or two people failed to notice the sensitive information.

Figure 8 shows that the Control group had the lowest fixation counts in P1D2 on data items, compared to the other groups, and the shortest average fixation duration on data items, similar to the Checkbox group. Consequently, the Control group had the lowest total fixation duration in P1D2, which explains the higher number of participants in this group who failed to notice the sensitive information.

Efficiency and satisfaction. Figure 10 shows how the average time spent on handling dialogues differed between each group in Phase 1 and also between the three consent forms within each group. As depicted in Figure 10, the time spent on consent forms decreased in all the four groups in Phase 1 as participants became more familiar with such forms. However, the time spent did not decrease at the same ratio level in all groups. For consent forms involving users actively with content, the average Time Decrease Ratio (TDR)¹⁰ was higher ($TDR - DAD = 37.17\%$,

⁸The p value estimated for adjusted standardized residuals in the contingency table.

⁹We used Fisher's Exact Test of Independence instead of the Chi-Square Test of Independence, because more than 20% of cells had expected frequencies < 5 .

¹⁰The decrease ratio percentage between time a and time b is calculated as $(abs(a - b)/a) * 100$.

Fig. 10. The average time spent on each dialogue in both phases in different groups.

Table 5. The Average Time (seconds) and Standard Deviation for Each Consent Form in Different Groups in Phase 1

	P1D1		P1D2		P1D3		AverageTP1	
	M	SD	M	SD	M	SD	M	SD
DAD	92.98	36.41	67.78	47.88	35.76	20.74	65.50	29.26
Swipe	86.32	47.24	54.99	23.87	40.00	21.73	60.44	22.70
Checkbox	82.57	45.73	48.98	24.03	46.94	27.99	59.50	23.26
Control	26.40	15.89	25.71	23.68	16.32	16.61	22.81	14.91

$TDR - Swipe = 31.77\%$, $TDR - Checkbox = 22.43\%$) than the average TDR for the Control group ($TDR - Control = 19.57\%$).

On average, it took 52.06 seconds ($SD = 28.42$) for each participant to handle a single dialogue in Phase 1. As evident in Table 5, the Control group spent much less time ($M = 22.81$, $SD = 14.91$) than the other groups. To test if the average time to handle a dialogue differed based on the four groups of consent forms, we conducted a Kruskal-Wallis H test, which showed that there is a significant difference in the time spent on a dialogue, measured by the AverageTP1 variable in Table 5, based on the four different consent forms, $\chi^2(3) = 31.621$, $p < 0.001$. A Dunn-Bonferroni post hoc comparison test revealed that the average time of the Control group differed significantly from the average time spent by each of the other groups (adjusted $p < 0.001$). The type of user engagement did not affect the time it took to handle a dialogue, but a lack of active user engagement significantly decreased the time.

Figure 11 shows participants' subjective opinion about the time they spent handling a single dialogue and how easy it was to handle a single dialogue in Phase 1 (see Section 5.3). The higher the score, the easier they perceived handling a consent form to be and the less time they perceived to have spent on a single dialogue. Are there statistically significant differences between the four groups regarding participants' scores for perceived ease of use and the time spent? The dependent variables (P1PerceivedEaseofUse, P1PerceivedTimeSpent) should be treated as ordinal variables, because one cannot meaningfully claim that the interval between the data points is even across the 10-point scales. Based on the results of a Kruskal-Wallis H test ($\chi^2(3) = 5.420$, $p = 0.144$), we conclude that participants' perceived "ease of use" did not differ based on the specific consent form they experienced.

Nonetheless, a similar test showed evidence of difference in users' perception of the time they spent based on the four different consent forms, $\chi^2(3) = 9.511$, $p = 0.023$, with a mean rank of 43.55

Fig. 11. Participants' perception of the time spent and ease of use on a dialogue in Phase 1. The numbers on the bars show how many participants gave a particular score in each group.

for DAD, 39.95 for Swipe, 28.35 for Checkbox, and 50.15 for the Control group. A Dunn-Bonferroni post hoc comparison test revealed that the Control group perceived that they had spent much less than the Checkbox group, who perceived to have spent a lot of time (adjusted $p = 0.016$). However, there was no evidence of a difference between the other combinations of the four groups.

6.2 Phase 2 and the Effects of Repeated Exposure

Overall in the habituation trial (P2D1 to P2D12), the time spent on handling consent forms in each group (shown in Figure 10) and the number of fixation counts (see Figure 6) on the AOIs both decreased, although there is some fluctuation between consent forms due to the different content of the three consecutive consent forms in each group (see Phase 2 in Section 5). Less visual scrutiny or sampling of regions of consent forms indicates a decrease in response to a stimuli, i.e., habituation. Does this necessarily mean that users habituated to consent forms in a way that negatively affected or endangered their privacy? This section helps us to answer this question and RQ2. Here, we report the results of indirect measures of attention to dialogues in Phase 2 and compare them with the results concerning direct measures of user attention. In addition, we report the calculated efficiency and satisfaction in Phase 2 and compare them with the results of Phase 1. Furthermore, we analyse the effects of repeated exposure using different measures.

Contrary to Phase 1, Control group participants did not differ noticeably from others in terms of the time they spent on each consent form and their fixations on data processing purposes in the habituation trial (see Figures 7 and 10). Although decreased during the habituation trial, users fixated the most on the data processing purposes in each group in a manner similar to Phase 1. Nonetheless, the difference between fixations on the data processing purposes and fixations on other consent form contents decreased with more exposure to dialogues, since the change ratio is not the same for all AOIs (see Figure 6).

Repeated exposure to consent forms did not affect the fact that fixation counts on AOIs for data items were higher in the DAD group than the others (see Figure 8). Fixation counts on data items in the habituation trial were more similar in the Swipe, Checkbox, and Control groups if compared to Phase 1, as shown in Figure 8. In the Checkbox and Swipe groups, similar to Phase 1, user fixations on the AOIs for the identity of the controller, data items, and policy information were almost the same. The Control group, however, fixated more on policy information during the habituation trial, both compared to Phase 1 and to the other groups.

At the beginning of Phase 2, our participants were biased towards paying attention to consent forms compared to the beginning of Phase 1 as they admitted in the closing phase of the study (see Section 5). This was shown through more fixations on policy information than on the purposes of processing or data items in Phase 2.

Data recall. As depicted in Figure 8, there was a big difference between the average number of fixation counts on data items in the DAD group ($M = 14.5$) and the Control group ($M = 6.4$) in P2D15. Nonetheless and contrary to Phase 1, a Kruskal-Wallis H test of the DRP2 dependent variable (range 0 to 3) did not show a significant difference in data recall scores between the four groups in Phase 2 ($\chi^2(3) = 0.364, p = 0.948$). After repeated exposure to consent forms, the significant effect of the type of consent form on the difference between data recall values disappeared. However, participants who experienced the consent forms with DAD ($M = 1.75, SD = 1.07$), Swipe ($M = 1.80, SD = 0.89$), and Checkboxes ($M = 1.70, SD = 0.98$) were slightly better at recalling what information they shared with the service provider compared to the Control group ($M = 1.60, SD = 1.05$).

Users in the DAD and Swiping groups were more certain about what they had shared/not shared, both in Phase 1 and after repeated exposure to consent forms in Phase 2, measured by DRP2-uncertainty. Nevertheless, similar to Phase 1, there was not a statistically significant difference regarding uncertainty ($\chi^2(3) = 1.266, p = 0.737$) between the DAD ($M = 0.85, SD = 1.09$), Swipe ($M = 0.95, SD = 0.76$), Checkbox ($M = 0.95, SD = 0.94$), and Control groups ($M = 1.10, SD = 0.97$) in Phase 2.

We compared the eye-tracker measures for P1D3 and P2D15 within the groups. As depicted in Figure 8, the average fixation duration between P1D3 and P2D15 did not change noticeably within the four groups (almost no change in the DAD and Control groups and less than 30 ms (<12%) decrease in the other groups), while the average number of fixation counts decreased around 40% in the DAD group, and increased around 24% in the Swipe, 20% in the Checkbox, and 65% in the Control groups. Consequently, the decrease in the total duration participants fixated on data items in the DAD group and the increase in the Control group contributed to more homogeneous data recall scores among groups in Phase 2.

Purpose recall. In Phase 2, similar to Phase 1, participants who experienced Swipe ($M = 3.80, SD = 1.15$) and Checkbox ($M = 3.55, SD = 1.99$) consent forms could better recall the data processing purposes in P2D15, measured by P2Purpose variable, than the Control ($M = 3.30, SD = 1.75$) and DAD groups ($M = 2.80, SD = 1.74$), although the differences were not statistically significant. Nonetheless, in all groups, on average, participants answered fewer questions correctly (of the six questions, see Section 5) on the purposes for which they shared their data in Phase 2 compared to Phase 1, due to repeated exposure to dialogues.

The distribution of fixation counts on data processing purposes in P2D15 (see Figure 8) was not the same across the groups ($\chi^2(3) = 10.028, p = 0.018$). The DAD group fixated much less on data processing purposes with the mean rank of 33.30 than the Checkbox group with a mean rank of 54.45. Although this significant difference did not lead to a significant difference between the purpose recall scores of the groups, it showed that users in the DAD group had more difficulty remembering the data processing purposes.

Policy recall. Similar to Phase 1, the Control group ($M = 2.80, SD = 1.11$) and the Checkbox group ($M = 2.75, SD = 1.41$) on average answered more questions correctly concerning policy information than the Swipe ($M = 2.25, SD = 1.52$) and DAD groups ($M = 2.20, SD = 1.28$), but the difference between the scores of different groups (defined by P2oR variable) was not deemed statistically significant in a Kruskal-Wallis H test ($\chi^2(3) = 3.494, p = 0.322$). However, Figure 9 shows that the Control group fixated more and for longer on policy information, especially compared to the DAD group.

As we asked participants the exact same policy questions in the Phase 1 and Phase 2 questionnaires, the policy recall scores (P1oR and P2oR) can serve as a dependent variable for a within-subject test to see whether there was a significant median difference in policy recall scores before

and after repeated exposure. Did participants pay more attention to policy information in Phase 2 after receiving some questions on this in Phase 1? An exact Sign Test¹¹ showed a statistically significant median increase (1 correct policy question) in the policy recall score in Phase 2 compared to Phase 1 ($Z = 5.131, p < 0.001$), which again shows that the intention to pay more attention to dialogues in Phase 2 affected the attention to policy information more than other content.

Notice adherence. As shown in Table 4, the total number of participants who did not pay attention to the dubious data processing purposes for data requested in P2D13 is highest ($n = 9$) in the Control group, followed by the number of DAD group participants who accepted the dialogue but did not notice the purposes ($n = 5$). A Fisher's Exact Test of Independence⁹ did not show evidence of a significant association between noticing the dubious purposes and the type of consent forms in Phase 2 ($p = 0.06$). Contrary to the results of Phase 1, participants using DAD consent forms were not more likely to miss dubious purposes, and the Checkbox group did not differ significantly from the others in attracting user attention to dubious purposes after repeated exposure. Eye-tracking data for P2D13, as depicted in Figure 7, show a possible explanation to why more participants in DAD and Control groups missed the dubious purposes than in the other two groups. In P2D13, the average fixation counts in DAD and Control groups are smaller than in the Swipe and Checkbox groups. In addition, in P2D13, the DAD group has the smallest average fixation duration while the average fixation duration in other groups are quite similar to each other. Consequently, participants in DAD and Control groups had smaller total fixation duration on data processing purposes compared to the other groups.

In total, similar to Phase 1, more participants paid attention to the sensitive information ($n = 74$) than to the dubious purposes ($n = 61$) in Phase 2. Both numbers increased slightly compared to Phase 1. A Fisher's Exact Test of Independence⁹ did not show a statistically significant relationship between the type of consent forms and attention to sensitive information ($p = 0.499$). The total number of participants who did not pay attention to the sensitive information in P2D14 was almost the same across the groups. The average fixation duration and the number of fixation counts on data items in P2D14 were much higher in the DAD than the Control group (see Figure 8), which clarifies why slightly fewer participants in the DAD and Swipe groups missed the sensitive information.

To determine whether the proportion of participants who noticed (did not notice) the critical content in each group differed significantly after repeated exposure to consent forms, we conducted Exact McNemar's tests. Exact McNemar's tests for participants in each group (with two dichotomous dependent variables between two related samples, i.e., the same participants in each group, before and after repeated exposure to dialogues) determined that there was not a statistically significant difference in the proportion of participants who paid attention to the critical fields in the pre and post habituation trial ($p > 0.05$ for all tests). Some performed slightly better, such as noticing dubious data processing purposes in the DAD group (from 11 who did not notice to 5) and noticing sensitive information in the Control group (from 7 who did not notice to 3) while some performed worse, such as noticing dubious purposes in the Checkbox group (from all who noticed to 3 who did not). Despite this, the number of fixation counts on the data processing purposes and data items dramatically decreased when comparing P1D1 with P2D13 and P1D2 with P2D14 as depicted in Figures 7 and 8, respectively.

Dubious purposes requested in P1D1 and P2D13 might be against what people usually expect from service providers. Despite the opportunity for participants to express their thoughts about

¹¹The distribution of the differences between the two policy recall values before and after repeated exposure to consent forms were not symmetrical in shape.

Table 6. The Average Time (seconds) to Handle a Consent Form in Different Groups in Phase 2

Group	D1	D2	D3	D4	D5	D6	D7	D8
	D9	D10	D11	D12	D13	D14	D15	AverageTP2
DAD	35.82	24.22	24.30	14.63	17.98	14.36	9.33	11.64
	13.78	11.30	13.03	12.84	17.28	12.79	22.06	17.02
Swipe	31.66	26.54	22.85	13.20	16.20	16.89	9.95	12.57
	14.53	9.63	12.20	13.53	22.01	15.40	28.09	17.68
Checkbox	33.12	25.56	19.86	10.46	13.58	13.08	10.86	11.67
	17.89	10.61	10.74	12.73	16.80	11.85	41.84	17.38
Control	32.43	21.71	15.69	11.46	13.63	14.64	9.72	10.94
	12.18	9.12	9.35	8.05	10.27	9.43	28.01	14.44

the study, we did not directly ask them if they expected to receive such requests and if such requests changed their thoughts about the cover story. Nonetheless, we asked them to provide their reasons for accepting or rejecting such requests. Most of the participants who accepted the dubious purposes mentioned that they did not notice or pay attention to them. Interestingly, one of the participant who noticed but accepted the dubious purpose (P2D13) said that she used the website before and she trusted the website. Another participant who accepted the dubious purpose (P2D13) said that she overlooked the request, because she knew the website and she was a member of it.

Moreover, sensitive information (e.g., credit card information) requested in P1D2 and P2D14 is not a typical information request users would receive in Google/Facebook permission dialogues. Two of the participants who accepted the requests for credit card information but noticed it in the permission dialogues justified their decision by saying that they knew that they did not disclose their credit card information to Facebook/Google so they would not have access to such information even if they accepted the request.

Efficiency and satisfaction. As depicted in Figure 10, compared to the first consent form in Phase 2, participants spent less time on each of the other forms in Phase 2 except for the last one. If a participant wanted to reject the last dialogue, then a message would appear saying that the last dialogue must be accepted. This therefore added to the time they spent handling the last dialogue (see Section 5.2 for more details on the study design).

Due to the different characteristics of the consent forms participants experienced in each group, the average time spent did not decrease at a constant rate. As shown in Figure 10, the decrease in time more or less follows a pattern: After a steep decrease, it increases for two consecutive forms and then again decreases. The content of consent forms in the habituation trial was not the same for all the dialogues and it was repeated every three consent forms from P2D1 to P2D12 (see Phase 2, Section 5.2). For example, the time spent on handling a consent form from P2D6 to P2D7 first decreased and then increased in all groups for handling P2D8 and P2D9. P2D7 requested just one mandatory item while the other two requested two mandatory items with one and two optional items in that order, which explains why the time spent decreased and increased over the habituation trial.

On average, it took each participant 16.63 seconds ($SD = 6.44$) to handle a single dialogue in Phase 2, much less than the 52.06 seconds in Phase 1. Table 6 shows a detailed overview of the time participants spent on each dialogue in the different groups and in total in Phase 2. As obvious from Table 6, the Control group on average spent less time ($M = 14.44$, $SD = 7.40$) than the other groups. To test if the time spent on handling a dialogue, measured by AverageTP2 variable in

Fig. 12. Participants' perception of the time spent and ease of use of a dialogue in Phase 2. The numbers on the bars show how many participants gave a particular score in each group.

Table 6, differed based on the four groups of consent forms in Phase 2, we conducted a Kruskal-Wallis H test, which did not show strong evidence of difference with $p = 0.311$.

A comparison of the average time participants in each group spent in Phase 1 (as presented in Table 5) and the average time they spent on handling a dialogue in Phase 2 shows a significant decrease in time in all groups after Phase 2, except for the Control group.

As shown in Figures 11 and 12, the number of participants who gave a score of more than five to PerceivedEaseofUse increased from Phase 1 ($n = 39$) to Phase 2 ($n = 49$) while the number who gave a score of more than five to PerceivedTimeSpent remained more or less the same ($n = 42$ in Phase 1, $n = 44$ in Phase 2). Similar to Phase 1, we conducted a Kruskal-Wallis H test, which showed that participants' ease of use scores did not differ based on the specific consent forms they experienced ($\chi^2(3) = 1.475, p = 0.688$). A Kruskal-Wallis H test, however, showed that there is a significant difference in users' perception of the time they spent based on the specific consent forms they handled similar to Phase 1, $\chi^2(3) = 16.966, p = 0.001$, with a mean rank of 43.80 for DAD, 48.10 for Swipe, 22.48 for Checkbox, and 47.62 for the Control group. A Dunn-Bonferroni post hoc comparison test revealed that the perceived time spent in the Checkbox group is significantly higher compared to all other groups (adjusted $p < 0.05$). Contrary to Phase 1, the Checkbox group's perception of the time they spent on handling a dialogue in Phase 2 was significantly higher, even when compared to the perceived time spent in the DAD and Swipe groups.

How did the participants' perceived ease of use and the time they spent handling a dialogue change from Phase 1 to Phase 2? We performed Sign Tests in each group for the two dependent variables of PerceivedTimeSpent and PerceivedEaseofUse. The results of the Sign Tests did not show a median difference in participants' perceived time to handle a dialogue before and after the habituation trial in any group ($p_{DAD} = 1.000, p_{Swipe} = 0.143, p_{Checkbox} = 1.000, p_{Control} = 0.754$). Likewise, the results of the Sign Tests did not show a median difference in participants' perceived ease of use before and after the habituation trial in any of the groups ($p_{Swipe} = 0.180, p_{Checkbox} = 0.791, p_{Control} = 1.000$) except for the DAD group. Interestingly, the habituation trial affected users' perceived ease of use of DAD consent forms, $p_{DAD} = 0.002$. After the habituation trial, 15 of the 20 participants in the DAD group showed an increased rating for perceived ease of use. Repeated exposure and repetition of the same action of dragging data items and dropping them to desired places to select data processing purposes helped participants to regard the ease of the task in the DAD consent forms more positively.

At the end of the post-test questionnaire in Phase 2, participants answered SUS questions regarding the consent forms they experienced. The consent forms in the Control group received the highest SUS values ($M = 66.5, SD = 12.71$) compared to DAD ($M = 56.75, SD = 15.58$), Swipe ($M = 61.50, SD = 15.90$), and Checkbox ($M = 54.40, SD = 17.29$) consent forms. However, a

one-way ANOVA¹² did not show a significant effect of the type of consent forms on SUS scores ($F(3, 76) = 2.375, p = 0.077$) similar to the lack of evidence for a significant difference in perceived ease of use based on the consent forms participants experienced.

7 DISCUSSION

The privacy notices in our study were the permission dialogues in the context of identity providers. However, the methods exploited to actively involve users apply to all contexts in which users need to make a choice and agree to the processing of their personal information. Moreover, because we evaluated user attention to different policy information including requested personal data, the purposes of processing, and the conditions of consent, our results are not limited to the permission dialogues of identity providers. Instead, they can be extended to other consent forms such as the permission dialogues of browser extensions, mobile apps, cookie consent forms, and so on. Nonetheless, it is worth mentioning that depending on the context, either the results of our study for Phase 1 or the results for Phase 2 may be applicable. Sometimes, in real scenarios, we encounter privacy notices, once in a while, such as registration forms for websites or permission dialogues of IdPs, for which the results of Phase 1 could be more relevant regarding user engagement. In contrast, privacy notices sometimes appear frequently when we surf the web, such as cookie consents, for which the results of the user engagement in Phase 2 could be more relevant if the frequent privacy notices are identical in their design layouts though not in their exact content.

Phase 1 and the first-exposure experiences. In this phase, we observed a significant difference between groups based on user attention to data items in the consent forms. However, there was not an association between noticing the sensitive information requested and the type of consent forms. Slightly more participants in the Control group accepted sharing their sensitive information without noticing it as they fixated for a shorter total time on AOIs for data items compared to the other groups. The DAD design showed the potential to be more effective for directing user attention to data items, which is supported by much higher fixation counts on data items in the DAD group compared to the others. Participants in the DAD group could recall better what they had shared or not shared compared to the Control group, but no significant difference was observed for other pairwise comparisons.

Participants' attention to the other parts of the consent forms was not the same. Although participants who experienced consent forms that involved them actively with the dialogues fixated much more on the data processing purposes, this did not help them to recall significantly better for which purposes they had shared their data compared to the Control group. The Swipe and Checkbox groups, however, recalled slightly better for which purposes they had shared their data. In Phase 1, significantly more participants in the DAD group than in the Checkbox group failed to notice the dubious purposes of processing. The fact that the purposes of processing for each data item would appear if the user clicked on the data item to drag it may have contributed to the complexity of extracting information related to the purposes of processing in the DAD group compared to other layouts in Phase 1. However, after their first experiences of consent forms, the number of participants who failed to notice the dubious purposes of processing did not vary significantly between the groups.

We did not observe a significant difference in user attention to the additional policy information, located at the bottom part of consent forms, between the groups in Phase 1, as is reflected in the rather similar recall scores for additional policy information and fixation counts on the corresponding AOI.

¹²The assumptions of normality and homogeneity of variance were satisfied in each group.

Consequently, based on the results obtained from direct and indirect measures of user attention to the policy content of consent forms including data items, purposes of processing, and additional policy information about the conditions of consent, we could not reject $H_{10} - \text{attention}$ in general.

Participants in the Control group on average spent much less time handling a consent form than those in all the other groups in Phase 1. Nonetheless, the self-reported subjective measurement of efficiency showed that although handling a consent in the DAD, Swipe, and Checkbox groups took participants almost equally long, the Checkbox group perceived spending much more time than the Control group. Therefore, we reject the $H_{10} - \text{time}$ and conclude that the Control consent forms are more efficient than the other three types of consent forms, although users' opinions on the time they spent may deviate from the actual time spent.

Participants' subjective opinion on the ease of handling a consent form did not differ between the various groups in our study. We, therefore, could not reject the $H_{10} - \text{satisfaction}$.

Phase 2, habituation trial, and its effects. Overall during the habituation trial, the time spent handling consent forms in each group and the number of fixation counts on our defined AOIs decreased. Repeated exposure to dialogues decreased the difference between groups both in time spent handling each dialogue and in users' fixations on the data processing purposes and data items. However, the decrease in time and fixations on AOIs did not affect the users' privacy negatively, because this did not lead to giving non-informed consent. We observed stimulus specificity in Phase 2, which helps an organism to reduce responses to innocuous stimuli, filtering them as unimportant, while still being responsive to novel stimuli [44]. After repeated exposure to dialogues, participants knew better how to manage their time and where to look in the dialogues to gain the knowledge they required for handling the consent forms.

The test session after the habituation trial showed that there was not a significant difference between different groups of people regarding the data items and data processing purposes they recalled. Noticing the critical fields in the consent forms was not associated with the type of consent forms people experienced. In addition, within each group, we did not find strong evidence that the habituation trial affected the number of users noticing the critical fields compared to Phase 1.

Similar to the results of our study, Xu and Sunder [60] reported that interactivity encourages the processing of interactive content and compromises the processing of non-interactive content. In our study, user involvement with the content in Swipe, Checkbox, and DAD groups does not cover all the content but only the requested data and the data processing purposes. Therefore, the layout in the Control group was more homogeneous. Everything is provided in text format without requiring active interaction from the users' side with any specific content. This characteristic of the Control consent forms seems to have affected user attention to the additional policy information about the conditions of consent, in the bottom part of consent forms, compared to the other groups in Phase 2. Participants in the Control group has a higher total number of fixations on the AOI for the additional policy information compared to other groups, although this did not lead to a significant difference between the Control group and the other groups regarding how well they answered the questions on the additional policy information. Potentially biased to pay more attention towards the content of consent forms in Phase 2 because of the questions in Phase 1, the effect was shown by more fixations on the additional policy information than fixations on purposes of processing or data items in Phase 2.

Consequently, we could not reject the $H_{20} - \text{attention}$. After repeated exposure to dialogues, none of the consent forms was significantly better than the others at attracting user attention to the different AOIs we defined in the consent forms.

The average time participants spent handling a consent form in Phase 2 did not differ significantly between the groups, and we could not reject the $H_{20} - \text{time}$ after repeated exposure.

Although users took much more time to handle consent forms in the DAD, Swipe, and Checkbox groups during Phase 1, after repeated exposure and becoming familiar with the consent forms the time spent did not differ between groups. Participants' perception of the time they spent handling a consent form in Phase 2, however, differs from what actually happened. Interestingly, the Checkbox group's perception of the time they spent is much higher than that of the other groups. Consent forms with integrated Checkboxes and consent forms without any options for users to opt-in or opt-out (i.e., Control group) are today prevalent, and participants might have experienced them in different contexts. Thus, we assume that when our participants in these two groups rated their perception of the time spent on a consent form in each group, they compared it with their perception of the time it usually takes them to handle such consent forms. The Swipe and DAD actions, however, are new to users in this context. Consequently, they did not have a point of comparison.

The participants' perception of the ease of handling a consent form in Phase 2, similar to Phase 1, did not differ based on the consent forms experienced. In addition, we did not observe a significant effect of the type of consent forms on SUS scores. Therefore, we could not reject $H2_0$ – *satisfaction* after repeated exposure to consent forms. Repeated exposure did not affect user perception of the ease of handling a consent form in all groups, except in the DAD group. Participants who experienced the DAD consent forms gradually became more familiar with their structure and this significantly improved their satisfaction after Phase 2 compared to Phase 1.

Actions involved in each interaction type. In the Control group, users just click a button to agree with the stated data sharing. However, in the other groups, before clicking on the button, users perform a combination of actions to make some choices. In the Checkbox and Swipe groups, users do not directly and actively interact with the data items. However, they are required to pay attention to data items to select the desired processing purposes for each data item. In the DAD group, contrary to Checkbox and Swipe groups, users have to actively interact with data items, in addition. We assume this difference can also explain why people in the DAD group fixated the most on the data items, in both phases, followed by people in the Swipe and Checkbox groups who fixated almost similarly on the data items and more than people in the Control group (in Phase 1).

To select the desired purpose of processing for a data item in the Checkbox group, the user first points to the related checkbox and then clicks on it (point-and-click). In the Swipe group, the user first points to the arrow (see Figure 2) and then clicks and holds the arrow (click-and-hold). Afterwards, the user drags and finally releases the arrow (drag-and-release), i.e., the mouse button. In the DAD group, to start the process of selection, first, the user points to a data item. Then, the user clicks and holds the data item (click-and-hold). Finally, the user drags and drops the data item (drag-and-release/drop), i.e., releases the mouse button.

In Phase 1, the set of actions in DAD, Swipe, and Checkbox groups, compared to the Control group affected the time required to handle the consent forms. However, surprisingly, although the set of actions in Swipe and DAD groups are more complex compared to the Checkbox group, no significant difference between these groups regarding the time to handle the consent forms was observed. In addition, although the set of actions to complete the process of selection seems similar between the DAD and Swipe groups, they have two key differences that can explain the different fixation counts or recall of information related to different policy items in these two groups.

The first key difference is related to the items on which actions are performed. The point, click-and-hold, and drag-and-release actions are all performed on processing purposes in the Swipe group. For DAD, the point, click-and-hold, and the drag actions, which can be briefed as *Drag* action, are performed on data items and may direct user attention to data items while the release action, i.e., the *Drop* action, also involves processing purposes and may direct user attention to where the items should be dropped, i.e., the purposes of processing.

The second key difference is related to drag-and-release action. In the Swipe group, the path for the drag action that should be followed is clear, direct, and along a horizontal line over a text conveying the purpose of processing. The point of release is also clear, which is the end of the rectangle surrounding the text. In contrast, the path is not necessarily direct and clear in the DAD group. In other words, each user should decide about the right trajectory to the desired processing purpose individually, which makes the release action more complex.

Compared to the Swipe and Checkbox groups, the set of actions in the DAD group likely involves user interaction with data items more than with purposes of processing. Our direct and indirect measures of attention to data items and processing purposes also support this assumption.

In comparison, we assume that the DAD requires much less attention to the label of the place where the data items should be dropped (i.e., the target). The attention to the target would be the minimum necessary attention to follow the right trajectory if there is just one target. However, the attention to the target may improve if participants encounter more than one target to select for dropping the dragged items. In our DAD design, participants had to select among different targets, i.e., different purposes of processing. Although our design could have boosted user attention to the targets compared to the one-target-option scenarios, it did not help enough to gain equal attention to targets similar to the attention participants paid to the dragged items.

Limitation. As discussed in Section 5.2, in Phase 1, participants were focused on their primary tasks without any reason to pay more attention to the content of consent forms (due to the characteristics of the study) than they would have paid to the content of privacy notices in everyday life. After Phase 1, participants might have thought that they needed to pay more attention to the content of the consent forms besides paying attention to the photos they had to rate because of the questions they had received. In other words, although we had a cover story, the questions after Phase 1 might have caused participants to form some interpretations about the goal of the study and to change their behaviour accordingly, which is called demand characteristics in Reference [39]. However, this does not negatively affect our study and contributes to filtering out inattentiveness not resulting from repeated exposure but, for example, resulting from users' general lack of attention, as they may not have a reason to care about privacy. The intention to pay more attention to consent forms in Phase 2 was reflected in more user attention to policy information and not other parts of the dialogues including personal data and data processing purposes, as reported in Section 6.2.

Despite our attempts to obtain a balanced sample of participants regarding age, gender, and educational background, we had a quite young sample of participants with academic backgrounds. Our participants were recruited among students and staff at our university and more female participants volunteered than males. Nonetheless, we tried to exclude people with computer-related backgrounds and had participants from various other fields of study (see Table 2).

8 CONCLUSION AND FUTURE WORK

We conducted a laboratory experiment with 80 people in four different groups (i.e., DAD, Swipe, Checkbox, and Control) in two phases (Phase 1 and 2). The results showed that, in Phase 2, after repeated exposure to the consent forms, the initially significantly more fixations observed in Phase 1 on data items in the DAD group and those on the purposes of processing in the DAD, Swipe, and Checkbox groups disappeared and fixations on content became more similar across all groups. Moreover, the time spent handling a consent form decreased in all groups and the difference in the average time spent handling a consent form became similar between all groups. The participants' perception of the ease of handling a consent form, after repeated exposure to dialogues, similar to before that, did not differ based on the consent forms experienced. However, repeated exposure improved user perception of the ease of handling a consent form with DAD.

After repeated exposure to the consent forms (i.e., habituation trial), despite the decreased number of fixations and time spent handling dialogues in all groups, participants were equally successful at noticing the critical fields in the consent forms and recalling what they shared, for which purposes, and under which conditions. It seems that during the habituation trial they had learned how to manage their time and resources to handle consent forms more effectively and efficiently. Thus, uniform consent forms with identical layouts across different services could train users to attend to the right pieces of information. Based on our results, we have derived some design recommendations.

Design recommendation. First, if engagement with policy information in consent forms is chosen as a means for increasing user attention to certain policy aspects, then this needs to be carefully designed, so that the design does not cause insufficient or less attention to any substantial policy information that allows users to understand the consequences of the data processing. To this end, we recommend using the same type of user engagement with all substantial policy information, as the different types of interactions may create some biases in user attention to different parts of the policy.

When choosing the right interaction type for engaging users, we recommend designers to pay attention to the context in which they design interfaces. If they are designing for frequently appearing consent forms such as cookie consent notices, then checkboxes suffice for both fulfilling the legal requirements and catching user attention. However, if the consent forms are shown seldom or once in a while, then we recommend using DAD actions.

Our study showed that the DAD action results in more user attention to the policy items that must be dragged (e.g., data items as in our study) than the policy items (e.g., purposes of processing) to which dragged items are dropped. Therefore, if the DAD is used to engage users with the content, we recommend designing the DAD in a way that substantial policy information gets dragged rather than being the target for the drop action. For example, users can drag the combination of both the data item and the purpose of processing and drop them to a single dedicated area for drop action, e.g., called *My Policy*.

Future Work. One possible direction for the future work is the replication of Phase 2 without the inclusion of Phase 1 and its questions and comparing it with the results of this study. How would users react after repeated exposure to the consent forms if they were not primed by questions that might have affected their attention? Another direction that is worth exploring is the effects of different actions for providing *explicit* consent on user attention. The GDPR, in certain cases, such as for processing sensitive data, requires *explicit* consent. The *explicit consent* is achieved if users expressively confirm consent. Our results in Phase 1 showed that even with active engagement with the content, users missed the critical fields and gave their consent, which could have severe consequences on users' privacy. Does, e.g., confirming consent with digital signatures result in more attention, as users may associate such action to a larger degree with legal commitments?

ACKNOWLEDGMENTS

The authors thank Marinette Grimbeek for proofreading the manuscript, Erik Framner for coding the interfaces of the study, and Elin Nilsson for her help in reaching more participants.

REFERENCES

- [1] Bonnie Brinton Anderson, Jeffrey L. Jenkins, Anthony Vance, C. Brock Kirwan, and David Eargle. 2016. Your memory is working against you: How eye tracking and memory explain habituation to security warnings. *Decis. Supp. Syst.* 92, C (2016), 3–13.

- [2] Bonnie Brinton Anderson, C. Brock Kirwan, Jeffrey L. Jenkins, David Eargle, Seth Howard, and Anthony Vance. 2015. How polymorphic warnings reduce habituation in the brain: Insights from an fMRI study. In *Proceedings of the CHI*. ACM.
- [3] Majid Arianezhad, L. Jean Camp, Timothy Kelley, and Douglas Stebila. 2013. Comparative eye tracking of experts and novices in web single sign-on. In *Proceedings of the CODASPY*. ACM.
- [4] Article 29 Data Protection Working Party. Revised and Adopted on 10 April 2018. Guidelines on consent under Regulation 2016/679. https://ec.europa.eu/newsroom/article29/document.cfm?action=display&doc_id=51030.
- [5] Article 29 Data Protection Working Party. Adopted on 25 November 2004. Opinion 10/2004 on more harmonised information provisions. https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2004/wp100_en.pdf.
- [6] Article 29 Data Protection Working Party. Revised and Adopted on 11 April 2018. Guidelines on transparency under Regulation 2016/679. https://ec.europa.eu/newsroom/article29/document.cfm?action=display&doc_id=51030.
- [7] American Psychological Association. 2002. Ethical principles of psychologists and code of conduct. *Amer. Psychol.* 57, 12 (2002), 1060–1073.
- [8] Mike Bergmann, Martin Rost, and John Sören Pettersson. 2006. Exploring the feasibility of a spatial user interface paradigm for privacy-enhancing technology. In *Advances in Information Systems Development*. Springer, Budapest, Hungary.
- [9] Rainer Böhme and Stefan Köpsell. 2010. Trained to accept?: A field experiment on consent dialogs. In *Proceedings of the CHI*. ACM.
- [10] Aga Bojko. 2013. *Eye Tracking the User Experience: A Practical Guide to Research*. Rosenfeld Media.
- [11] Cristian Bravo-Lillo, Lorrie Cranor, Saranga Komanduri, Stuart Schechter, and Manya Sleeper. 2014. Harder to ignore? In *Proceedings of the SOUPS*. USENIX Association, Menlo Park, CA.
- [12] Cristian Bravo-Lillo, Saranga Komanduri, Lorrie Faith Cranor, Robert W. Reeder, Manya Sleeper, Julie Downs, and Stuart Schechter. 2013. Your attention please: Designing security-decision UIs to make genuine risks harder to ignore. In *Proceedings of the SOUPS*. ACM.
- [13] John Brooke et al. 1996. *SUS-A Quick and Dirty Usability Scale*. Taylor & Francis.
- [14] Fred H. Cate. 2010. The limits of notice and choice. *IEEE Secur. Privacy* 8, 2 (2010), 59–62.
- [15] Lynne Cooke. 2006. Is eye tracking the next step in usability testing? In *Proceedings of the IPCC*. IEEE.
- [16] Lorrie Faith Cranor, Praveen Guduru, and Manjula Arjula. 2006. User interfaces for privacy agents. *ACM TOCHI* 13, 2 (2006), 135–178.
- [17] Steven Cullipher and Hannah Sevan. 2015. Atoms versus bonds: How students look at spectra. *J. Chem. Edu.* 92, 12 (2015), 1996–2005.
- [18] Angelika Dimoka, Fred D. Davis, Alok Gupta, Paul A. Pavlou, Rajiv D. Banker, Alan R. Dennis, Anja Ischebeck, Gernot Müller-Putz, Izak Benbasat, David Gefen, et al. 2012. On the use of neurophysiological tools in IS research: Developing a research agenda for NeuroIS. *MIS Quart.* 36, 3 (2012), 679–702.
- [19] Olive Jean Dunn. 1964. Multiple comparisons using rank sums. *Technometrics* 6, 3 (1964), 241–252.
- [20] European Union Agency for Fundamental Rights and Council of Europe. May 2018. Handbook on European data protection law. Retrieved from <https://fra.europa.eu/en/publication/2018/handbook-european-data-protection-law>.
- [21] Susanne Furman and Mary Theofanos. 2014. Preserving privacy—More than reading a message. In *Proceedings of the Conference on UAHCI*. Springer.
- [22] Joshua Gluck, Florian Schaub, Amy Friedman, Hana Habib, Norman Sadeh, Lorrie Faith Cranor, and Yuvraj Agarwal. 2016. How short is too short? Implications of length and framing on the effectiveness of privacy notices. In *Proceedings of the SOUPS*. USENIX Association.
- [23] Jeffrey A. Gray. 1982. Précis of the neuropsychology of anxiety: An enquiry into the functions of the septo-hippocampal system. *Behav. Brain Sci.* 5, 3 (1982), 469–484.
- [24] Marian Harbach, Markus Hettig, Susanne Weber, and Matthew Smith. 2014. Using personal examples to improve risk communication for security & privacy decisions. In *Proceedings of the CHI*. ACM.
- [25] Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost Van de Weijer. 2011. *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford University Press, Oxford.
- [26] Leif-Erik Holtz, Harald Zwengelberg, and Marit Hansen. 2011. *Privacy Policy Icons*. Springer, 279–285.
- [27] Chris Jay Hoofnagle, Bart van der Sloot, and Frederik Zuiderveen Borgesius. 2019. The European Union general data protection regulation: What it is and what it means. *Info. Commun. Technol. Law* 28, 1 (2019), 65–98.
- [28] Robert J. K. Jacob and Keith S. Karn. 2003. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *Mind's Eye* 2, 3 (2003), 573–605.
- [29] Yousra Javed and Mohamed Shehab. 2016. Investigating the animation of application permission dialogs: A case study of Facebook. In *Proceedings of the DPM*. Springer.
- [30] Yousra Javed and Mohamed Shehab. 2017. Look before you authorize: Using eye-tracking to enforce user attention towards application permissions. In *Proceedings of the PETS*.

- [31] Farzaneh Karegar, Nina Gerber, Melanie Volkamer, and Simone Fischer-Hübner. 2018. Helping john to make informed decisions on using social login. In *Proceedings of the 33rd Annual ACM SAC*. ACM.
- [32] Farzaneh Karegar, Daniel Lindegren, John Sören Pettersson, and Simone Fischer-Hübner. 2018. User evaluations of an app interface for cloud-based identity management. In *Advances in Information Systems Development*. Springer, Larnaca, Cyprus.
- [33] Patrick Gage Kelley, Lucian Cescă, Joanna Bresee, and Lorrie Faith Cranor. 2010. Standardizing privacy notices: An online study of the nutrition label approach. In *Proceedings of the CHI*. ACM.
- [34] Jialiu Lin, Bin Liu, Norman Sadeh, and Jason I. Hong. 2014. Modeling users' mobile app privacy preferences: Restoring usability in a sea of permission settings. In *Proceedings of the SOUPS*. USENIX Association.
- [35] Daniel Lindegren, Farzaneh Karegar, Bridget Kane, and John Sören Pettersson. 2019. An evaluation of three designs to engage users when providing their consent on smartphones. *Behav. Info. Technol.* (2019), 1–17. DOI:<https://doi.org/10.1080/0144929X.2019.1697898>
- [36] Ewa Luger, Stuart Moran, and Tom Rodden. 2013. Consent for all: Revealing the hidden complexity of terms and conditions. In *Proceedings of the CHI*. ACM.
- [37] Aleecia M. McDonald, Robert W. Reeder, Patrick Gage Kelley, and Lorrie Faith Cranor. 2009. A comparative study of online privacy policies and formats. In *Proceedings of the PETS*. Springer.
- [38] Council of the European Union and European Parliament. 2016. Regulation (EU) 2016/679 of the European parliament and of the council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/EC (General Data Protection Regulation). *Offic. J. Eur. Union* L119 (2016), 1–88.
- [39] Martin T. Orne. 1962. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *Amer. Psychol.* 17, 11 (1962), 776–783.
- [40] Andrew S. Patrick and Steve Kenny. 2003. From privacy legislation to interface design: Implementing information privacy in human-computer interactions. In *Proceedings of the PETS*. Springer.
- [41] John Sören Pettersson. 2015. A brief evaluation of icons in the first reading of the european parliament on COM (2012) 0011. In *Privacy and Identity Management for the Future Internet in the Age of Globalisation*. Springer, Patras, Greece.
- [42] John Sören Pettersson, Simone Fischer-Hübner, Ninni Danielsson, Jenny Nilsson, Mike Bergmann, Sebastian Clauss, Thomas Krieglstein, and Henry Krasemann. 2005. Making PRIME usable. In *Proceedings of the SOUPS*. ACM.
- [43] Irene Pollach. 2007. What's wrong with online privacy policies? *Commun. ACM* 50, 9 (2007), 103–108.
- [44] Catharine H. Rankin, Thomas Abrams, Robert J. Barry, Seema Bhatnagar, David F. Clayton, John Colombo, Gianluca Coppola, Mark A. Geyer, David L. Glanzman, Stephen Marsland, et al. 2009. Habituation revisited: An updated and revised description of the behavioral characteristics of habituation. *Neurobiol. Learn. Mem.* 92, 2 (2009), 135–138.
- [45] Keith Rayner. 1998. Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* 124, 3 (1998), 372–422.
- [46] Robert W. Reeder, Adrienne Porter Felt, Sunny Consolvo, Nathan Malkin, Christopher Thompson, and Serge Egelman. 2018. An experience sampling study of user reactions to browser warnings in the field. In *Proceedings of the CHI*. ACM.
- [47] Jennifer D. Ryan, Robert R. Althoff, Stephen Whitlow, and Neal J. Cohen. 2000. Amnesia is a deficit in relational memory. *Psychol. Sci.* 11, 6 (2000), 454–461.
- [48] Florian Schaub, Rebecca Balebako, and Lorrie Faith Cranor. 2017. Designing effective privacy notices and controls. *IEEE Internet Comput.* 21, 3 (2017), 70–77.
- [49] Florian Schaub, Rebecca Balebako, Adam L. Durity, and Lorrie Faith Cranor. 2015. A design space for effective privacy notices. In *Proceedings of the SOUPS*. USENIX Association.
- [50] Bart W. Schermer, Bart Custers, and Simone van der Hof. 2014. The crisis of consent: How stronger legal protection may lead to weaker consent in data protection. *Ethics Info. Technol.* 16, 2 (2014), 171–182.
- [51] B. Shneiderman. 1996. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the VL/HCC*. IEEE.
- [52] Mike Stieff, Mary Hegarty, and Ghislain Deslongchamps. 2011. Identifying representational competence with multi-representational displays. *Cogn. Instruct.* 29, 1 (2011), 123–145.
- [53] S. Shyam Sundar, Saraswathi Bellur, Jeeyun Oh, Qian Xu, and Haiyan Jia. 2014. User experience of on-screen interaction techniques: An experimental investigation of clicking, sliding, zooming, hovering, dragging, and flipping. *HCI* 29, 2 (2014), 109–152.
- [54] Madiha Tabassum, Abdulmajeed Alqhatani, Marran Aldossari, and Heather Richter Lipford. 2018. Increasing user attention with a comic-based policy. In *Proceedings of the CHI*. ACM.
- [55] Hui Tang, Elizabeth Day, Lisa Kendhammer, James Moore, Scott Brown, and Norbert Pienta. 2016. Eye movement patterns in solving science ordering problems. *J. Eye Move. Res.* 9, 3 (2016).
- [56] Richard F. Thompson and William A. Spencer. 1966. Habituation: A model phenomenon for the study of neuronal substrates of behavior. *Psychol. Rev.* 73, 1 (1966), 16–43.

- [57] Anthony Vance, Jeffrey L. Jenkins, Bonnie Brinton Anderson, Daniel K. Bjornn, and C. Brock Kirwan. 2018. Tuning out security warnings: A longitudinal examination of habituation through fMRI, eye tracking, and field experiments. *MIS Quart.* 42, 2 (2018), 355–380.
- [58] Na Wang, Jens Grossklags, and Heng Xu. 2013. An online experiment of privacy authorization dialogues for social applications. In *Proceedings of the CSCW*. ACM.
- [59] Michael S. Wogalter, Vincent C. Konzola, and Tonya L. Smith-Jackson. 2002. Research-based guidelines for warning design and evaluation. *Appl. Ergonom.* 33, 3 (2002), 219–230.
- [60] Qian Xu and S. Shyam Sundar. 2016. Interactivity and memory: Information processing of interactive versus non-interactive content. *Comput. Hum. Behav.* 63 (2016), 620–629.

Received April 2019; revised August 2019; accepted November 2019