PriBots: Conversational Privacy with Chatbots

Hamza Harkous
École Polytechnique Fédérale
de Lausanne, Switzerland
hamza.harkous@epfl.ch

Kassem Fawaz
The University of Michigan
kmfawaz@umich.edu

Kang G. Shin
The University of Michigan
kgshin@umich.edu

Karl Aberer
École Polytechnique Fédérale
de Lausanne, Switzerland
karl.aberer@epfl.ch

ABSTRACT

Traditional mechanisms for delivering *notice* and enabling *choice* have so far failed to protect users' privacy. Users are continuously frustrated by complex privacy policies, unreachable privacy settings, and a multitude of emerging standards. The miniaturization trend of smart devices and the emergence of the Internet of Things (IoTs) will exacerbate this problem further. In this paper, we propose **Conversational Privacy Bots (`PriBots`)** as a new way of delivering notice and choice through a two-way dialogue between the user and a computer agent (a chatbot). `PriBots` improve on state-of-the-art by offering users a more intuitive and natural interface to inquire about their privacy settings, thus allowing them to control their privacy. In addition to presenting the potential applications of `PriBots`, we describe the underlying system needed to support their functionality. We also delve into the challenges associated with delivering privacy as an automated service. `PriBots` have the potential for enabling the use of chatbots in other related fields where users need to be informed or to be put in control.

1. INTRODUCTION

Privacy *notices* inform users about how websites, devices, apps, or service providers handle their data. They convey the details of how much data is collected, how long it is kept, and with which parties it is shared. Such notices manifest in several forms, from the (typically lengthy) privacy policies to the (often ambiguous) app permissions [3]. Notices also pave the way for informed *choices* to be made by users, as in opting-in for data collection, authorizing the transfer of their data to third-party ad networks, or controlling the extent to which their data is shared.

1.1 Inadequacy of Current Models

The *notice and choice* concept has so far failed at achieving its intended purposes [3]. Rarely have privacy policies deviated from traditional multi-page documents that are front-loaded with legal jargons. Privacy policies have been playing the contradicting roles of being legally-binding and being comprehensible by a layman, with the former role dominating. This has further bolstered the asymmetric relationship between the powerful service providers and the often ill-prepared/educated users. Faced with notice complexity, lack of choices, and notice fatigue, users tend to ignore these notices with time and opt to use the services directly [13].

State-of-the-art approaches to improving this model have included standardizing privacy notices (via labels [4, 10], icons [5, 9], etc.), and (semi-)automatically summarizing existing policies [14]. However, these attempts have also seen limited spread/usage. One reason for this is the rare adoption from the service providers' side, especially with the lack of incentives and the absence of regulations. Another reason is the difficulty of shaping a standard interface that appeals to the vast majority of users coming from different countries and educational backgrounds. This became even more challenging with the miniaturization trend of electronic devices that started with mobile phones and reached its peak with the Internet of Things (IoTs) [7]. The question of whether we can provide users with an easy-to-comprehend interface to learn how their data will be handled and to better control it is still an open problem.

1.2 The Rise of Conversational UI

On another note, in the recent years, we have been witnessing what might be a major paradigm shift in the evolution of user interfaces, resulting from the global rise of *Conversational UI*. With over 2.5 billion users that currently have at least one messaging app installed [1], chat has emerged as the interface understood by a large user base. This lured the big technology players, such as Microsoft and Facebook, to build new ecosystems on top of the chat interface, in the form of chatbots. In essence, chatbots (or *bots* for short) are computer agents that are designed to simulate a conversation with human users via auditory or textual methods. The primary advantage of bots stems from its familiar conversational UI: the user's messages are on the right, others' messages are on the left, and there is an input field on the bottom to compose messages. This simplicity is opposite to the custom interactions and workflows built by traditional GUIs. Fueled by advances in NLP and AI, conversational UI is well-positioned to bring us closer to the ultimate human-computer interface, which has been also termed as the "No Interface" [11] or the "Calm Technology" [8].

1.3 Conversational Privacy Bots

In this paper, we explore the potential of conversational UI in the usable privacy field by introducing the concept of

Workshop on the Future of Privacy Indicators, at the Twelfth Symposium on Usable Privacy and Security (SOUPS) 2016, June 22–24, 2016, Denver, Colorado.

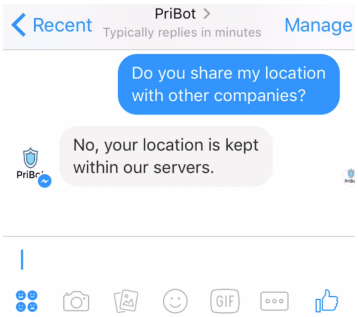

Figure 1: Example of a PriBot's answer to a user query

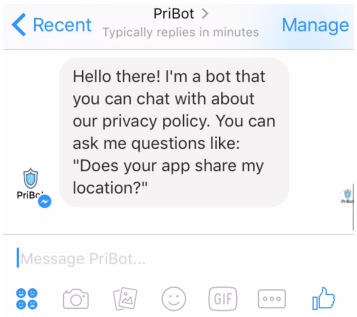

Figure 2: A PriBot initiating a dialog with the user. In this context, it the primary way of delivering the policy.

Conversational Privacy Bots (PriBots). PriBots are a new way of delivering notice and choice by creating a two-way dialogue between the user and a computer agent (a chatbot). We show an example of the operation of a PriBot in Figure 1. A user can utilize a PriBot by posing his questions about the privacy policy (notice mode) or setting his desired choices in natural language text (choice mode). Accordingly, the PriBot either replies with the relevant information or executes the relevant actions.

1.4 Motivation

The motivation for PriBots stems from two main limitations of the existing choice and notice mechanisms, which we discuss below.

Unifying Interface: PriBots alleviate the interface diversity concern through their familiar, intuitive, and widely used interface. They appeal to the two main user categories: new adopters and existing users of technology. New technology adopters are those individuals who have used texting before (via traditional SMS) but are new to the realm of applications, websites, and smart devices. Our hypothesis is that the transition from text messaging to PriBots is easier than training all these users on a completely new standard of privacy notices. Second, PriBots appeal to the existing users of traditional apps as they replace the complexity of current privacy notices with the common chat model that users are already accustomed to.

Voicing User Concerns: Another issue with the current privacy notices is that they enact a one-way conversation between service providers and users. Providers present information in both the order and the level of detail (or ambiguity) that is most convenient them. As a tool to satisfy regulatory guidelines, such notices have not been designed with users in mind. On the other hand, not only do PriBots grant users the possibility for querying the privacy-related "knowledge base" of the provider, but they also activate the backward channel from users to providers so that users can relay their concerns to the service providers.

In the following section, we discuss the role of PriBots in the process of delivering privacy notices and in enabling a new interface to set privacy choices. In Section 3, we give an overview of the architecture of the underlying system that can support PriBots, along with the various deployment options. We present in Section 4 the challenges associated with the deployment of PriBots along with the possible mitigation

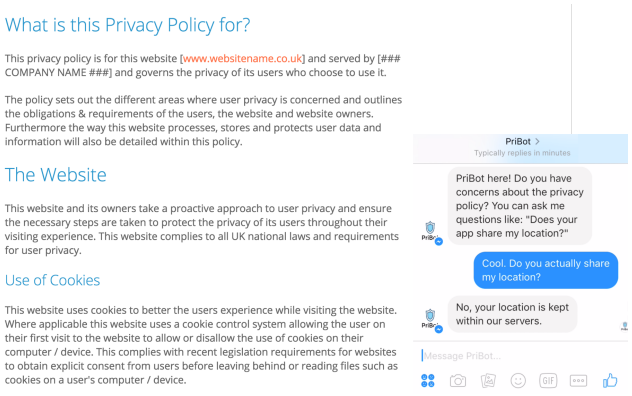

Figure 3: A PriBot serving as a complementary way of being informed about the policy

strategies. Finally, the paper will conclude with our future plans.

2. NOTICE AND CHOICE WITH PRIBOTS

As we specified earlier, PriBots run in two modes: (1) delivering privacy notices and (2) setting privacy preferences. In what follows, we elaborate on these two modes.

2.1 Delivering Privacy Policies

A PriBot can act either as the *primary* or as the *secondary* method of delivering privacy notices to the user. In the first case, the user authorizing a specific application would be given a dialog that immediately acts as a chatbot (see Figure 2 for an example). In the second case, a PriBot is given as a complementary method to the existing privacy notices. That is, a privacy policy can be shown to the user, along with the option to ask a PriBot about certain aspects of the policy (as in Figure 3). Hence, the chatbot will be acting as an advanced, semantically-aware search method.

Along the same lines, a PriBot can provide notices at two stages: *pre-authorization* and *post-authorization*. In the first stage, the user can chat with the PriBot to alleviate his privacy concerns before he authorizes the application. The PriBot can also initiate that discussion by sending a message that includes a summarized version of the privacy policy and prompting the user for any questions (as in Figure 2). In the post-authorization stage, the user can consult the PriBot at any time for any question about the privacy practices of the

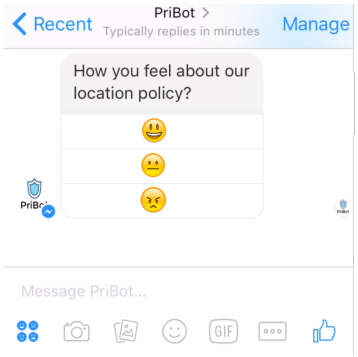

Figure 4: Example of structured messages used to solicit the user's opinions about the policy information provided by a PriBot

app (as in Figure 1).

Moreover, the user's input to the PriBot does not have to be restricted to privacy-policy queries. An additional opportunity enabled by PriBots is that users can essentially send any kind of input. One way to leverage this backward channel is to solicit feedback from users about the answers provided by the PriBot. This can be in a free textual form, which can be later analyzed using sentiment analysis techniques. It can also be in the form of responses to structured messages (as in Figure 4). In all cases, such input can be used later to improve the data handling policies themselves, thus fostering trust between the two parties.

While different alternatives do exist to realize the various PriBot features, the application context, the developers' goals, and the desired complexity will eventually decide on the appropriate alternative.

2.2 Setting Privacy Preferences

Accepting a privacy policy and setting privacy preferences in an application are two different stories, with the latter typically being user-initiated. Privacy preferences menus, provided by different applications, can be hard to locate and cumbersome to use. The user has to navigate through complex menus to locate a particular preference and is also expected to properly understand the implications of his privacy choice. The only feedback about a certain privacy preference the user gets is after the fact — from its repercussions in the wild.

PriBots improve on this current state significantly; they make the task of setting privacy preferences more usable. The user can ask a PriBot about the value of a certain preference, modify it, and inquire about its implications. By utilizing the natural language interface of PriBots, the user need not navigate through the complex menus of applications and their technical language. Figure 5 shows the steps needed to change the visibility of the user's birthday on Facebook. In the desktop version, the user has to navigate three pages before reaching the setting, assuming he knows where the setting is located to start with. In the mobile version, the flow for achieving the same goal is even different, due to the modified interface. Alternatively, by using a PriBot, as in Figure 6, the user just needs to ask a question about who can access the birthday and request limiting this access through natural language. In that sense, PriBots are platform-agnostic as the chat interface is essentially the same across various screen sizes.

PriBots open the door for applications beyond a simple question-answer mechanism. By employing advanced NLP techniques, the bot can infer the user's intentions and suggest changes to the privacy preferences. For example, a user could ask the bot if a service provider shares his location with third-party entities, and the bot will provide the user with an opt-out option in response. The bot can anticipate the user's decisions to reduce the interaction time. At some point, the user would be simply approving/denying suggested actions instead of coming up with them.

3. PROPOSED SYSTEM

In this section, we will present a high-level overview of a PriBot's architecture (shown in Figure 7), including the main components and the underlying interactions.

3.1 Architecture

We first consider the mode of privacy notices. The first step of a PriBot is to analyze the user input, provided in a free textual form via natural language processing techniques. The outcome of this analysis is semantically categorizing the user intent and determining whether he is (1) soliciting advice or (2) making a statement.[1]

In the first case, the user's query is transformed into a structured query, which is comprehensible by the *Retrieval Module*. This module allows automated reasoning about the privacy policy and returns an answer to the user's query, up to a certain confidence level. If that level is sufficiently high, the server converts the retrieved result into a natural language answer. At this stage, the answer can be toned to match the preset tone and character of the PriBot. It is important for this stage to not be a deterministic map from results to predefined sentences. Diversifying the sentences reduces the habituation effect and can result in higher user engagement and satisfaction rates. On the other hand, if the confidence level is low, the server can respond with an apology message and a link to the full policy (see Figure 8 for example). The confidence threshold is an important parameter to control, and the developer should be aware of the involved trade-offs. Increasing the required confidence threshold will result in less useful answers to users' queries. Decreasing it will result in a higher likelihood of inaccurate answers, which can have its usability and legal repercussions (see Section 4 for more about this legal aspect).

The retrieval module can benefit from a knowledge base containing the privacy policy. In addition, it might contain complementary resources that can assist in responding to queries about that policy. The policy can simply be in the original *unstructured* textual form while the algorithm applies machine learning and NLP techniques to find the part of the text relevant to the query. Alternatively, the policy can in a *structured* form, with the text annotated by various high-level, data-handling categories. Reasoning about this format can then benefit from ontologies designed specifically for privacy policies. We believe that further

[1]It is also possible that the classifier cannot decide the class with high confidence. In such a case, the PriBot can ask the user to further clarify his input.

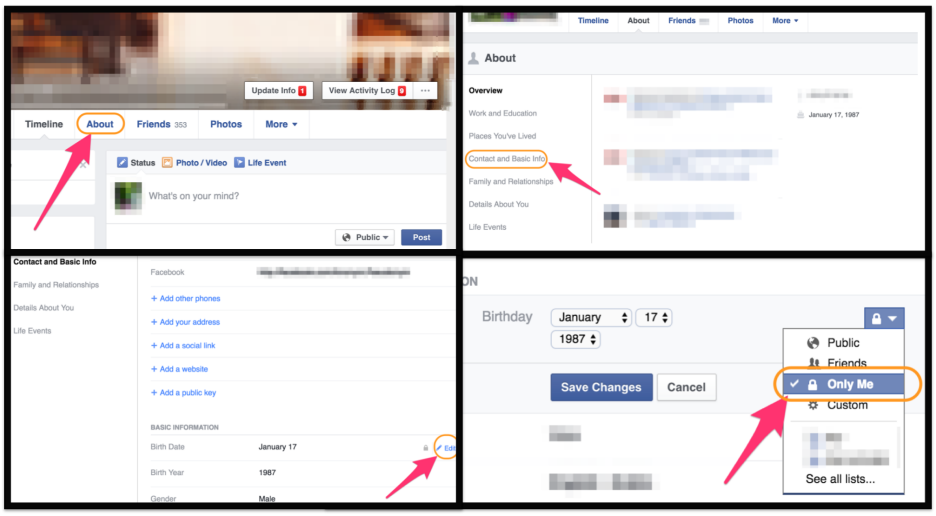

Figure 5: Steps needed to change the visibility of a user's birthday in Facebook

Figure 6: Example of changing the visibility of a user's birthday with a `PriBot`

advances in question-answering systems in general and in semantic reasoning on top of privacy policies in specific will be important pillars for supporting `PriBots`[2].

As we have indicated earlier, the user's input can be classified as a statement instead of being a query waiting for an answer. This statement can be added to a *Feedback Database*. The policy provider can run different types of analysis on this database periodically. For instance, it can explore the sentiments expressed by users towards the privacy practices. This database includes also all the previous responses to users' queries. Hence, other analyses can be run to discover the most frequently requested types of information or the most frequently missed answers. This allows the provider to further improve and amend the knowledge base supporting the retrieval module. At some point, a `PriBot` would be able to notify the user about new answers to previously missed questions based on the updated knowledge base.

[2]Notably, the recent related works within the *Usable Privacy Project* [12] can be of significant use in this regard.

On the other hand, the same system can be easily modified to support controlling privacy settings via `PriBots`. The main differences lie in the retrieval engine and the knowledge base. These should be tailored to support retrieving privacy settings instead of privacy practices. Following this, the two modes of notice and choice for `PriBots` can be combined under one system. This can be mainly achieved via a user input classification algorithm that differentiates between a user intending to change a setting and a user requesting information about a data practice.

3.2 Deployment Options

Up till now, we have discussed `PriBots` as being deployed by the provider itself. However, it is possible that a third party develops a `PriBot` to deliver privacy as a service. In such a case, the system can answer queries about various providers. These providers may not necessarily be aware of or participate in such a service[3]. One additional advantage that a third party can bring is the ability to generate new queries about multiple providers' policies. For example, this enables the user to compare multiple providers based on a specific data handling practice.

Finally, a `PriBot` is not limited to a chatbox; it can utilize voice interfaces to deliver the privacy notifications and allow users change their preferences. Needless to say, voice naturally lends itself for natural language interactions. This is particularly useful for IoT devices lacking traditional input interfaces (such as smart appliances [7]) or when utilizing interfaces is impractical (during exercising for instance) or even dangerous (while driving). `PriBots` can capitalize on the now-popular speech recognition and speech synthesis technologies to set and communicate privacy options to the user.

4. CHALLENGES

In addition to the benefits that `PriBots` bring, they still come with several challenges. In this section, we list these challenges along with possible mitigation strategies.

[3]This is conceptually similar to what services like `tosdr.org` and `explore.usableprivacy.org` do.

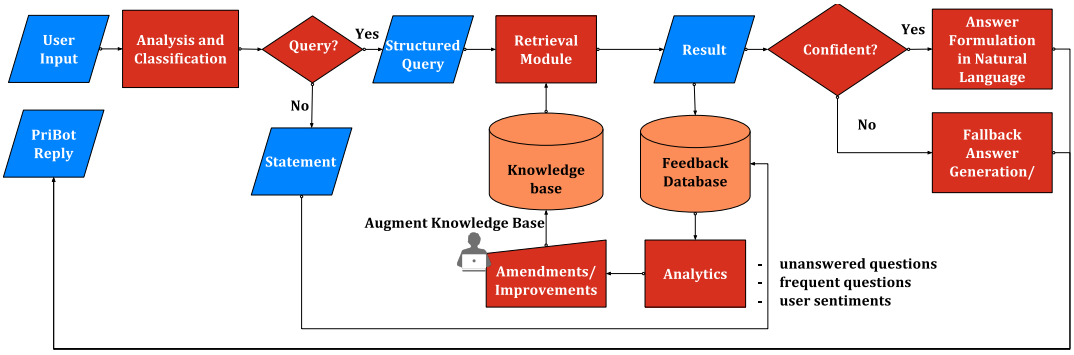

Figure 7: Proposed system architecture that can enable the functionality of PriBots in both notice and control modes

Figure 8: Example of a graceful fallback message that a PriBot sends when the confidence level is low

Mature User Understanding Techniques: The first obstacle for deploying such a system in practice is developing mature enough technologies that can support the text processing and question answering modules. Despite the significant progress in these two fields over the previous decade, we are still far from near-perfect accuracy levels. Furthermore, as we are dealing with a domain-specific system, generalized techniques cannot be directly extrapolated. Specialized datasets have to be collected in order to improve the reasoning about privacy policies and settings. Until then, PriBots can still deliver significant advantages, but it is important to control user expectations and to have graceful fallback strategies.

Legal Challenges: Traditional privacy notices are complex because they are meant to be legally binding and to be enforced like contracts on the users. When it comes to automatically generated answers based on these policies, there is a room for false positives and false negatives. The automated reasoner can tell the user that the app does share his location while, in reality, the app does not. It can alternatively tell the user that the app does not share his location while the app actually does so. Hence, the question arises on whether these answers are legally binding in the first place. Upon determining the answer to this question, PriBots should come with a clarification about this aspect. Furthermore, the case when a PriBot is run by a third party comes with additional legal challenges. Will relaying accidentally wrong information about a provider considered as a defamation? Also, how to differentiate between an accidental and intentional (but disguised) defamation? Such a question needs to be answered before PriBots can take off as a complementary means of delivering notice and/or choice.

Trusting the Machine: The next challenge is related to whether users trust the automated agent to give them answers about critical aspects, such as how their data will be used. If the automated agent had been restricted to preset answers that have been verified *a priori*, the user's trust would have been less of a concern. However, given the limitations of this static strategy, it is inevitable that PriBots will err at some point. In that case, what would happen when the user knows that he has been given an inaccurate answer? Will that result in a backlash and in an uninstallation of the app? Or would it result in totally abandoning the use of PriBots by the user? Again the answers to these questions would require further user studies. The only way to minimize those incidents from the PriBot's side is by regulating the confidence thresholds and to more frequently resort to graceful fallback answers.

Gaming the System: As PriBots can utilize users' feedback to improve future responses, there is a risk of malicious users creating the opposite effect: turning PriBots into useless agents. That raises the issue of whether the feedback loop should be fully automated or a human validation step is needed before factoring users' responses into the system. No company would desire a scenario similar to Microsoft's Tay Twitter bot, which has been recently manipulated by the users to reply with racist tweets. A PriBot's answer to a question about a privacy policy should not, for example, change due to user's negative reaction; the policy itself should be revised.

PriBots' Personality: Chatbots have the advantage of mimicking human-to-human conversations. This makes the PriBot closer to a personal privacy assistant than to a smarter search interface. Accordingly, the PriBot should be developed to have a positive tone and a consistent virtual personality. In addition to fulfilling user queries, such practices will lead to a higher user trust in the app. The PriBot also has the potential to reduce the habituation effect if the style and the content of its messages to users are diversified. This capability can be built into the engine that transforms the raw answer into complete human-style sentences.

Adapting to the Rise of Chatbots: One might go further to conjecture that chatbots, in the future, might significantly affect users' expectations of information exchange, regardless of their prior experience. As Nicholas Carr argues in his book "The Shallows: What the Internet is Doing to Our Brains", repeated exposure to a specific technology alters how the brain behaves in general. When taking web browsing as an example, he says [2]:

> "Whether I'm online or not, my mind now expects to take in information the way the Net distributes it: in a swiftly moving stream of particles. Once I was a scuba diver in the sea of words. Now I zip along the surface like a guy on a Jet Ski."

Carr builds his thesis on research around the topic of "*neuroplasticity*" by Norman Doidge and others [6], and his following argument about the effect of web browsing can be analogously extended to the potential effect of chatbots:

> "As particular circuits in our brain strengthen through the repetition of a physical or mental activity, they begin to transform that activity into a habit. The paradox of neuroplasticity, observes Doidge, is that, for all the mental flexibility it grants us, it can end up locking us into 'rigid behaviors.' The chemically triggered synapses that link our neurons program us, in effect, to want to keep exercising the circuits they've formed. Once we've wired new circuitry in our brain, Doidge writes, 'we long to keep it activated.' "

Ultimately, we might reach a stage where delivering any kind of information via chatbots becomes the default way that people expect, thus forcing a lot of interfaces to be reshaped to fit that new paradigm. In that case, the challenge would be to adapt `PriBots` to serve as primary techniques of notice and choice, instead of complementing existing mechanisms.

5. FUTURE WORK

In this position paper, we have introduced the idea of conversational privacy bots (`PriBots`), which can serve as an automated privacy-counseling agent. We believe that `PriBots` have the potential to at least complement traditional techniques for notices and choice. Via their familiar interface, they pave the way for a lot of users who care about their privacy to easily get answers to their concerns or requests. This will be enabled by advanced text processing techniques that understand the user's language and transform it into a structured format. Still, `PriBots` are at the ideation stage. We have developed a rule-based prototype that demonstrates the various scenarios discussed in this paper. In the future, we will work on moving to a customized machine-learning based system that implements the basic ideas discussed in Section 3. We will also explore the potential and the usability of `PriBots` via user studies in order to find the best context where they can bring value. Furthermore, the approach we followed with `PriBots` can be directly extended to security applications. It can allow employees, for example, to query about the security practices recommended by the company. It can even allow users to configure security settings. This will definitely bring a lot of other challenges, like whether users should be able to change their passwords from the chatbot's input field, for example. However, the perceived benefits from an interface point of view are definitely there. We believe that the work on `PriBots` will open the door to these possibilities in the security field and in general to any type of notice and choice, regardless of the context.

6. REFERENCES

[1] Bots, the next frontier | The Economist. http://www.economist.com/news/business-and-finance/21696477-market-apps-maturing-now-one-text-based-services-or-chatbots-looks-poised.

[2] N. Carr. *The shallows: What the Internet is doing to our brains*. WW Norton & Company, 2011.

[3] F. H. Cate. The limits of notice and choice. *IEEE Security Privacy*, 8(2):59–62, March 2010.

[4] L. Cranor, M. Langheinrich, M. Marchiori, M. Presler-Marshall, and J. Reagle. The platform for privacy preferences 1.0 (p3p1. 0) specification. *W3C recommendation*, 16, 2002.

[5] L. F. Cranor, P. Guduru, and M. Arjula. User interfaces for privacy agents. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 13(2):135–178, 2006.

[6] N. Doidge. *The brain that changes itself: Stories of personal triumph from the frontiers of brain science*. Penguin, 2007.

[7] Federal Trade Commission. Internet of Things, Privacy & Security in a Connected World. https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf, Jan. 2015.

[8] M. Hohl. Calm technologies 2.0: Visualising social data as an experience in physical space. *Parsons journal for information mapping*, 1(3):1–7, 2009.

[9] L.-E. Holtz, H. Zwingelberg, and M. Hansen. Privacy policy icons. In *Privacy and Identity Management for Life*, pages 279–285. Springer, 2011.

[10] P. G. Kelley, J. Bresee, L. F. Cranor, and R. W. Reeder. A nutrition label for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security*, page 4. ACM, 2009.

[11] G. Krishna. *The Best Interface is No Interface: The Simple Path to Brilliant Technology*. Pearson Education, 2015.

[12] N. Sadeh, A. Acquisti, T. D. Breaux, L. F. Cranor, A. M. McDonalda, J. R. Reidenbergb, N. A. Smith, F. Liu, N. C. Russellb, F. Schaub, et al. The usable privacy policy project. Technical report, Technical Report, CMU-ISR-13-119, Carnegie Mellon University, 2013.

[13] F. Schaub, R. Balebako, A. L. Durity, and L. F. Cranor. A design space for effective privacy notices. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*, pages 1–17, Ottawa, July 2015. USENIX Association.

[14] S. Zimmeck and S. M. Bellovin. Privee: An Architecture for Automatically Analyzing Web Privacy Policies. *23rd USENIX Security Symposium (USENIX Security 14)*, pages 1–16, 2014.