A Usability Study of Doppelganger, A Tool for Better Browser Privacy

Chris K. Karlof
Umesh Shankar

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2007-116
http://www.eecs.berkeley.edu/Pubs/TechRpts/2007/EECS-2007-116.html

September 11, 2007

Acknowledgement

This work was supported in part by National Science Foundation award number CCF-0424422 (The TRUST Center).

A Usability Study of Doppelganger, A Tool for Better Browser Privacy

Chris Karlof[1] and Umesh Shankar[2]

[1] ckarlof@cs.berkeley.edu
UC Berkeley
[2] ushankar@google.com
Google, Inc.

Abstract. We present the results of a usability study of Doppelganger, a novel system for managing HTTP cookie policies in a web browser. Doppelganger's goal is to infer personalized, privacy-preserving cookie policies in a mostly automated fashion, interrupting the user only rarely and asking intuitive questions when it does so. Using eighteen subjects, our study compared Doppelganger to two existing browser policies: the Default, allow-all policy, and the Ask policy, which requires users to make cookie decisions manually. We asked subjects to represent the stated privacy preferences of a hypothetical person while they completed a script of common web browsing tasks. We measured traditional usability metrics, such as task completion rate, but unlike most previous cookie usability studies, we also evaluated privacy performance, measured by the number of sites whose cookies were accepted during the session. In terms of the privacy metric, we found that Doppelganger performed better than the fully-manual Ask policy and far better than the Default policy. Ease of use was in between the two. We discuss usability changes suggested by subjects' performance and direct comments as well as lessons we learned to make future usability studies of Doppelganger and other cookie management tools more effective.

1 Introduction

Privacy software is fundamentally different from ordinary application software. The latter directly serves to make accomplishing a task—typically creating, editing, transmitting, or viewing content—easier for the user. The former, meanwhile, does not help the user accomplish the task at all; instead, it tries to control *metaproperties* of the content. For privacy, that means controlling information disclosure. The unhappy result is that privacy software can never make it easier to accomplish a content-related task. (Although one might define a "task" to include privacy goals, we use it here in the sense of the essential content-focused goal, for example viewing or editing a document.)

As observed by Smetters and Grinter[37] and, later, Kuo et al. [24], the differing role of privacy software also changes the way in which we must conduct usability studies. In a conventional software usability study, task completion and user satisfaction are the objects of study: how long it takes the user to complete the task, how well the task was completed (quality), how happy the user was with the experience, and so forth. For privacy related studies we must include an additional metric: how well privacy goals were met while the user performs the task.

In this paper, we report the results of a controlled usability study of Doppelganger [36], a new web browser privacy tool we have developed. Doppelganger aims to help users generate effective, personalized HTTP cookie policies with low manual effort. We were motivated to develop Doppelganger because of the poor set of cookie management choices afforded by existing browsers: users must either accept significant privacy compromise, forgo many important features, or answer an enormous number of questions trying to strike a balance between the two.

While the original intent of HTTP cookies was to provide a session state mechanism for the stateless HTTP protocol, cookies have since been used not just for things like shopping carts and authentication, but also for tracking users' web surfing habits and building targeted advertising profiles. The problem is deciding which cookies are worth accepting and which are not. Ideally, a user should be able to compare the privacy cost of a cookie with the functionality benefit the cookie enables. However, most users are not equipped to make these decisions manually and accept the global defaults in their browsers, which generally apply a single policy to all sites. Since web site features such as shopping carts and logins often require cookies and users may become confused or annoyed if these features don't work, the default browser policies liberally accept cookies.

To try to decide which cookies are useful, Doppelganger simulates a world in which the user has accepted cookies and compares it to the (default) world in which the user has not. If there is no change in the user's experience between the two worlds, then we can fairly say that the cookies are not useful. Thus, Doppelganger essentially creates a hidden twin of the user who explores the value of cookies on the sites the user browses and informs the user when accepting cookies may be a good tradeoff; useless cookies are rejected by default, to preserve privacy. Another key component of Doppelganger is an automated error recovery module, which users may invoke with a single click. The module attempts not only to correct the cookie policy, but also takes action to restore the user's session to a good state, as though cookies had been accepted from the start.

Our study compares Doppelganger against two existing browser settings for managing cookies: the *Default*, which accepts all cookies, and the *Ask* policy, which requires user approval for each cookie. The first of these represents the worst case

for privacy, since no cookies are rejected, and the second represents the potentially best case for privacy since the user can theoretically reject every cookie that she does not want.

Most previous usability studies of cookie management tools have focused on user awareness of cookies and user satisfaction, but they did not evaluate how well these tools actually help users identify useful or necessary cookies and reject useless ones. In contrast, our goal was to obtain both objective and subjective evaluations of these cookie management options. We would expect there to be tradeoffs between three measured variables: (1) ease of use, including intrusiveness and ease of decision-making with respect to the mechanism itself; (2) functionality, which here means that users were able to complete the tasks they were given to perform; and (3) privacy, measured by the number of sites from which cookies were accepted during the session. We solicited additional subjective and free-form feedback to try to understand what worked and why, as well as areas for improvement.

Although subjects found the Default policy easy to use, it also resulted in the most number of sites from which they accepted cookies, including third-party sites. Using the Ask policy, the subjects were able to protect their privacy by reducing the number of cookies accepted, but subjects found it more difficult to use, and some subjects made mistakes in their decisions which prevented them from completing all the tasks. As a compromise, Doppelganger enabled subjects to significantly reduce the number of sites from which they accepted cookies, while maintaining a relatively high ease of use.

Our subjects' comments and performance also suggested several user interface improvements to Doppelganger. Subjects indicated interest in a richer interface to Doppelganger's internals, e.g., being able to better understand what it is doing, as it is doing it. We also discuss lessons we learned from our experience to improve future usability studies of Doppelganger and other cookie management tools.

2 HTTP cookies

HTTP cookies are a general mechanism for web servers to store and retrieve persistent state on web clients [30]. When a client makes an HTTP request to a server, the server has the option of including one or more `Set-Cookie` headers in its response. The optional `expires` field in the `Set-Cookie` header indicates how long a cookie is valid. If the `expires` field is omitted, then the cookie is called a *session cookie* and should be deleted when user closes the web browser. Cookies with an `expires` field are called *persistent cookies* and persist across browsing sessions.

Cookies are also characterized by the context in which they are sent or received. After the browser receives an HTML page from a web server, it parses the page for references to elements needed to render the page (e.g., images), and issues additional HTTP request for these elements. Some of these requests may be to the same domain of the requested document, but some requests may be to different domains. The latter is often the case with advertisements. Content whose URL is from the same domain as the main page (i.e., the one in the URL bar) is considered to be *first-party*. Elements from other domains are in *third-party* context.

2.1 Cookie management

Some applications of cookies are beneficial to users. For example, web sites can use cookies to remember users' preferences and settings, implement authentication [20], and maintain session state such as shopping carts. However, web sites can also use cookies to track users and their actions. Using tracking cookies, web site operators and Internet advertisers can construct sophisticated profiles of users for targeted advertising, data mining, and information sharing with other companies.

Tracking cookies also make cookie management difficult. To prevent her web surfing habits from being tracked, a privacy conscious user might decide not to accept or send any cookies, but blocking all cookies causes a significant loss in functionality on the web and is consequently impractical for most users. Rather than blocking all cookies, the average privacy-conscious user would probably be willing to accept some cookies from the web services she derives some benefit from, but would like to block cookies that compromise her privacy "too much" or provide her no value. In this section, we discuss various browser options for managing cookies.

The Default policy. Since many web applications require cookies and fixing a mistake in a cookie policy requires navigating several confusing browser menus, the default policy in most web browsers is to accept all cookies, including persistent and third-party cookies. Although this policy requires no user interaction, it also compromises users' privacy the most.

The "Ask" policy. The "Ask" policy is a fine-grained cookie policy where the browser prompts the user for every cookie decision. With this policy, when the browser receives a cookie from a web site `foo.com`, it opens a dialog notifying the user it has received a cookie from `foo.com`, and asks the user whether it should accept the cookie, accept the cookie for each session only, or block it. An example of such a dialog box is shown here:

The user can click on the "Show Details" button to get the cookie's text and headers, and there is an option for the browser to apply the user's decision for a particular dialog to all cookies from that domain. Enabling the Ask policy requires explicit user configuration in the browser menus.

Other options. Browsers offer other configurable cookie policies as well. For example, users can configure their browsers to accept only first-party cookies, accept only session cookies, or both. Accepting only first-party cookies is a good start, but click tracking services and advertisers use HTTP redirection [31] to evade third-party cookie blockers. For example, as a user browses `xyz.com`, the server can redirect all requests through a third-party click tracking service, e.g., `trackyou.com`. Since the redirect is for a top-level request, a "first-party only" cookie policy will allow `trackyou.com` to set and receive cookies. The danger is that if `abc.com` and `xyz.com` both use `trackyou.com`, then `abc.com` and `xyz.com` can collude with `trackyou.com` to determine their common users and track their browsing habits.

Accepting only session cookies also seems like a good idea, but blocking all persistent cookies denies users the option of web site personalization and persistent authentication. Also, a session cookie used over the course of a long browsing session (say, a week) could violate a user's privacy as much as a persistent cookie.

3 Related work

3.1 Cookie Management Tools

Millett et al. [26] develop the notion of "informed consent" online, whereby users understand what is being disclosed about them and their actions, and can consent to this disclosure. The authors apply this framework to the cookie interface design in the Netscape and Internet Explorer browsers from 1995 to 2000, generally finding them lacking. Friedman et al. [19] continue with this theme, identifying a specific problem with cookie management: users need awareness but not intrusiveness for a cookie management tool to be useful. The authors implemented a "Cookie-Watcher" plugin for the Mozilla browser [27] that showed the user's cookies to the user along with color coding based on the type of cookie. Cookies could be removed, and there were also help screens to explain cookie concepts. Ultimately, though, the system did not really help users figure out which cookies were useful, and the authors only measured the system on user satisfaction rather than privacy outcomes.

Several other tools try to increase users' peripheral awareness of cookies, improve their ability to make informed decisions about cookie policies, and make the user interface for managing cookies less cumbersome [4, 5, 12–14, 26, 29, 39]. These tools help users understand the risks of accepting cookies in some cases and help alleviate the difficulty and annoyance of navigating the browser menus to manage cookie policies. However, they do little to help users evaluate the benefits of accepting a cookie, nor do they cast the problem in more intuitive terms; they are primarily better interfaces for direct cookie-level management of the cookie store.

A more promising system is Acumen [21], which uses social recommendations to help determine cookie policies. Such a system, with appropriate anonymization and additional personalization features, would be complementary to Doppelganger and could serve as another layer of informed automation before users are burdened with manual decisions.

3.2 Lessons about the usability of security and privacy software

Smetters and Grinter[37] point out that traditional measures of usability like completion time and user satisfaction are insufficient for security features, which are often a byproduct or tangential to the actual task at hand. At the least, we must see if the security goal was in fact achieved. They go on to suggest that to make progress, user interface advances are unlikely to be good enough unless coupled with underlying changes that make success easier. Kuo et al. [24] had similar criticisms of the application of conventional techniques to usability of security features. In response, they advocate a combination of four approaches: mental model interviews, surveys, "contextual inquiries" (observation of natural work patterns), and usability studies to gauge the ability of participants to complete tasks.

DeWitt and Kuljis [16] found that users were "apathetic" about security and would routinely bypass security mechanisms to get work done faster. Users don't like to be constantly interrupted with questions or alerts; and when this happens, they will tend to disable or ignore the offending mechanism [22, 32, 41]. Zurko et al. [42] analyzed user behavior when faced with security decisions in Lotus Notes. They found that users often simply did not understand the implications of security decisions, and would make incorrect choices as a result.

However, this may well be due to the poor usability of existing security and privacy software. Adams and Sasse argue [3], users can be motivated to care about security and even do a good job of it, if they are educated properly and use security mechanisms that match what they need to do. Both findings accord well with the rational incentive model of security that we adopt with Doppelganger; if users feel like security will have a big payoff, and it is not too disruptive to achieve it, they will take the necessary steps. We have architected Doppelganger to maximize the value of each user decision, and to explain the implications of each.

3.3 P3P: site-specified privacy costs

The Platform for Privacy Preferences (P3P) Project [38] is a protocol developed by the World Wide Web Consortium to help inform users of the privacy guarantees of the web sites they visit. P3P envisions users configuring their web browsers with specifications of their privacy requirements while surfing the web. Then, when a user visits a web site, that site will send a compact P3P policy specifying how it uses personal information, and the browser will determine whether the user's and site's policies are compatible. If not, the browser would inform the user of the incompatibility.

P3P seems useful for helping users make informed decisions about their cookies policies—and, as the only formally-specified privacy policy specification, we use it to present privacy costs to the user in Doppelganger—but in practice P3P has many problems [10], not least the difficulty in constructing policies and the lack of any enforcement guarantee. As a result, Egelman et al. [18] found that as of 2005, only 13.6% of popular sites had P3P policies.

Privacy Bird [9] attempts to make P3P more useful by showing an icon that indicates the level of privacy protection offered by the site being viewed. Privacy Finder [7] extends this idea by annotating search results to help users decide between search results based on privacy.

There are also products to show and analyze P3P policies. Privacy Fox [6] can parse a P3P policy and present it to the user in a more readable form. Byers et al. [8] and Levy and Gutwin [25] describe tools for automated understand of sites' privacy policies. It is our hope that either P3P or some other privacy standard becomes more widely adopted, so that a cost-benefit analysis can be more accurately performed by users and their browsers.

3.4 Economic analysis of privacy attitudes

There is still no consensus on how people make privacy decisions, partly because it is hard to find a good natural experiment; it is hard to measure privacy in a laboratory setting without using and potentially compromising subjects' personal information. There are useful results, however. Acquisti and Grossklags [1, 2] identified some obstacles to economically efficient outcomes. One that is very relevant to our discussion is the authors' finding that users' lack of information about privacy threats makes it difficult to make a good decision. Furthermore, it is hard to know in advance what the real cost of a privacy violation is going to be.

There are many ways for sites to get users to accept privacy risks, at least in study conditions. If a site agrees to protect personal data from disclosure to third parties, users are more likely to accept the risk [15]. Subjective feelings of trust in the site have also been shown to induce users to accept more privacy costs [11]. We hope that by exposing relevant privacy information to the user, Doppelganger will make privacy a more efficient market, giving sites an incentive to offer real benefits in exchange for privacy costs, and users a way to avoid the "lemons market" for privacy [40] by separating high-value sites from low-value ones.

4 How Doppelganger works

In this section, we give an overview of Doppelganger, which is drawn from a more detailed presentation in [36]; the interested reader may also download the source code directly from `http://www.umeshshankar.com/doppelganger` [34]. If a user wants to decide whether or not a particular cookie is beneficial, she must determine whether the benefit she receives from accepting the cookie outweighs the attendant privacy loss she suffers. Doppelganger helps each user perform this cost/benefit analysis and formulate her ideal cookie policy. Doppelganger relies on the following principle to identify useful cookies: if a cookie from a domain confers some benefit, it should be evident in the user's experience. If no such benefit is found, then we may assume that cookies from that site may be blocked.

A basic assumption we made in designing Doppelganger is that users don't care about cookies *per se* so much as privacy and functionality. Instead of requiring users to make uninformed low-level decisions about cookies directly, Doppelganger reformulates cookie decisions as cost-benefit analyses between privacy loss and functionality gains, which are presented to the user. This enables users to make informed decisions regarding their privacy and accept privacy loss when there is commensurate compensation.

Doppelganger uses two main techniques to identify cookies beneficial to the browsing experience: mirroring and user initiated error recovery (Figure 1). We describe these mechanisms in the following sections.

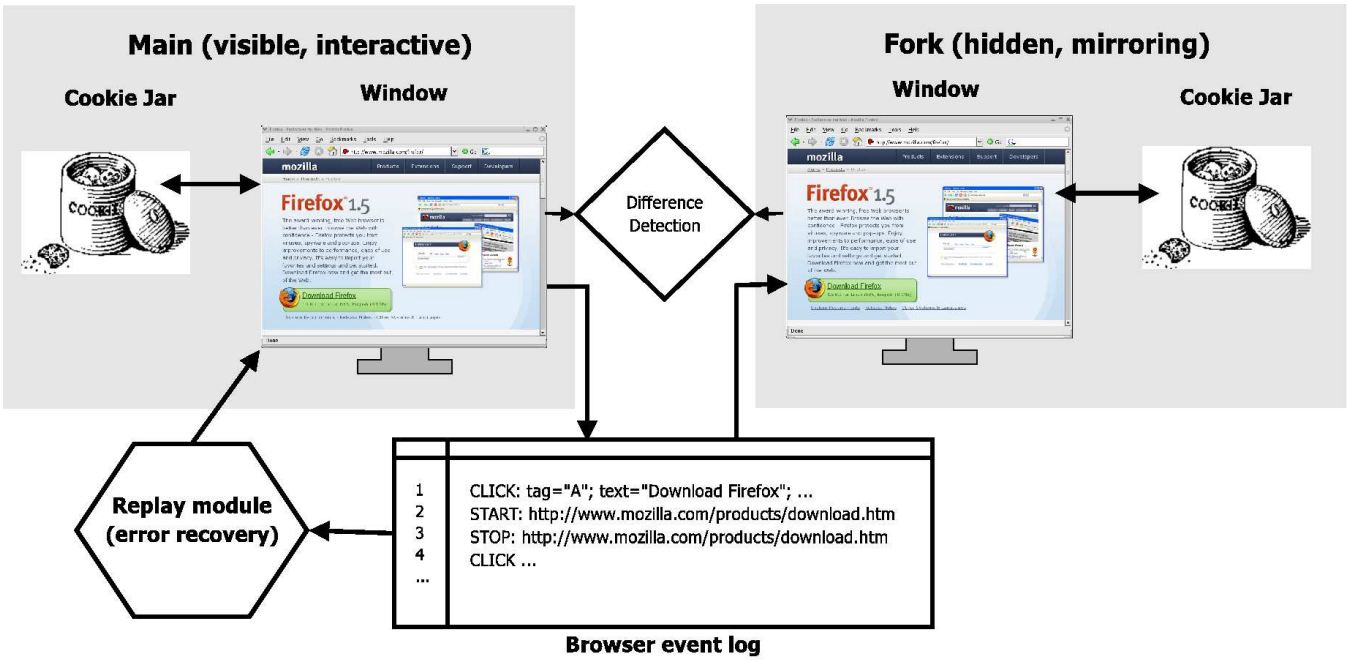


Fig. 1. An overview of Doppelganger. Doppelganger mirrors the user's web session in a hidden fork session whose only configuration difference is the cookies accepted and sent. When Doppelganger detects a difference between the contents of the main window and fork window, it reveals the fork window and asks the user to compare the two (see Figure 2). Doppelganger also maintains a log of the user's actions for error recovery.

4.1 Mirroring

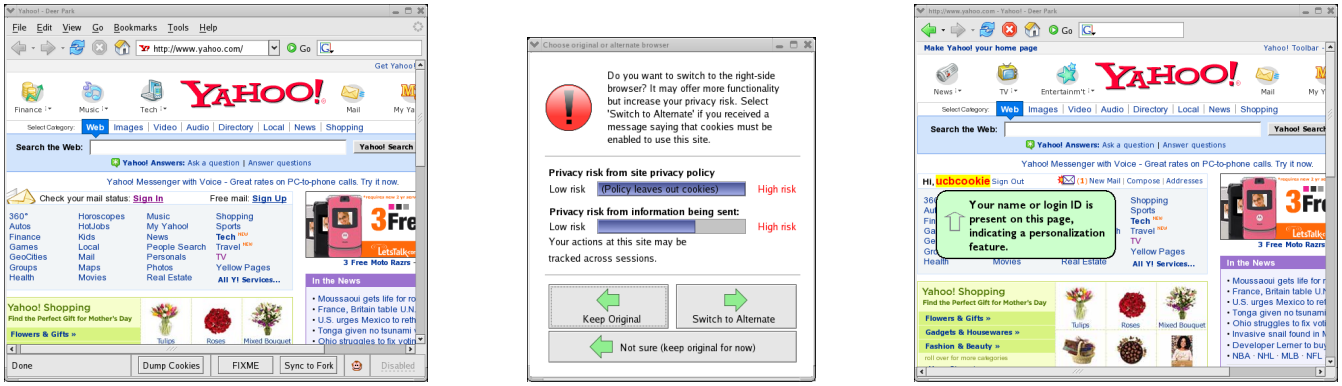
When Doppelganger encounters a domain D in the user's browsing session for which it hasn't determined a cookie policy, it mirrors the session in a hidden, parallel session whose only difference is the cookies accepted and sent. We refer to this hidden parallel session as the *fork window* since it represents a forking of the browser state. Doppelganger mirrors the session by replicating the user's main window actions in the fork window. Mirroring user events is non-trivial; we discuss it in depth in our full paper on Doppelganger [36].

As Doppelganger mirrors the user's browsing session at D, it looks for differences between the two. When Doppelganger detects a difference between the main window and fork window, it reveals the fork window and asks the user to compare the two, highlighting benefits and informing the user of the privacy risks. For an example comparison screenshot, see Figure 2. If Doppelganger detects no difference after a fixed number of page loads, Doppelganger concludes that session cookies from D provides no benefit, stops mirroring, and sets the cookie policy to deny cookies from D.

Now, suppose Doppelganger has determined session cookies from D are beneficial, and the user closes her browser and revisits D the next day. The browser may have persistent cookies from D from the previous session. Since Doppelganger has not yet determined whether persistent cookies from D are beneficial, it begins to "fork" on these cookies. Doppelganger loads persistent cookies for D from the previous session into the fork cookie space and clears all of D's cookies from the main cookie space.

Doppelganger then proceeds as it was when forking on session cookies, except now, both windows accept session cookies instead of just the fork window. The difference is the fork window may have persistent state from the previous session which positively affects the user's experience. Doppelganger tries to detect this. Similar to before, if it detects a difference, it prompts the user for a decision whether the difference is beneficial. If Doppelganger detects no difference after a fixed number of page loads, Doppelganger concludes that persistent cookies from D provide no benefit, stops mirroring, and sets the cookie policy to deny persistent cookies from D.

Difference detection. Doppelganger must be able to detect when the fork and main windows significantly differ in function or personalization enough to warrant interrupting the user for a decision. At present we use a coarse mechanism: we compare page titles (to detect obvious errors) and we look for the presence of the user's name or login ID in the fork window (and its absence in the main window) to detect personalization. A better heuristic is the source of ongoing work.



(a) Main window (b) Comparison dialog (c) Fork (mirroring) window

Fig. 2. A screenshot of the Doppelganger's comparison dialog. When Doppelganger detects a significant difference between the main and fork windows, it prompts the user for a decision. Doppelganger provides some indication of the difference and a measure of the privacy risk from accepting cookies. In this case, Doppelganger detects the presence of a personalization feature and alerts the user to it.

Exposing the cost of cookies. When Doppelganger detects a potential benefit of accepting cookies at a domain D, it tries to measure and expose the privacy risks when it prompts the user to compare the fork and main windows. One measure of the risk is the type of cookies Doppelganger must enable for the user to benefit (i.e, session or persistent). We also assess risk by interpreting the domain's P3P policy, if one exists; we borrowed some P3P parsing code from [6] for this purpose.

Logins. When Doppelganger detects a user logging into a domain, it automatically enables session cookies for that domain. The rationale for this policy is that if a user has a relationship with a site which requires a login, then accepting session cookies is unlikely to cause additional privacy loss, and we want to avoid unnecessary user interruptions.

There are many other features and implementation details of Doppelganger which we omit here for clarity, including how we handle ephemeral site visits and user options for how to approach session cookies. The interested reader is referred to the full Doppelganger paper [36] and the source code and documentation of the tool [34].

4.2 User initiated error recovery

The second technique Doppelganger uses to identify beneficial cookies is user initiated error recovery. Fix Me is a rewind-and-playback mechanism. Doppelganger maintains a log of a user's actions and browser state changes, and invokes the Fix Me mechanism when the user indicates to the system that something is wrong, perhaps due to a cookie error message or missing functionality which the mirroring system missed. The idea is that if a lack of cookies was the problem, then we may enable cookies and replay the user's actions, simulating what the user's session *would* have been if cookies had been enabled in the first place. The user interface is simple: Doppelganger installs a single button labeled Fix Me on the browser status bar that the user can click when necessary.

Doppelganger handles recovery differently depending whether it is mirroring a session or not. If Doppelganger is mirroring a session, it simply uses the mirroring comparison dialog to show the user what recovery would look like. If Doppelganger is not currently mirroring a session, it enables the next most permissive cookie policy setting (as the fork window would have) and replays the user's session at the current site from the beginning by replaying all user-initiated UI events (e.g., clicks, form submissions). We do not replay across site boundaries.

Of course, strict replaying is not the goal: we want the result to be different (and better). Doppelganger manages the replay with a state machine which watches page loads and sends user events. If Doppelganger cannot replay a user event, an expected page does not load, or an unexpected page loads, Doppelganger stops the replay. Since one of these events is evidence of a page not present in the original sequence, Doppelganger optimistically assumes the problem is fixed; since the desired outcome is one that we have not yet seen, there is no way to know if it is the correct one automatically. If the problem has in fact not been fixed, the user may click the button again, and Doppelganger will enable the next most permissive cookie setting (if possible) and replay again. If this, too, fails, then likely a lack of cookies was not the source of the problem. For further discussion of our recovery mechanism, including handling of Back and Forward buttons, and `POST` requests, refer to our full paper on Doppelganger [36].

Hours	Count
0-5	0
5-10	2
10-20	9
20+	7

(a) "How many hours a week do you use a web browser?"

Level of concern	Count
(1) Not concerned	2
(2) Somewhat concerned	11
(3) Concerned	5
(4) Paranoid	0
Average level	2.2

(b) "On a scale of 1-4, how concerned are you about your privacy online?"

Level of concern	Count
(1) Not concerned	3
(2) Somewhat concerned	8
(3) Concerned	6
(4) Paranoid	1
Average level	2.3

(c) "In particular, how concerned are you about your browsing habits and actions being recorded by the sites you visit or by third party sites?"

Fig. 3. Usability study survey results — General questions

5 Study setup

5.1 Profile of the subjects

We conducted our study with the help of the XLab (Experimental Social Science Laboratory) at UC Berkeley, which recruited and scheduled subjects, and provided space and laptop computers for conducting the study. The XLab maintains a pool of prospective subject candidates for experiments; subjects can register for the studies online. Per the XLab protocol, all of the subjects were students, staff, or faculty at UC Berkeley. Subjects were not selected for any other characteristics. No attempt was made to control for gender or any particular privacy preference; this choice was justified by the finding by Hann et al. [23] that users' willingness to choose one site over another with respect to privacy is unaffected by all the measured personal characteristics of the user, including gender. (They found that preferences did vary with site characteristics, however.)

We conducted our experiment with 19 subjects; however, one subject's data could not be used due to data corruption, yielding a final total of 18. Subjects were regular users of web browsers, most browsing for 10-20 hours a week. Tabulated data for these survey responses is provided in Figures 3 and 4. They were "somewhat concerned" about their online privacy in general, and slightly more concerned about tracking in particular. All the subjects had heard of cookies; 5 had heard of them but didn't understand how they work; 9 basically understood how they work; and the remaining 4 had a deeper knowledge, understanding the differences between types of cookies. All but 4 subjects had taken steps to manage or monitor their own cookies; this is consistent with findings in a more general population that show that as many as 58% of users delete their cookies [28], with 40% doing so each month.

Choice	Count
Not at all	0
I had heard of them, but didn't really understand how they work	5
I basically understood how cookies work	9
I understood terms like 'persistent cookies' and 'third party cookies'	4

(a) "Before this study, how familiar were you with web browser cookies?"

Choice	Count
Yes	14
No	4

(b) "Have you taken any steps to manage or monitor the cookies in your browser? For example, have you deleted cookies, examined cookies, or changed your cookie preferences, or used third-party software that did so?"

Fig. 4. Usability study survey results — Familiarity with cookies

5.2 Method

Subjects were given a packet of information that explained what browser cookies are as well as both the useful features and privacy risks that are enabled by accepting cookies. The packet also included instructions on how to use Doppelganger and explained the dialogs shown by the Ask setting (see Section 2.1 for a description of the Ask setting). Subjects were also given a task list containing the web browsing tasks they would perform. They were instructed to perform the entire list three times—three different "scenarios"—each with a different privacy setting. The first scenario used the Default (allow all) setting, and the second and third used Ask and Doppelganger in some order.

Doppelganger has multiple modes of operation, which have different session cookie policies. We used *medium paranoia* mode in our experiments. Medium paranoia uses mirroring, but when a difference is detected, automatically enables session

cookies without asking the user. Since the privacy risks of session cookies are generally low, the net benefit of accepting them is likely positive at a domain where the mirroring process detects a benefit, and we can avoid interrupting the user to make a decision. In addition, medium paranoia mode enables session cookies when a `POST` request is seen. A main benefit of the medium paranoia setting is that it automatically denies cookies from tracking sites which are visited using redirection, but never requires users to make left-or-right comparisons for session cookies.

During the browsing session, we asked subjects to represent the privacy preferences of a hypothetical user rather than their own preferences. The full task list is shown in the Appendix; the trust levels of the hypothetical user are shown after the site URLs. After performing the tasks in each of the three scenarios, the users took an exit survey. During the browsing sessions, we recording screen-capture movies for each user, so we could review the sessions later if necessary. We also installed an extension in the browser that recorded the contents of the cookie jar, including session cookies, to a file. This extension was used in each of the three scenarios.

6 Usability study results

6.1 Ease of use (tabular data in Figure 5).

Subjects indicated that performing the tasks under the Default scenario was quite easy, averaging 5.8 on a scale of 1 to 6, 6 being "Very easy". This was to be expected, and indeed was an intentional study design decision. We did not want the tasks themselves to be difficult to perform, so that we could ascribe essentially all the difficulty the subjects had in the Doppelganger and Ask scenarios to those respective cookie management mechanisms.

Completion of the tasks while using Doppelganger was judged to be somewhat more difficult than for Default, averaging 4.8, and still more difficult with Ask Me, averaging 3.8. In order to understand what made the latter two settings more difficult, we asked users a survey question about what they thought of the number and kind of questions they were asked:

> How did you feel about the number and difficulty of cookie management tasks and decisions you faced in Scenario #2 [or 3]? By "easy" we mean that it was generally clear which button to push and when, and that you knew the correct choice for each dialog box. By "difficult" we mean that it was generally unclear what was the right action in each situation.
>
> 1. There were too many tasks (dialogs and button presses) but they were easy to handle
> 2. There were too many tasks and they were difficult to handle
> 3. There were not too many tasks and they were easy to handle
> 4. There were not too many tasks but they were difficult to handle

Since the primary difference between Default and the other two settings was the dialog boxes presented to the user, we felt that this question would capture some of the reasons behind the usability gap. The results were striking:

Setting	Too many	Not too many	Difficult	Easy
Doppelganger	2	16	3	15
Ask	16	2	10	8

While users felt that Doppelganger did not impose too many tasks on them, and that those tasks were easy, they felt quite the opposite about Ask. We may also see how the ease-of-completion numbers vary with subjects' answers to these questions:

Subjects' evaluation	Ease of completion (Doppelganger)	Ease of completion (Ask)
Questions were easy	4.9	4.4
Questions were hard	4	3.3
Not too many questions	4.9	5
Too many questions	4	3.6

Users who had trouble with the decisions in volume or clarity also had a significantly more difficult time completing the tasks in both scenarios. This suggests that each of these two metrics may be a good predictor of overall ease of use.

The distributions of the ease-of-completion numbers for Ask and Doppelganger were of different shapes (see Figure 5). While 12 of the 18 subjects decided that completion with Ask was "relatively easy" (the rest were uniformly distributed from

Ease	Count
(1) Very difficult	0
(2) Difficult	0
(3) Relatively difficult	0
(4) Relatively easy	1
(5) Easy	3
(6) Very easy	14
Average	5.8

(a) "How easy was it for you to complete the tasks in scenario [Default]?"

Ease	Count
(1) Very difficult	0
(2) Difficult	0
(3) Relatively difficult	2
(4) Relatively easy	6
(5) Easy	4
(6) Very easy	6
Average	4.8

(b) "How easy was it for you to complete the tasks in scenario [Doppelganger]"

Ease	Count
(1) Very difficult	1
(2) Difficult	1
(3) Relatively difficult	2
(4) Relatively easy	12
(5) Easy	1
(6) Very easy	1
Average	3.8

(c) "How easy was it for you to complete the tasks in scenario [Ask]?"

Fig. 5. Survey results — ease of completion. Doppelganger and Ask were the second and third scenarios; for 10 of the 18 subjects, Ask came before Doppelganger.

"very difficult" to "very easy"), 10 of 18 found completion with Doppelganger to be "easy" or "very easy". Only 2 subjects deemed completion with Ask "easy" or "very easy" despite the task list's intrinsic ease (as measured by Default). This seems to indicate that Ask is not satisfactory for a large majority of users, but that Doppelganger may be well suited for a significant portion of the population.

Subjects were given the opportunity to provide free-form feedback for each scenario in response to the question "What did you like or not like about Scenario #N?". Comments were generally positive for Default; the only criticism offered was by one subject who said that s/he "did not know which cookies were being accepted by [his/her] browser". Subjects were much more critical of the Ask scenario. Several complained about the number of dialog boxes. Another common refrain was that subjects were confused; they hard a hard time figuring out what cookies they were enabling, and at the end were not sure if they had made the right choices. A few noted that the presence of dialogs asking about cookies from other sites (i.e., third party cookies) was confusing, and they did not always notice when this happened.

Several users indicated Doppelganger was easy to use and liked the privacy protection it provides. One subject said s/he "liked the interface because it was easy to use and it seemed like [s/he] was getting more protection from cookies" compared to the Default setting. Another said s/he "found it easy to use for the level of security it seemed to be providing". The full text of all survey questions and answers may be found in the Appendix of Shankar's Ph.D. dissertation [35].

6.2 Performance

We measured performance in a few ways: completion of each scenario; whether the user encountered any problems along the way (even if the scenario was completed); and the number sites whose cookies were accepted.

Every subject was able to complete the Default scenario without any problems. Six subjects were unable to complete Task 2 (Netflix) using Ask, and 3 were unable to do so using Doppelganger. In each case the subject had chosen to deny cookies from Netflix, making it impossible to access the site. This is not surprising, given that Netflix had been marked as having a low trust level, but we should take that into account in interpreting the cookie numbers, which were deflated relative to subjects that accessed Netflix. Subjects did not report problems other than their annoyance or confusion with the dialog boxes, which was much more prevalent with Ask than with Doppelganger (see above).

The accepted-cookies data, shown in Figure 6, is revealing. For every type of cookie, the Default setting allowed the most sites to set cookies by a wide margin, followed by Ask and then Doppelganger. The Default setting also led to the acceptance of a large number of third-party sites' cookies (almost 8 sites on average). Notably, subjects did not accept any third-party cookies using Doppelganger; this represents a significant improvement, since even a small number of third-party sites' cookies can enable tracking at an enormous number of other sites. The variance was much larger for the Ask setting than for the other two. This is to be expected given subjects' confusion about what to do, leading to, in many cases, arbitrary decision-making. Whereas

the flexibility of Ask might lead to large variance in the real world, since users many have very different preferences, we would expect to see much less variance in a controlled environment where users were told to represent the same privacy preferences. That we did not see small variance with Ask, but did see it with Doppelganger, says a lot about the need to help users make informed choices. Variance in the Default setting was zero for first-party cookies, which is to be expected: each of the three first-party sites set as many cookies as it wanted. Third-party variance was due to randomness in the set of advertisements being displayed.

As we discussed above, in order to do a more fair comparison, we would need to consider the effects of the missing Netflix cookies for some subjects. Since only one site is affected, and it is first-party, the maximum increases we could see in the averages are 6/18 = 0.33 for Ask in each of the first-party categories, and 3/18 = 0.17 for Doppelganger. Neither would make a significant difference in our interpretation.

7 Discussion

7.1 Overall ease of completion

Perhaps unsurprisingly, users rated Doppelganger's ease of use in between that of the Default setting and that of the Ask setting. Since Doppelganger also performed as well or better than Ask in terms of privacy protection for every cookie category, we may fairly say that the Doppelganger dominates the Ask setting. The remaining question is whether the privacy protections offered by Doppelganger—and the data suggest that they are significant—outweigh the usability decline suffered by users in using it relative to the Default setting. Since users care enough about cookies to take action to manage them, there is reason to believe that if we offer a reasonable option, users will actually use it.

7.2 How Doppelganger might be improved

Suggested user interface improvements There were also a few comments that suggested that a better layout of some of the Doppelganger dialogs would be helpful. Our observations of the subjects as they used Doppelganger suggested that its asynchronous, active nature may require better explanations of what it is doing, as it is doing it. For example, one subject "would have liked to have an idea what the program was syncing".

Some subjects also indicated that they were not sure when the Fix Me button was applicable. Subjects stated "it was confusing because I don't really know what FIXME/SYNCED is all about" and "the FIXME button was helpful but it didn't look like it was applicable when entering the Netflix site".

Part of the difficulty is that there are two reasons that the Fix Me button might not be available and also cases in which, when pressed, it may not solve the user's problem. First, Fix Me button is only available when the fork browser has "caught up" to the main one; that is, when both browsers have finished loading the same page. Also, once cookies have been enabled for a site, there is nothing further to do. Finally, some site problems are not due to disabled cookies, and in those cases clicking Fix Me will not solve the problem. A tooltip explanation of what clicking the button might accomplish could alleviate some of these concerns.

Additional feedback to the user Comments from the subjects seemed to indicate that they like to see how they are doing; that is, how their choices are affecting their privacy. One subject "didn't really see any significant advantage or benefit in using Doppelganger over the default browser". Visual feedback may be appropriate in at least two ways: (1) showing the cookie policy for and number of cookies accepted from the current site; and (2) some sort of historical count or average indicating how many sites' cookies the user has blocked. The latter might induce the user to improve her privacy "score".

Showing which cookies are currently accepted is useful because cookies may be used to track the user without any visual indication. For example, `weather.com` remembers the zip code that a user types in using a persistent cookie, even if the user does not choose to be remembered and is asked to re-enter the zip code on future visits.

One subject's comments indicated that he/she wasn't sure if selecting the fork browser in a left-or-right choice committed the entire browsing session to that choice, or just the site. Again, a visual indicator of the current policy may help with this distinction. In addition, the left-or-right dialog box could make the implications of the choice more precise.

7.3 Effect of role-playing

We used role-playing in our study, and in the exit survey we did ask participants how well they thought they represented the privacy preferences of the person we asked them to represent. Researchers have found that role-playing in privacy studies can sometimes be problematic [17, 33]. To minimize risk of real privacy violations, role-playing participants are given fictitious usernames, passwords, and form input, as well as instructions to act on behalf of the fictitious identity. Since it is fairly clear

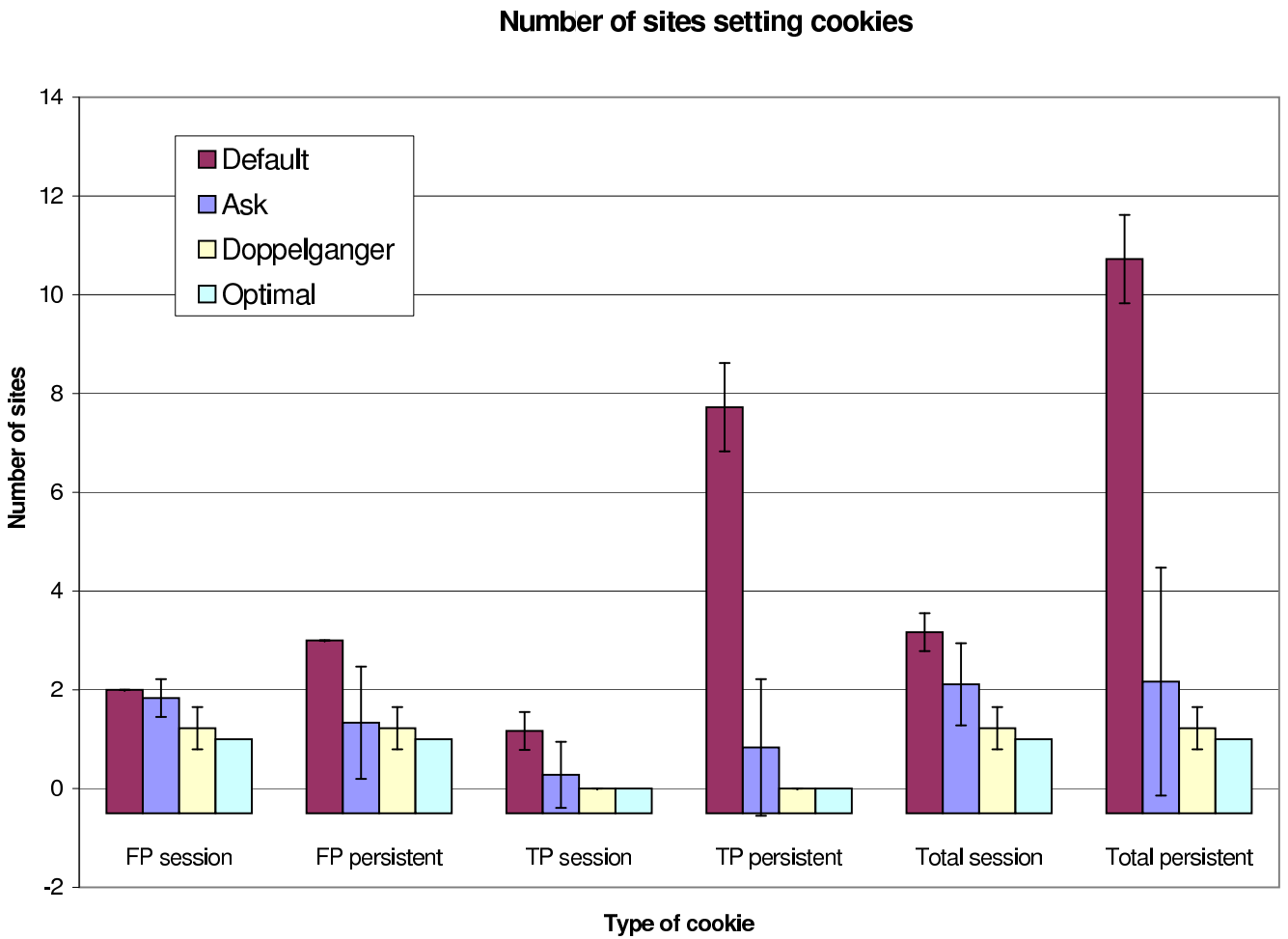


Fig. 6. Cookies accepted for each cookie management setting. The figures used here are the averages across all users, taken from the cookie jar as of the end of task list (Figure 7). Each type of cookies from each site counted as "1" in the tally; i.e., we did not distinguish between accepting two first-party session cookies from a site and accepting three first-party session cookies from the site. The end of the task list represents the end of the second browsing session, since users had to close and restart the browser after Task 3. Error bars represent the standard deviation. The "optimal" number for each category represents the minimum number of cookies which would both allow the tasks to be accomplished and reflect the privacy preferences of the hypothetical user that subjects were supposed to emulate.

that their own privacy is at minimal risk, participants may behave differently than if they acted on their own behalf. However, the alternative is to expose study participants to real privacy risk, which has obvious ethical implications, and could also limit the study pool in artificial ways, since not every participant would have a relationship with every web site. There are two factors contributing to the accuracy of representation: converting our written description to a mental privacy model, and translating that model into actions during browsing; since we're largely concerned with the latter, role-playing seems like a reasonable choice.

7.4 Further areas for study

Many of the remaining questions we want to answer are related to the long-term usefulness of Doppelganger—would users have the patience to learn how to use it? would it provide lasting value? would it fall prey to sites designed to defeat it? would users' behavior change when their personal privacy is at stake?—and cannot be measured in a laboratory in a short, controlled experiment.

Nonetheless, there are some ways in which we might have learned more useful information. First, we did not include a hands-on training segment. Users were given an information packet to read, but did not have the opportunity to try using either Doppelganger or the Ask setting before the study started. It is not clear our approach accurately emulates how users would start using either of these mechanisms. Users often do not read initially documentation for new software; they often just start trying it out. To create a more realistic environment, it may have been prudent to give users time to acquaint themselves with each mechanism before the study and give less written information on how each tool works.

Although most our users used Doppelganger effectively during their first encounter with it, longer term results will ultimately be more conclusive and may expose more interesting user behavior. For example, based on personal experience of the authors, when we first started using the Ask setting, we thought somewhat carefully about each cookie decision, but after making a mistake or two, the inconvenience of navigating the browser menus to correct them caused us to gradually make more liberal decisions. A longer term study might reveal similar tendencies in Doppelganger. For example, many of our users carefully considered their options before pressing Fix Me, but over time it is possible users might try pressing Fix Me more aggressively to correct problems unrelated to cookies and thus form more liberal cookie policies. Also, some subjects were hesitant about being required to make excessively many left-or-right comparisons. One subject stated that "if I were to do that every single time I wanted to browse a website, I think would get a little bit sick of it". A longer term study would help clarify this issue; users were experiencing the startup phase of using the tool. In the steady state, they would be expected to make many fewer decisions, since the policies for each site would have been determined.

7.5 Sources of error

There were some potential sources of error and uncertainty in our study. First, the subjects were not representative of the general web browsing population; since all were affiliated with UC Berkeley, they were likely more educated than average. In addition, it is possible that more users are needed to confirm our findings. A possible source of bias is that users were informed that we (the experimenters) were also the authors of Doppelganger; this may have affected their experience or reports of their experience.

8 Conclusion

Ultimately, we would like to conduct a large-scale, long-term deployment. It is difficult to predict the steady-state usability of Doppelganger from a one-hour study session in which the subject is not only seeing Doppelganger for the first time, but may well be learning quite a bit about cookies and online privacy as well. At the same time, a controlled study lets us take meaningful aggregates and more precise comparisons across users. We believe that the study made a fairly good case for the viability of Doppelganger, and users made good suggestions that can be used to improve the system. People found it to be significantly more usable that its privacy-comparable counterpart, Ask; none accepted third-party cookies, while the majority did accept third-party cookies with Ask. Doppelganger also fared well in some absolute sense; most users found it "easy" or "very easy" to complete the browsing tasks while using it.

9 Acknowledgements

This work was supported in part by National Science Foundation award number CCF-0424422 (The TRUST Center).

References

1. ACQUISTI, A. Privacy in electronic commerce and the economics of immediate gratification. In *EC '04: Proceedings of the 5th ACM conference on Electronic commerce* (New York, NY, USA, 2004), ACM Press, pp. 21–29.
2. ACQUISTI, A., AND GROSSKLAGS, J. Privacy and rationality in individual decision making. *IEEE Security and Privacy 3*, 1 (2005), 26–33.
3. ADAMS, A., AND SASSE, M. A. Users are not the enemy. *Communications of the ACM 42*, 12 (1999), 40–46.
4. Add & Edit Cookies. `http://addneditcookies.mozdev.org/`.
5. ALSAID, A., AND MARTIN, D. Detecting web bugs with bugnosis: Privacy advocacy through education. In *Proceedings of the 2002 Workshop on Privacy Enhancing Technologies* (2002).
6. ARSHAD, F. Privacy Fox - A JavaScript-based P3P Agent for Mozilla Firefox. `http://privacyfox.mozdev.org/PaperFinal.pdf`, 2004.
7. BYERS, S., CRANOR, L., KORMANN, D., , AND MCDANIEL, P. Searching for privacy: Design and implementation of a p3p-enabled search engine. In *2004 Workshop on Privacy Enhancing Technologies (PET2004)* (2004).
8. BYERS, S., CRANOR, L. F., AND KORMANN, D. Automated analysis of P3P-enabled web sites. In *ICEC '03: Proceedings of the 5th international conference on Electronic commerce* (New York, NY, USA, 2003), ACM Press, pp. 326–338.
9. BYERS, S., CRANOR, L. F., KORMANN, D. P., AND MCDANIEL, P. D. Searching for privacy: Design and implementation of a p3p-enabled search engine. In *Privacy Enhancing Technologies* (2004), pp. 314–328.
10. CENTER, E. P. I., AND JUNKBUSTERS. Pretty poor privacy: An assessment of P3P and Internet privacy. `http://www.epic.org/reports/prettypoorprivacy.html`, June 2000.
11. CHELLAPPA, R. K., AND SIN, R. G. Personalization versus privacy: An empirical examination of the online consumer's dilemma. *Inf. Tech. and Management 6*, 2-3 (2005), 181–202.
12. Cookie Button. `http://basic.mozdev.org/cookiebutton/`.
13. Cookie Culler. `http://cookieculler.mozdev.org/index.html`.
14. Cookie Toggle. `http://cookietoggle.mozdev.org/`.
15. CULNAN, M. J., AND ARMSTRONG, P. K. Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation. *Organization Science 10*, 1 (1999), 104–115.
16. DEWITT, A. J., AND KULJIS, J. Aligning usability and security: a usability study of polaris. In *SOUPS '06: Proceedings of the second symposium on Usable privacy and security* (New York, NY, USA, 2006), ACM Press, pp. 1–7.
17. DHAMIJA, R., TYGAR, J. D., AND HEARST, M. Why phishing works. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2006), pp. 581–590.
18. EGELMAN, S., CRANOR, L. F., AND CHOWDHURY, A. An analysis of p3p-enabled web sites among top-20 search results. In *ICEC '06: Proceedings of the 8th international conference on Electronic commerce* (New York, NY, USA, 2006), ACM Press, pp. 197–207.
19. FRIEDMAN, B., HOWE, D., AND FELTEN, E. Informed consent in the Mozilla browser: Implementing value sensitive design. In *35th Annual Hawaii International Conference on System Sciences (HICSS'02)* (2002).
20. FU, K., SIT, E., SMITH, K., AND FEAMSTER, N. Dos and Don'ts of client authentication on the web. In *10th USENIX Security Symposium* (August 2001), pp. 251–268.
21. GOECKS, J., AND MYNATT, E. D. Social approaches to end-user privacy management. In *Security and Usability: Designing Secure Systems That People Can Use*, L. F. Cranor and S. Garfinkel, Eds. O'Reilly, 2005, ch. 25, pp. 523–545.
22. GOOD, N., DHAMIJA, R., GROSSKLAGS, J., THAW, D., ARONOWITZ, S., MULLIGAN, D., AND KONSTAN, J. Stopping spyware at the gate: A user study of notice, privacy and spyware. In *Symposium on Usable Privacy and Security (SOUPS) 2005* (July 2005).
23. HANN, I.-H., HUI, K.-L., LEE, T. S., AND PNG, I. P. L. Online information privacy: Measuring the cost-benefit trade-off. In *Proceedings of the Twenty-Third International Conference on Information Systems* (2002), pp. 1–8.
24. KUO, C., PERRIG, A., AND WALKER, J. Designing an evaluation method for security user interfaces: lessons from studying secure wireless network configuration. *interactions 13*, 3 (2006), 28–31.
25. LEVY, S. E., AND GUTWIN, C. Improving understanding of website privacy policies with fine-grained policy anchors. In *WWW '05: Proceedings of the 14th international conference on World Wide Web* (New York, NY, USA, 2005), ACM Press, pp. 480–488.
26. MILLETT, L., FRIEDMAN, B., AND FELTEN, E. Cookies and web browser design: Toward realizing informed consent online. In *Proceedings of the CHI 2001 Conference on Human Factors in Computing Systems* (April 2001), pp. 46–52.
27. Mozilla web browser. `http://www.mozilla.org`.
28. O'MALLEY, G. Jupiter analyst: Nielsen research confirms users delete cookies. `http://publications.mediapost.com/index.cfm?fuseaction=Articles.san&s=2%8883&Nid=12855&p=297686`.
29. Permit Cookies. `http://gorgias.de/mfe/`.
30. Persistent client state: HTTP cookies, Preliminary specification. `http://wp.netscape.com/newsref/std/cookie_spec.html`.
31. POULSEN, K. Microsoft cookies jump domains. `http://www.securityfocus.com/news/83`, September 2000.
32. SALTZER, J. H., AND SCHROEDER, M. D. The protection of information in computer systems. *Proceedings of the IEEE 63*, 9 (September 1975), 1278–1308.
33. SCHECHTER, S., DHAMIJA, R., OZMENT, A., AND FISCHER, I. Emperor's new security indicators: An evaluation of website authentication and the effect of role playing on usability studies. In *Proceedings of the 2007 IEEE Symposium on Security and Privacy* (May 2007).
34. SHANKAR, U. Doppelganger Homepage. `http://www.umeshshankar.com/doppelganger`.

35. SHANKAR, U. *Bridging the Gap Between People and Policies in Security and Privacy*. PhD thesis, EECS Department, University of California, Berkeley, December 21 2006.
36. SHANKAR, U., AND KARLOF, C. Doppelganger: Better browser privacy without the bother. In *Proceedings of the 13th ACM Conference on Computer and Communications Security (CCS 2006)* (New York, NY, USA, 2006), ACM Press.
37. SMETTERS, D., AND GRINTER, R. E. Moving from the design of usable security technologies to the design of useful secure applications. In *New Security Paradigms Workshop* (Virginia Beach, VA, September 2002).
38. The Platform for Privacy Preferences Project (P3P). `http://www.w3.org/TR/P3P/`.
39. View Cookies. `http://www.bitstorm.org/extensions/view-cookies/`.
40. VILA, T., GREENSTADT, R., AND MOLNAR, D. Why we can't be bothered to read privacy policies models of privacy economics as a lemons market. In *ICEC '03: Proceedings of the 5th international conference on Electronic commerce* (New York, NY, USA, 2003), ACM Press, pp. 403–407.
41. YEE, K.-P. Guidelines and strategies for secure interaction design. In *Security and Usability: Designing Secure Systems That People Can Use*, L. F. Cranor and S. Garfinkel, Eds. O'Reilly, 2005, ch. 13, pp. 247–273.
42. ZURKO, M. E., KAUFMAN, C., SPANBAUER, K., AND BASSETT, C. Did you ever have to make up your mind? What Notes users do when faced with a security decision. *Annual Computer Security Applications Conference 2002* (2002), 371.

A Appendix

Open the Firefox browser, using the icon corresponding to the scenario you are on.
Task 1: www.excite.com Trust Level: HIGH Excite.com is a web portal which offers services such as email, news, and stock quotes
0) If you are using Doppelganger, take a moment to look at the lower right corner of your browser. You should see the FIXME button and a status indicator saying "Synced". You can read more about how to use these in the Doppelganger scenario description, above. 1) Visit http://www.excite.com 2) Log in to an email account (ID: ucbtest; password: gobears) 3) Read the message from Testy McTest (ucbcookie@yahoo.com). Make sure you can see the "secret phrase". 4) Return to www.excite.com 5) Find a news article and click it. 6) If you can read the article, then you are done with Task 1.
Task 2: www.netflix.com Trust Level: LOW Netflix is an online DVD movie rental service 1) Visit http://www.netflix.com 2) Choose "Browse Selection" 3) Search for the movie "Shrek". 4) If you can see the name of the director of "Shrek", you are done with Task 2
Task 3: www.weather.com Trust Level: LOW Weather.com provides weather forecasts and historical data 1) Visit http://www.weather.com 2) Find the forecast for zip code 11217 using the "Local Weather" box 3) If you see the city name for zip code 11217, you are done.
Close the browser and wait a few seconds. Now, re-open the browser, using the same link.
Task 4: Revisit www.excite.com 1) Visit http://www.excite.com 2) Check your email again, and reply to the message from ucbcookie@yahoo.com. Write "I'm almost done!"
Task 5: Revisit www.weather.com 1) Visit www.weather.com 2) Find the current temperature for zip code 11217 again.

Fig. 7. Task list used in the Doppelganger usability study.