# Trained to Accept? A Field Experiment on Consent Dialogs

**Rainer Böhme**
International Computer Science Institute
Berkeley, California, USA
rainer.boehme@icsi.berkeley.edu

**Stefan Köpsell**
Technische Universität Dresden
Dresden, Saxony, Germany
stefan.koepsell@tu-dresden.de

## ABSTRACT

A typical consent dialog was shown in $2\times2\times3$ experimental variations to 80,000 users of an online privacy tool. We find that polite requests and button texts pointing to a voluntary decision *decrease* the probability of consent—in contrast to findings in social psychology. Our data suggests that *subtle positive* effects of polite requests indeed exist, but *stronger negative* effects of heuristic processing dominate the aggregated results. Participants seem to be habituated to coercive interception dialogs—presumably due to ubiquitous EULAs—and blindly accept terms the more their presentation resembles a EULA. Response latency and consultation of online help were taken as indicators to distinguish more systematic from heuristic responses.

### Author Keywords

informed consent, privacy notices, EULA, default button, user behavior, field experiment, AN.ON/JonDonym

### ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: User Interfaces—*interaction styles, user-centered design, theory and methods*; H.1.2 Models and Principles: User/Machine Systems—*human factors, human information processing*; J.4 Social and Behavioral Sciences: *psychology*

### General Terms

Experimentation, Human Factors, Security

## BACKGROUND AND THEORY

In many situations, information systems require user inputs which result in substantial economic, social or legal consequences. Hence, finding the right response requires considerable attention and cognitive effort. But users are unlikely to make this effort when inputs are requested in *interception dialogs*, which by definition interrupt the user's current primary task. Examples for such situations include the acceptance of end-user license agreements (EULAs) at software install time or upon service subscription [7, 6], and reactions to virus alerts and active browser warnings [3, 4, 12]. Also in the area of privacy and data protection, strong emphasis is put on the principle of informed consent across jurisdictions [5]. Anecdotal evidence, intuition (from introspection), and empirical studies on various forms of interceptions all suggest that interception dialogs fail to reach their purpose. Too many users simply do not bother to read notices and do not heed warnings. Reasons are lack of choice [14], inaptitude [7], or chiefly habituation: over the years, interface designers have trained their users to click dialogs away to get their primary task done [4]. In hope of a remedy, most HCI research on this subject innovated on the presentation of interceptions to attract users' attention [7, 6, 4, 12]; yet with limited success. Though statistically significant relative improvements were found, the absolute level of attention often remains unsatisfactory. Moreover, when innovative interfaces become mainstream, habituation might kick in again and take away the reported (short-term) benefits.

The existing evidence is largely based on laboratory experiments. While this is the method of choice to evaluate novel interfaces, it comes at the cost of small samples and ecological (in)validity. Small samples impede analyses of higher-order interactions between experimental variations and inter-individual differences. However, understanding individual differences is essential to improve beyond catalogues of case-specific relations between stimuli and aggregate responses, and to move towards better models of human decision making. Similar approaches have led to major insights in related disciplines, like social psychology and survey research [10]. A very simple model is to assume a dual-path process: individuals either make *systematic* decisions, taking into account all available information, or resort to *heuristics* as convenient shortcuts [2]. Which of the two paths is chosen in a specific decision depends on the type of decision, the individual's state of involvement, its personality traits, and contextual cues, such as properties of consent dialogs.

This note contributes results from a large-scale field experiment, in which we modify subtle properties of a typical consent dialog and measure differences in behavior. The properties were selected to form contextual cues that stimulate either systematic or heuristic responses, and at the same time are general enough for our focus on common processes, which likely take place in the specific instances of interception dialogs studied in prior work (EULAs, privacy notices).

## STUDY DESIGN

An opportunity for our field data collection was given between May 20th and July 20th, 2008, when the operators of the free Internet anonymity service AN.ON/JonDonym (`http://anon.inf.tu-dresden.de`) initiated a study to mea-


*CHI 2010*, April 10–15, 2010, Atlanta, Georgia, USA.

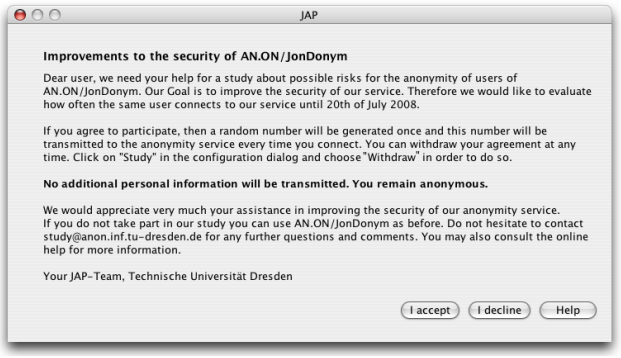


**Figure 1. Example of our consent dialog (Mac OS X, English language)**

sure the security of their service against a new kind of threat [1]. Their research required an update of the client software and the operators found it appropriate to let the users decide whether they would like to participate in the measurement process or not. The operators agreed to implement this consent dialog with $2\times2\times3$ experimental conditions designed by us, and assigned them to users randomly and independently. User reactions were measured by the participation rate (i. e., fraction of users who approved the request), whether or not further information was consulted in the online help, and by the time elapsed to make a decision (response latency). The client software update was enforced for all users, each at a random time (for server-side load balancing). The intended user experience was an interruption of anonymous web access. This mimics the distraction from primary tasks that is characteristic for interception dialogs.

Our sample of users of an anonymity tool is certainly not representative. It is biased towards above-average computer literacy and concerns about online privacy: AN.ON users make efforts to install the client software, reconfigure their web browsers, accept a small bandwidth, and differ in attitudes and motivations with regard to online privacy [11]. The bias is not a serious problem, as our results can be interpreted as best-case bounds. If even highly concerned individuals are unresponsive to relevant decisions about their security, then the average user cares even less. By contrast, other research has been criticized of diluting results by including too many indifferent subjects. Sunshine et al. [12] complemented their study with a sample of security experts.

## Experimental Variations

The consent dialog was designed to show a brief message of about 200 words, all accessible without scrolling (Fig. 1). Our variations did not alter the overall semantic of the message, but rather tried to stimulate the systematic or heuristic path with contextual cues. So we modified the perception of the dialog between a typical EULA—which usually coerce into accepting terms to continue using the software—and a truly voluntary decision to participate [5]. The dialog was translated into English and German to cater to the needs of the majority of AN.ON users. The language and the exact appearance depended on the user's operating system and its localization. (The software is portable to all major platforms and its GUI builds on the Java Swing library.) Our results are robust despite this heterogeneity, which reduces the risk of over-interpreting artifacts of appearance details.

### *Variation 1: Heading*

We varied the wording of the consent dialog's heading between a neutral form ("Improvements to the security ...") and a polite request for help ("Please help us to improve the security ..."). The motivation is twofold, and we expect a complex relationship. On the one hand, social-psychological processes of altruism and helping tendency have been confirmed as influencing factors to compliance with survey requests in public opinion [8] and persuasion research [9]. This suggests a *higher* participation rate in the polite ask-for-help condition. However, the above results assume systematic processing rather than a heuristic response to apparent EULA dialogs. Since typical EULAs are coercive, a polite request contrasts with prior beliefs and might stimulate the systematic path. Considering this non-coersive request, users learn that objection is a feasible option suggesting *lower* participation rates in the ask-for-help condition.

### *Variation 2: Button Text*

We also varied the wording of the button text between a version resembling a typical EULA dialog with labels "I accept" and "I decline", and one indicating the voluntary nature of the decision ("I take part" and "I do not take part"). Interceptions require the user to make a decision. The response options are typically displayed as button text and—unlike all other parts of the dialog—must be read to continue. We hypothesize that variations of the button text have a much larger impact than the two other variations. Egelman et al. [4] varied the button text of browser warnings (though confounded with other features) and found that "providing clear choices" positively affects their effectiveness.

### *Variation 3: Default Button*

As a reference to gauge the magnitude of effects, we also varied the default button (i. e., the button that is pre-selected and can be activated with the return key on most platforms) between consent, objection, and no default at all. According to common design guidelines (e. g., Principle 13:14 of [13]), the most frequently used option should be made the default. This serves as a kind of norm or anchor to users, indicating how other users might have decided, and thereby guide their own decision. We hypothesize that default buttons receive more hits than normal buttons.

## RESULTS

Our main data set contains 81,920 user responses. Table 1 shows the marginal participation rates for the three experimental variations. As expected, the button text is most influential: labels that resemble a EULA receive $26.8\,\%$-pts. higher participation rates than those indicating real choice. By contrast, the overall effect of a polite heading is negative ($-1.9\,\%$-pts.). This suggest that the distinction from typical EULAs dominates the hypothesized positive effect of helping tendency. The default button creates an effect of similar magnitude with $5.7\,\%$-pts. between the extremes. Due to the large sample, all effects are highly significant ($p \leq 0.001$) in individual $\chi^2$ contingency tests. Joint significance tests have been run by regressing the experimental variations on the participation rate using a linear model with a logistic link function (see Tab. 2, Model 1).

**Table 1. Individual effect of experimental variations**

| Cue | | Participation rate [%] |
|---|---|---|
| **Heading** | neutral | 48.3 |
| | ask for help | 46.4 |
| **Button text** | "accept" | 60.7 |
| | "take part" | 33.9 |
| **Default button** | no default | 48.3 |
| | consent | 49.8 |
| | objection | 44.1 |
| Total ($N$ = 81,920) | | 47.4 |

**Table 2. Logistic regression coefficients for decision to participate**

| Predictor | Model 1 | Model 2 | Model 3 |
|---|---|---|---|
| *Intercept* [a)] | −0.67 | −0.68 | −0.68 |
| Heading (ask for help=1) | −0.09 *** | −0.09 *** | −0.09 *** |
| Button text ("accept"=1) | 1.10 *** | 1.10 *** | 1.11 *** |
| Def. button (consent=1) | 0.15 *** | 0.15 *** | 0.16 *** |
| Help clicked (yes=1) | | 0.08 | 0.51 *** |
| Heading ∧ Help cl. | | | −0.03 |
| But. text. ∧ Help cl. | | | −0.97 *** |
| Def. button ∧ Help cl. | | | −0.05 |

$N$ = 81,920; [a)] no null hypothesis; significance level: *** $p \leq 0.001$.

A straight interpretation is that even privacy-concerned individuals are very much accustomed to EULA dialogs so that they agree to everything that superficially looks like one. As soon as cues exist which conflict with this prior belief, fractions of users—roughly proportional to the salience of the cue—reconsider and state their true opinion rather than casting a reflexive agreement. To further support this interpretation, we refine the analysis and look at indicators which proxy the individual user's likely path of processing.

### Differentiating Systematic and Heuristic Processing

*Indicator 1: Consultation of Online Help*

A 'Help' button displayed on all dialog variations allowed users to retrieve background information on the anonymity measurement study. Although less than 1 % consulted online help, the absolute number of 733 users is still large enough to analyse this subset. While seeking online help has no direct influence on the participation rate (see Tab. 2, Model 2), it substantially reduces the sensitivity to variations in the button text (see the large and highly significant negative interaction term in Model 3). When we assume that users who read the help page take the systematic path, then the result supports our view that the main effects are largely due to heuristics of users who confuse our dialog with a EULA.

*Indicator 2: Response Latency*

A better indicator for the systematic path is the time needed to complete the dialog. Figure 2 shows the distribution of response latencies. More than 50 % of the users take less than 8 seconds, which is clearly too short to read the entire notice. In the figure and all subsequent analyses, we censor the right tail of the latency distribution to exclude cases

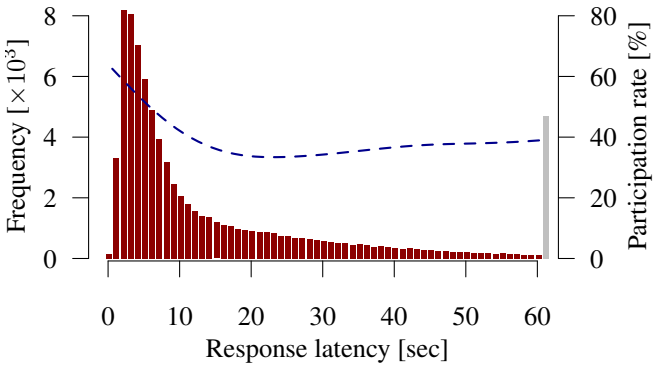


**Figure 2. Histogram of response latency (LHS, rightmost bin open) and participation rate as a function of response latency (RHS, smoothed)**

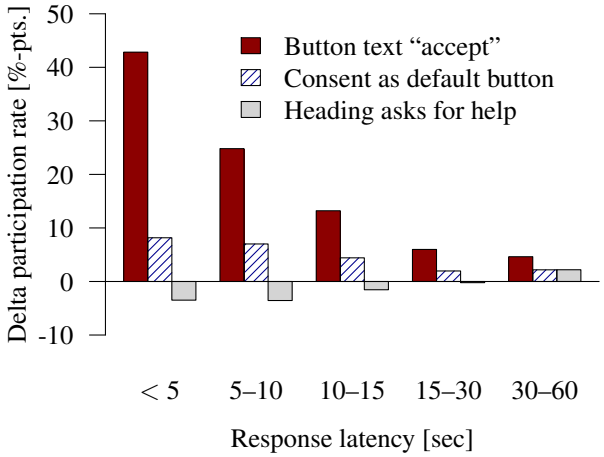


**Figure 3. Sensitivity of participation rate as function of latency**

where users apparently switched to other tasks before completing the dialog. More interestingly, Figure 3 breaks down the relative effect of the experimental variations for different batches of response latencies. The sensitivity to all variations decays with longer latency, so the observed effects are supposedly due to heuristic processing. Moreover, the effect of a polite heading turns positive for users who take the systematic path. This solves the apparent contradiction and corroborates with findings from social psychology and survey research, which assume a primary task and thus systematic processing. All interactions between the cues and response latency are statistically significant (tests omitted for brevity).

### Robustness

We were initially concerned by possible endogeneity of our indicators for the processing path. For example, the decision to consult online help might not be independent of the experimental condition. Similarly, the response latency could depend on cues in the heading or the button labels. However, a check for endogeneity in Table 3 mitigated our concerns: although tendencies of endogeneity are observable, they are relatively small in magnitude. At most one in ten users who clicked help might have done so in reaction to the experimental condition, and the differences in response latency are much smaller than the brackets used in Figure 3. The response to Variation 3 further supports our conjecture that users perceive default buttons as recommendations and are more likely to seek help if they miss one.

**Table 3. Check for endogeneity of indicators for processing path**

| Cue | Help clicked | Response latency mean [a] [sec] | Response latency > 60 sec [%] |
|---|---|---|---|
| **Heading** | | | |
| neutral | 0.9 % | 12.5 | 5.8 |
| ask for help | 0.9 % | 12.1 | 5.6 |
| **Button text** | | | |
| "accept" | 0.7 % | 11.9 | 5.4 |
| "take part" | 1.1 % | 12.7 | 6.0 |
| **Default button** | | | |
| no default | 1.0 % | 12.4 | 5.8 |
| consent | 0.8 % | 12.1 | 5.7 |
| objection | 0.9 % | 12.4 | 5.7 |
| Total | 0.9 % | 12.3 | 5.7 |
| Help clicked | | 19.7 | 31.4 |

[a] robust mean of response latencies: right tail > 60 seconds excluded

**Table 4. Robustness check: Participation rates at independent operators**

| Cue | AN.ON/JonDonym operator [a] CookieCooker | AN.ON/JonDonym operator [a] Speedpartner–ICPP |
|---|---|---|
| **Heading** | | |
| neutral | 49.3 % | 40.3 % |
| ask for help | 45.6 % | 37.2 % |
| **Button text** | | |
| "accept" | 61.5 % | 52.7 % |
| "take part" | 33.5 % | 24.8 % |
| **Default button** | | |
| no default | 47.8 % | 38.7 % |
| consent | 51.3 % | 40.0 % |
| objection | 43.3 % | 37.7 % |
| Total | 47.4 % | 38.8 % |
| | ($N$ = 9,378) | ($N$ = 36,923) |

[a] identifiers are cascade names; consult the AN.ON website for details

Two more operators who use the AN.ON/JonDonym technology agreed to roll out the client update and provide us with data. Each data set contains thousands of cases and can be used to check the robustness of our result in two unrelated samples. Table 4 reports the key facts (cf. Tab. 1). Observe that the effects carry the same sign and are of similar magnitude, although the level of the overall participation rate varies somewhat. We have also repeated all other analyses and were surprised to find qualitatively the same effects. So we are pretty confident about the validity of our results.

## CONCLUDING DISCUSSION

This note complements the existing laboratory studies on interception dialogs by a large-scale field experiment of more than 80,000 real users. The sample size allows to distinguish between systematic and heuristic processing on the individual level. Our approach is also less susceptible to typical issues of laboratory experiments, such as demand characteristics, self-selection bias, obedience to authority, and artificially strong task focus. The latter is crucial for research on interception dialogs, because task focus stimulates systematic processing. Such results must be interpreted with caution in situations where real users tend to use heuristics, as our findings show. On the downside, the field study setup did not allow us to administer exit surveys to learn more about the users' (self-reported) motivation and knowledge.

On the bottom line, we have new evidence supporting the hypothesis that ubiquitous EULAs have trained even privacy-concerned users to click on "accept" whenever they face an interception that reminds them of a EULA. This behavior thwarts the very intention of informed consent [5]. So we are facing the dilemma that the long-term effect of well-meant measures goes in the opposite direction: rather than attention and choice, users exhibit ignorance. Even worse, this ignorance seems to spill over from moderately relevant topics (e. g., EULAs) to more critical ones (online safety and privacy). In the light of our results, a last resort to prevent habituation is economizing consent decisions and thus reserving users' scarce decision capacity for the really important choices. There, interface design still matters, and we call for more field experiments to complement laboratory studies.


## ACKNOWLEDGEMENTS

Thanks are due to Jens Grossklags for valuable comments. The first author was supported by a post-doctoral fellowship of the German Academic Exchange Service (DAAD).


## REFERENCES


1. S. Berthold, R. Böhme, and S. Köpsell. Data retention and anonymity services. In V. Matyas et al., editor, *The Future of Identity in the Information Society*, pages 92–106, Boston, 2009. Springer.
2. S. Chaiken and T. Yaacov. *Dual-process Theories in Social Psychology*. Guilford Press, New York, 1999.
3. R. Dhamija, J. D. Tygar, and M. Hearst. Why phishing works. In *Proc. of CHI*, pages 581–590, Montréal, Canada, 2006. ACM.
4. S. Egelman, L. F. Cranor, and J. I. Hong. You've been warned: an empirical study of the effectiveness of web browser phishing warnings. In *Proc. of CHI*, pages 1065–1074, Florence, Italy, 2008.
5. B. Friedman, P. Lin, and J. K. Miller. Informed consent by design. In S. Garfinkel and L. F. Cranor, editors, *Security and Usability: Designing Secure Systems that People Can Use*, pages 495–522. O'Reilly, 2005.
6. N. Good, J. Grossklags, D. K. Mulligan, and J. A. Konstan. Noticing notice: a large-scale experiment on the timing of software license agreements. In *Proc. of CHI*, pages 607–616, San Jose, CA, 2007.
7. J. Grossklags and N. Good. Empirical studies on software notices to inform policy makers and usability designers. In *Financial Cryptography*, LNCS 4886, pages 341–355. Springer, 2007.
8. R. M. Groves, R. B. Cialdini, and M. P. Couper. Understanding the decision to participate in a survey. *Public Opinion Quarterly*, 56(4):475–495, 1992.
9. J. C. Mowen and R. B. Cialdini. On implementing the door-in-the-face compliance technique in a business context. *Journal of Marketing Research*, 17:453–458, 1980.
10. R. E. Petty and W. B. G. Jarvis. An individual differences perspective on assessing cognitive processes. In N. Schwarz and S. Sudman, editors, *Answering Questions*, pages 221–257. Jossey-Bass, 1996.
11. S. Spiekermann. The desire for privacy. *Int'l Journal of Technology and Human Interaction*, 1:74–83, 2005.
12. J. Sunshine, S. Egelman, H. Almuhimedi, N. Atri, and L. F. Cranor. Crying wolf: An empirical study of SSL warning effectiveness. In *Proc. of the 18th USENIX Security Symposium*, 2009.
13. U. S. Dep. Health and Human Services. *Research-based Web Design & Usability Guidelines*. 2006.
14. T. Vila, R. Greenstadt, and D. Molnar. Why we can't be bothered to read privacy policies. In L. J. Camp and S. Lewis, editors, *Economics of Information Security*, pages 143–153. Springer, 2004.