Ethical Implications and Consequences of Phishing Studies in Organizations – An Empirical Perspective

Marc Busch
AIT Austrian Institute of Technology
Vienna, Austria
marc.busch@ait.ac.at

Yung Shin Van der Sype
KU Leuven
Leuven, Belgium
yungshin.vandersype@law.kuleuven.be

Michaela Reisinger
AIT Austrian Institute of Technology
Vienna, Austria
michaela.reisinger@ait.ac.at

Peter Fröhlich
AIT Austrian Institute of Technology
Vienna, Austria
peter.froehlich@ait.ac.at

Christina Hochleitner
AIT Austrian Institute of Technology
Vienna, Austria
christina.hochleitner@ait.ac.at

Manfred Tscheligi
AIT Austrian Institute of Technology
Vienna, Austria
&
University of Salzburg
Austria
manfred.tscheligi@sbg.ac.at

Abstract
With employees being still the weakest link in organizational information security, phishing studies are becoming increasingly important and are more frequently employed as a research method. Ensuring the validity of results often calls for the use of deception in phishing research. Yet, deception as a research practice has severe ethical implications: researchers and practitioners have to account for possible emotional harm and distress of participants. Unfortunately, empirical data to estimate this potential harm and distress is still rare. In an ongoing study, we are collecting quantitative and qualitative data on emotional and social effects on employees participating in an organizational phishing study. From this data, we will derive guidelines to estimate possible negative effects and suggest interventions for remediation.

Author Keywords
Ethics; Deception in Research; Phishing Studies

ACM Classification Keywords
H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous; K.4.1 [Computers and Society]: Public Policy Issues | Ethics

Introduction

Deception as a research paradigm is a much debated topic in psychology (but also in other fields such as economics)[4]. It is employed whenever there is a reason to assume that knowledge about the intention or goal of a study might bias participants' behavior and/or self-reports of attitudes and behavior. Deception has been shown to be a useful tool to gain scientific knowledge while maintaining the validity of results. Probably the most well-known example is the 'Milgram experiment' [11], where participants were told the experiment was about the influence of punishment on learning success, whereas the real intention was to study the impact of authority on compliance.

Deception in HCI has been mainly discussed from an interaction perspective, when an interaction is deliberately designed to deceive the user, e.g. showing progress bars to cover system delays [14]. What is missing is a discussion of deception and related ethical consequences for HCI studies, for example, how deception affects the emotional and social well-being of participants.

HCI for information security is a hot topic for modern organizations, especially when it comes to data threats caused by phishing mails: as a recent PwC survey[1] shows, employees' low awareness and lack of knowledge is still the largest threat to security in organizations. As a consequence, a notable number of phishing studies focusing on users have been conducted. Phishing studies investigate which factors in- or decrease an individual's susceptibility to fall for scam messages[2], what characteristics constitute a successful phishing message, and which interventions can prevent phishing. To estimate a person's susceptibility to fall for phishing or evaluate the efficacy of phishing interventions in a valid way, researchers send phishing messages to participants and register the responses (e.g., clicking on links, disclosure of information) to that message.

Results of phishing studies are heavily biased when participants know that they will be phished [2,7,8,13], thus deception is an important tool [5] for phishing research. Phishing studies are also relevant and common outside academia, where they are performed as part of organizational penetration testing, with commercial software allowing automated tests of employees' vulnerability to fall for phishing (e.g. with Wombat Security[3]).

Deception has severe ethical implications. Several bodies and institutions published guidelines and best practices (e.g., European Commission [18], American Psychological Association [1]) that suggest what to consider and how to proceed before, within, and after studies in which the use of deception is justified. Ethical guidelines require researchers to asses if there is any harm to the physical and mental well-being of study

[1] http://www.pwc.com/gx/en/issues/cyber-security/information-security-survey.html

[2] mostly via email, but also possible via social media or instant messaging services

[3] https://www.wombatsecurity.com/

participants (before they conduct the study) [7]. Yet in most cases, researchers don't have any empirical basis to estimate possible harm and rely on anecdotal evidence or simply 'guessing'.

Concerning possible harm to participants, only few studies have investigated and discussed effects of participating in experiments and studies involving deception in general [3,6,12,15,17], and even fewer investigated possible effects of participating in phishing studies [10]. Results from studies with student or children samples in non-organizational contexts [12,15] can be transferred only partly to organizational phishing studies. We argue that organizations have to be considered as a special and socially sensitive context: it is possible that participation (and 'performance') in phishing studies can have severe consequences regarding organizational and social climate (e.g., climate of 'mistrust'). This could be caused by the fear of participants that they face negative consequences when their ´performance´ should get public. A consequence could thus be a negatively affected organizational climate.

Additionally, guidelines on how to set up deception [1,18] and phishing experiments [7] need further substantiation and specification as well as empirical grounding to be valid.

Research Questions

There is a lack of empirical data on the consequences of using deception in organizational phishing studies. We close this research gap and investigate the following research questions:

1. What is the emotional (mood) and social (in terms of 'suspiciousness') impact of participating in a phishing study?
2. How does this impact depend on falling for phishing?
3. Are there any educational effects of participating in such a study?
4. If there is significant (negative) emotional or social impact on participants: which interventions can successfully remedy those negative impacts?

Previous Work

Several studies have investigated the effects and perceptions of participating in social psychology experiments involving deception: Boynton et al. [3] conducted a laboratory experiment with students and found that when an experimenter behaved unprofessionally, participants experienced negative emotions. Epley and Huff [6] replicated a standard psychology experiment and found little negative effects caused by the deception itself, but significant negative impact from receiving false negative feedback on performing tasks. Smith and Richardson [15] found that undergraduate students who participated in deception psychology experiments indicated a more positive experience than those who had not been deceived. Noel et al. [12] investigated the perception of deception in pediatric research roughly two years after participation and came to the result that the children experienced the deception experiment as positive. Another study on the effect of deception in studies with children showed that they generally perceived participating in an experiment involving deception as positive [17].

However, these results cannot be directly transferred to phishing studies in organizational 'real world' environments which provide an entirely different context and rely on different kinds of participants. While the aforementioned studies (focusing on deception, not on phishing) were conducted with students or children in laboratory environments, field studies conducted in an organizational setting need to consider the individual and social effects and possible repercussions when results of the phishing studies become available to the participants, their colleagues or their employers.

There is little information regarding the effect of deception in phishing studies. Jagatic et al. [10] conducted a phishing experiment with students in which they set up an online forum to collect participant responses after being phished: some were angry, all of them denied that they themselves fell for the attack, some misunderstood the experimental setting and most misunderstood how the researchers were able to tailor the messages to their context[4]. While this provides a first insight into effects of phishing, a more structured and controlled way to collect reliable data is needed.

Method

In our planned study, we will answer the research questions outlined above by setting up a large scale eMail phishing study in various organizations. Employees (n=1000) will be recruited from a market research panel that advertises a long-term study on 'the use of ICT at the workplace'. This (fake) market research study is the deceptive element in the actual phishing study. We set up this deceptive element to ensure that participants do not know about the real intention of the study from the beginning. We use a double informed consent approach: a first consent will be administered when participants register for the market research panel and agree to participate in the 'ICT at the workplace' research study. This first consent will not include information about the use of phishing messages in the study. After registering on the platform, participants will receive an online survey in which they will be asked about attitudes towards and use of information and communication technologies at the workplace.

Three months after taking this survey, participants will receive a phishing message via email. The phishing message will be contextualized to their company and include a link. If participants 'fall for the phish' and click on the link, they will be forwarded to a Web site which provides debriefing and full disclosure of the goal of the study. It also contains a second, full informed consent which allows participants to withdraw from the study. If participants do not 'fall for the phish' (i.e., do not click on the link), they will receive a separate mail[5] containing the debriefing, disclosure and second informed consent.

Participants who do not withdraw from the study will proceed to the part of the study where we collect the data to answer the research questions.

After being 'phished' (or being informed about the phishing via mail), participants will immediately be prompted with an online questionnaire in which we will

[4] Researchers harvested this data from social network sites

[5] The email will be sent out three days after the phish was sent out, to allow them enough time to read the phishing mail.

ask them about their experience of the deception in the study and how they expect to be affected by the study in the next months. This includes both, their emotional well-being, but also how they react to mails in daily working life (e.g. ´overly suspicious´). We will include standardized indicators of emotional experience, such as PANAS (positive and negative affect schedule) [16]. After one month and after two months, participants will receive further questionnaires in which they will be asked how the study actually affected their emotional well-being and how their attitudes towards emails changed (RQ1,2). Additionally, they will indicate if participating in the study educated them about phishing (RQ3).

About one month after participating in the phishing study, we will employ problem-centric interviews to get a deeper understanding of participants' feelings and thoughts about the study. We will also conduct focus groups, which will especially be used to study the social dynamics and consequences of the phishing study. We will make use of the group dynamics in the focus groups that arise by including both, employees who fell for the phish and employees who did not. Additionally, interviews and focus groups will be mainly used to inform the design of the interventions (RQ4).

We chose a mixed methods approach, in which we combine quantitative surveys with all participants (n=1000) and qualitative methods such as interviews (~n=30) and a minimum of two focus groups (with each ~n=6) with a subset of the full sample. As this research topic in this special context is still underexplored and as we aim at gaining deep insights into feelings and perceptions of participants, our method will be guided by grounded theory [9].

Expected Results and Impact

We expect to establish an empirical grounding for researchers and practitioners to estimate the emotional, social and educational impact of organizational phishing studies on employees. This grounding should serve as a solid basis for ethical considerations of the ´pros´ and ´cons´ of phishing studies and could potentially feed into existing guidelines (e.g., APA or the European Commission) on how to justify the use of deception in organizational phishing studies.

Furthermore, we will establish empirically-grounded suggestions on measures that should be taken after performing a phishing study in an organization. These guidelines will include interventions that could remedy possible negative effects from participating in phishing studies. In future work, we plan to evaluate the effectiveness of these interventions in randomized controlled trials.

In the long term, we would like to contribute by stimulating ethical discussions of deception as a method in HCI studies in general and specifically in organizational security studies.

Acknowledgements

This work is partially funded by the European Union Horizon 2020 program (H2020-DS-2014-1) under grant agreement 653618 (DOGANA – aDvanced sOcial enGineering And vulNerability Assessment Framework).

References

1. American Psychological Association. 2002. Ethical principles of psychologists and code of conduct.
2. Vivek Anandpara, Andrew Dingman, Markus

Jakobsson, Debin Liu, and Heather Roinestad. 2007. Phishing IQ Tests Measure Fear, Not Ability. In *Financial Cryptography and Data Security*. 362–366.

3. Marcella Boynton, David Portnoy, and Blair Johnson. 2013. Exploring the ethics and psychological impact of deception in psychological research. *IRB Ethics and Human Research* 35, April: 7–13.
4. Larry Christensen. 1988. Deception in psychological research: When is its use justified? *Personality and Social Psychology Bulletin* 14, 4: 664–675.
5. Rasha Salah El-din. 2012. To deceive or not to deceive! Ethical questions in phishing research. *BCS HCI 2012 Workshops: HCI Research in Sensitive Contexts: Ethical Consideration*.
6. Nicholas Epley and Chuck Huff. 1998. Suspicion, affective response, and educational benefit as a result of deception in psychology research. *Personality and Social Psychology Bulletin* 24, 7: 759–768.
7. Peter Finn and Markus Jakobsson. 2007. Designing and Conducting Phishing Experiments. *In IEEE Technology and Society Magazine, Special Issue on Usability and Security*: 1–21.
8. Steven Furnell. 2007. Phishing: Can we spot the signs? *Computer Fraud and Security* 2007, 3: 10–15.
9. Barney G Glaser and Anselm Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Wilhelm Fink Verlag, München.
10. Tom Jagatic, Nathaniel Johnson, Markus Jakobsson, and Filippo Menczer. 2007. Social phishing. *Commun. ACM* 50, 10: 94–100.
11. Stanley Milgram. 1963. Behavioral study of obedience. *The Journal of Abnormal and Social Psychology* 67, 4: 371–378.
12. Melanie Noel, Katelynn Boerner, Kathryn Birnie, et al. 2015. Acceptability by parents and children of deception in pediatric research. *Journal of Developmental & Behavioral Pediatrics* 36, 2: 75–85.
13. Malcolm Pattinson, Cate Jerram, Kathryn Parsons, Agata McCormac, and Marcus Butavicius. 2012. Why do some people manage phishing e- mails better than others? *Information Management & Computer Security* 20, 1: 18–28.
14. Lumpapun Punchoojit and Nuttanont Hongwarittorrn. 2015. Research ethics in human-computer interaction. 180–185.
15. Stevens Smith and Deborah Richardson. 1983. Amelioration of deception and harm in psychological research: The important role of debriefing. *Journal of Personality and Social Psychology* 44, 5: 1075–1082.
16. David Watson, Lee Clark, and Auke Tellegen. 1988. Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology* 54, 6: 1063–1070.
17. Carol Weissbrod and Thomas Mangan. 1978. Children's attitudes about experimental participation: The effect of deception and debriefing. *The Journal of Social Psychology* 106, 1: 69–72.
18. European textbook on ethics in research. 2010. https://ec.europa.eu/research/science-society/document_library/pdf_06/textbook-on-ethics-report_en.pdf