To Click or Not to Click? Deciding to Trust or Distrust Phishing Emails

Pierre-Emmanuel Arduin

To Click or Not to Click?
Deciding to Trust or Distrust Phishing Emails

Pierre-Emmanuel Arduin

Université Paris-Dauphine, PSL, DRM UMR CNRS 7088
Place du Maréchal de Lattre de Tassigny, 75775 Paris Cedex 16, France
pierre-emmanuel.arduin@dauphine.psl.eu

Abstract. While the email traffic is growing around the world, such questions often arise to recipients: to click or not to click? Should I trust or should I distrust? When interacting with computers or digital artefacts, individuals try to replicate interpersonal trust and distrust mechanisms in order to calibrate their trust. Such mechanisms rely on the ways individuals interpret and understand information.
Technical information systems security solutions may reduce external and technical threats; yet the academic literature as well as industrial professionals warn on the risks associated with insider threats, those coming from inside the organization and induced by legitimate users.
This article focuses on phishing emails as an unintentional insider threat. After a literature review on interpretation and knowledge management, insider threats and security, trust and distrust, we present a methodology and experimental protocol used to conduct a study with 250 participants and understand the ways they interpret, decide to trust or to distrust phishing emails. In this article, we discuss the preliminary results of this study and outline future works and directions.

Keywords: Insider Threats · Trust · Interpretation · Knowledge Management · Decision-making.

1 Introduction

Technical and externally centred Information Systems security solutions allow the prevention of intrusions [17], the detection of denial of service attacks [68], and the strengthening of firewalls [46]. Nevertheless, the academic literature as well as industrial professionals consider that a predominant threat is neither technical nor external, but human and inside the organization [54, 29, 64, 66]. Such an insider threat may be intentional or non intentional, malicious or non malicious [35, 65, 5].

In fact, according to audit and advisory surveys such as [28], more than 33% of reported cyber-attacks between 2016 and 2018 used phishing, just behind those who used malware (36%). The proportion of insiders among threats increased from 46% in 2016 to 52% in 2018. According to cybersecurity ventures, employees should be trained to recognize and react to phishing emails [43]. More than 90% of successful attacks rely on phishing, *i.e.* emails leading their recipients to interpret and decide to trust them [21]. Recipients are invited, suggested or requested to click on a link, open a document or forward information to someone they should not.

For authors such as [2], security concerns and actual behaviour are disconnected due to the "lack of comprehension". It may be difficult for users to understand security warnings [14], as well as to identify ongoing attacks [20]. In this paper we argue that such understanding difficulties may be studied by focusing on trust and distrust elements users rely on when receiving an email. A logo, a date, a number or an email address, those elements and others are used to decide either an email may be trusted or not.

In the first section of this paper, background theory and assumptions are presented: First, the ways individuals interpret and understand information relying on the knowledge management literature; Second, insider threats and their different categories; Third trust and distrust with a particular focus on individual psychological processes. In the second section of this paper, a study that involved 250 participants is presented: First, the methodology and experimental protocol; Second, a discussion of the preliminary results; Third, a presentation of future works. The overall purpose of this paper is to share observation statements, preliminary results, and future expectations on ways to prevent insider threats by identifying how we decide to trust or distrust phishing emails.

2 Background Theory and Assumptions

In this section, we first draw from the knowledge management literature to present the ways individuals interpret and understand information. Second, we discuss the importance of considering insider threats and their different categories. Third, we expose individual psychological processes leading to trust or distrust.

2.1 On Interpretation and Understanding

To describe the complex interpretation machinery, some authors talk about a "mental model" [23] or a "neural apparatus" [25], a place of chemical reactions that can be analysed. Others believe that interpretation involves above all the socio-individual [67], resulting from our history, a place for expressing a form of intellectual creativity specific to each person. We all act as interpretative agents, information processors interacting with the world that surrounds us through a filter. In fact, this indescribable filter through which we interact with the world may be called an "*interpretative framework*" [63].

Information is transmitted by talking, writing or acting during a *sense-giving* process. We collect data from this information by listening, reading or watching during a *sense-reading* process. *Sense-giving* and *sense-reading* processes are defined by [50] as follows: "Both the way we endow our own utterance with meaning and our attribution of meaning to the utterances of others are acts of tacit knowing. They represent sense-giving and sense-reading within the structure of tacit knowing" [50, p. 301]. When he studied the processes of *sense-giving* and *sense-reading*, [63] highlighted the idea that knowledge was the result of the interpretation by an individual of information.

Information is continuously created during sense-giving processes and interpreted during sense-reading processes. Knowledge can then be:

- *made explicit*, *i.e.* it has been made explicit by someone within a certain context, it is sense-given and socially constructed. Individuals, as well as computers are "information processing systems" [19, p. 9];
- *tacit*, *i.e.* it has been interpreted by someone within a certain context, it is sense-read and individually constructed. Relying on [49]: "We can know more than we can tell".

So that made explicit knowledge is tacit knowledge that has been made explicit by someone within a certain context. It is information source of tacit knowledge for someone else. It is "what we know and can tell" answering to [49] quoted above. Every piece of information can be seen as a piece of knowledge that has been made explicit by someone within a certain context and with their own intentions.

When a person P_1 structures his/her tacit knowledge and transmits it, he/she creates made explicit knowledge, *i.e.* information created from his/her tacit knowledge. A person P_2 perceiving this information and absorbing it, potentially creates new tacit knowledge for him/herself (see Figure 1). Knowledge is the result of the interpretation by an individual of information. This interpretation is done through an interpretative framework that filters the data contained in the information and with the use of pre-existing tacit knowledge [63].

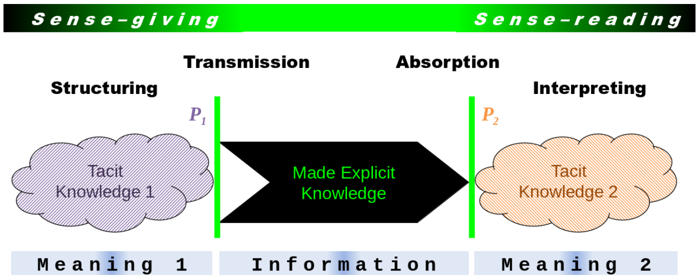


Fig. 1. Sense-giving and sense-reading: the ways we create, interpret and understand information

This interpretation leads to the creation of meaning that can vary from one individual to another: this is *meaning variance* [3, 4]. This question of meaning variance is central in organizations, notably for deciding whether an email may be trusted or not to prevent insider threats.

2.2 On Insider Threats

At the beginning of the 1990s, the literature on information systems security had already affirmed that there was “a gap between the use of modern technology and the understanding of the security implications inherent in its use” [35, p. 173]. The massive arrival of microcomputers was also accompanied by questions regarding the security of interconnected systems where computer science was previously mainframe oriented.

Indeed, the number of technological artefacts has exploded and this increase has gone hand in hand with the evolution of their various uses [9]. Yesterday, a terminal connected the user to the computer, while today entry points into the information system are multiple, universal, interconnected and increasingly discreet. Employee’s social activity can be supported by social networks and their health maintained using connected watches.

The taxonomy of threats targeting the security of information systems proposed by [35] presented in Figure 2 is disturbingly topical, with regard to the four dimensions that make up his angle of analysis: (1) sources, (2) perpetrators, (3) intent, and (4) consequences. It should be recognized that independent of the sources, perpetrators, and intent of a threat, the consequences remain the same: disclosure (of profitable information), modification or destruction (of crucial information), or denial of service (by hindering access to resources). These consequences are covered in the 2013 ISO/IEC 27001 standard: information security management, which defines information security management systems as ensuring (1) confidentiality, (2) integrity and (3) availability of information [22].

A business’s firewall constitutes a protection against external threats, which appear on the left branch in Figure 2. Authors such as [66] represent a part of the literature on information systems security that tends to pay attention to insider threats, more particularly those whose perpetrators are humans with the intention to cause harm (upper right branch in Figure 2).

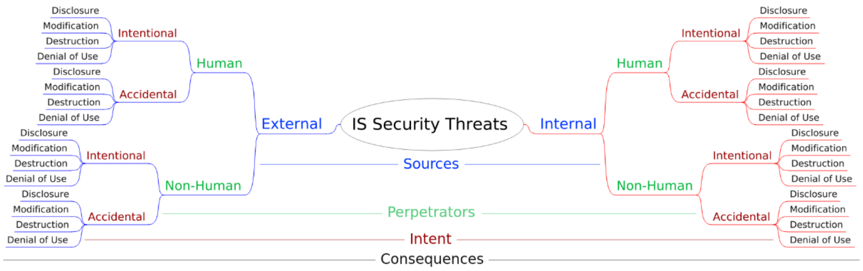


Fig. 2. Taxonomy of IS security threats (inspired from [35])

For authors such as [5], insider threats may be categorized along two dimensions: (1) whether the character of the threat is intentional or not, and (2) whether its character is malicious or not. From the point of view of the employee, who may constitute the entry point into the system, an insider threat can be:

1. *unintentional*: wrong actions taken by an inexperienced or negligent employee, or one manipulated by an attacker; for example, an inattentive click, input error, accidental deletion of sensitive data, etc. [59];
2. *intentional and non-malicious*: deliberate actions by an employee who derives a benefit but has no desire to cause harm; for example, deferring backups, choosing a weak password, leaving the door open during a sensitive discussion, etc. [15];
3. *intentional and malicious*: deliberate actions by an employee with a desire to cause harm; for example, divulging sensitive data, introducing malicious software into the computer system, etc. [58].

The study presented in this article focuses on the manipulation and social engineering techniques that exploit unintentional insider threats (category 1 above). Even though the attacker is outside the system and the organization, he makes an employee, a component of the system, unintentionally facilitate his/her infiltration: the latter has, for example, clicked on a link or even opened the door to a self-proclaimed delivery person with a self-proclaimed task. A social engineer is an attacker who targets a legitimate user from whom he/she obtains a direct (rights of access, harmful link visited, etc.) or indirect (vital information, relationship of trust, etc.) means to get into the system [42].

As new technological solutions are developed, the exploitation of hardware or software weaknesses becomes more and more difficult. Attackers are then turning toward another component of the system susceptible to attack: the human one. For authors as [56]: “Security is a process, not a product”. For others such as [42, p. 14], breaching the human firewall is “easy”, requiring no investment, except for occasional telephone calls and involves minimum risk. Every legitimate user constitutes thus an unintentional insider threat to the information system’s security.

Individuals are not trained to be suspicious of others. Consequently, they constitute a strongest threat to the security of the information system insofar as any well-prepared individual can win their trust.

2.3 On Trust and Distrust

Even if trust is recognized as particularly important in security issues of computer networking environments [26], very little studies on information systems deal with both trust and security [51]. Some authors focus on end-users’ trustworthiness [1, 60] and others on trustworthy information systems’ design [48, 53, 62].

In human and social sciences, authors as [37] and [47] consider that the psychological functionality of trust is to reduce the perceived uncertainty, *i.e.* the perceived risk in complex decision-making situations. Trust induces a mental reduction of the field of possibilities leading to take a decision without considering the outcome of each possible alternative [33].

Some authors consider concepts such as interpersonal trust and organizational trust, between respectively two or more people [13, 24, 57]. Others consider systemic trust, toward institutions or organizations [37], and trust in technologies [30, 39].

An overall definition of trust seems to be lacking when tackling the literature. Relying on the taxonomy of [6], [44, 45] defined trust as expectations: expectation of persistence of the natural and moral social orders, expectation of competence, and expectation of responsibility.

[51, p. 116] proposed an operational definition of trust as a "state of expectations resulting from a mental reduction of the field of possibilities". Such a definition appears to be consistent with the concept of distrust, which is a "confident negative expectation regarding another's conduct" [32, p. 439]. Distrust is often presented as relying on [36] and his suggestion that those who choose not to trust "must adopt another negative strategy to reduce complexity" [27, p. 24]. So, you trust when you have positive expectations, you distrust when you have negative expectations. Distrust should not be confused with mistrust, which is "either a former trust destroyed, or former trust healed" [61, p. 27] and is not considered in the study presented in this is article.

[10, p. 7] went deeper when they stated that "the quantitative dimensions of trust are based on the quantitative dimensions of its cognitive constituents". These constituents are the beliefs on which we rely to trust, and they may explain the contents of our expectations. Examples related to trusted humans are: benevolence, integrity, morality, credibility, motives, abilities, expertise [38, 40]. Examples related to trusted technologies are: dependability, reliability, predictability, failure rates, false alarms, transparency, safety, performance [16, 18, 34, 55].

An appropriate distrust fosters protective attitudes [32] and reduces insider threats. Nevertheless, authors such as [51, p. 118] consider that "trust and distrust are alive, they increase or decrease depending on how expectations are met (or unmet [...])". Initial trust is notably based on information from third parties, reputation, first impressions, and personal characteristics such as the disposition to trust [41]. Then, from facts, understanding of the trustee's characteristics, predictability and limits notably [31, 47], the trustor calibrates his/her trust [45, 11]. Trust is adjusted, meaning expectations are adjusted.

3 Research Proposal and Experimental Protocol

In this section, we first present the methodology and experimental protocol we used to conduct a study with 250 participants and the ways they interpret, decide to trust or distrust phishing emails. Second, we discuss the preliminary results. Third, we outline future works and directions following this work-in-progress.

3.1 Description of the Study

The study has been conducted with 250 students of the Paris-Dauphine university. Half of them were Computer Science students and the other half Management Science students. Half of them were Bachelor students and the other half Master's degree students. Participants were given course credits for participating in the study. In the following we refer to the students involved in the study as the "participants". The average age of participants is 20.2 years.

A research engineer scheduled the presence of participants in a room with 10 computers and copy-protection walls. For the first waves of answers, a member of the research team was here to explain the purpose of the study and answer questions. Then the research engineer continued to manage the response room during one month in order to collect data and he was available to answer questions participants might have.

A short video presented the study to participants who were then given 20 emails, 8 of which were in English. They viewed each email one at a time on the computer screen and they were then asked to: (1) click on the areas leading them to trust it, (2) click on the areas leading them to distrust it, and (3) comment their choices in a general remarks field. Finally, participants should answer some profiling questions (age, academic level, etc.).

Participants arrived in the response room and gave informed consent to participate. Once installed, the researcher asked them if they had any questions. A short video gave them instructions:

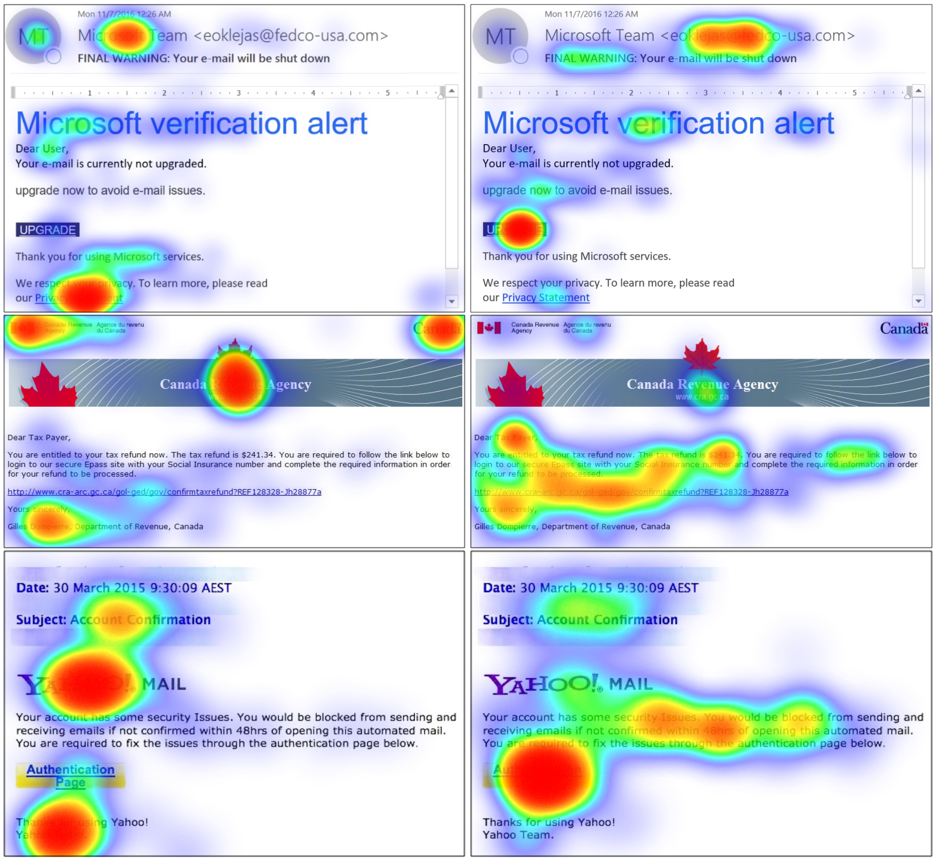
> *This animation will present you the objective of this questionnaire, as well as the perspectives of the study. It will introduce the way you have to understand the questions in order to improve the usefulness of your answers for our investigation. Each question is composed of three parts:*
>
> 1. *click on the areas of the email that make you think that it is official;*
> 2. *click on the areas of the email that make you think that it is fraudulent;*
> 3. *comment in a few words.*
>
> *The results of the first part will allow to identify elements of trust carried by the mail, whereas the results of the second part will allow to identify elements of distrust carried by the mail.*
> *The free text box "comments" allows you every time to explain your feeling.*
> *When you are ready, click "start" below.*

Then participants completed the task, as described above.

3.2 Discussion of the Preliminary Results

Data has been managed by the online reaction time experiments solution Qualtrics (see [7]) notably in order to produce heatmaps [8] such as shown in Figure 3. General remarks fields have been manually tagged by the research team and when interpretation doubts were encountered, participants were contacted to clarify their meaning.

In this article, we consider the subset of 8 English emails in order to restrict the amount of data to process. Thus the collected data represents for each participant and each email, trust and distrust areas, meaning: $(250 \times 8) \times 2 = 4\ 000$ images of emails with clicked areas. These images have been aggregated by the Qualtrics solution to $8 \times 2 = 16$ heatmaps: 8 trust-leading areas images and 8 distrust-leading areas images. Figure 3 shows three examples of trust-leading and distrust-leading areas in emails. As outlined in the next section, the research team is now analysing such images, notably to find invariant elements used as trust or distrust givers by individuals, *i.e.* elements leading participants to decide to trust or to distrust.


Mon 11/7/2016 12:26 AM

MT Microsoft Team <eoklejas@fedco-usa.com>

FINAL WARNING: Your e-mail will be shut down

Microsoft verification alert

Dear User,

Your e-mail is currently not upgraded.

upgrade now to avoid e-mail issues.

UPGRADE

Thank you for using Microsoft services.

We respect your privacy. To learn more, please read our Privacy Statement

Canada Revenue Agency Agence du revenu du Canada Canada

Canada Revenue Agency www.cra.gc.ca

Dear Tax Payer,

You are entitled to your tax refund now. The tax refund is $241.34. You are required to follow the link below to login to our secure Epass site with your Social Insurance number and complete the required information in order for your refund to be processed.

http://www.cra-arc.gc.ca/gol-ged/gov/confirmtaxrefund?REF128328-Jh28877a

Yours sincerely,

Gilles Dompierre, Department of Revenue, Canada

Date: 30 March 2015 9:30:09 AEST

Subject: Account Confirmation

YAHOO! MAIL

Your account has some security Issues. You would be blocked from sending and receiving emails if not confirmed within 48hrs of opening this automated mail. You are required to fix the issues through the authentication page below.

Authentication Page

Thanks for using Yahoo!
Yahoo Team.

(a) (b)

Fig. 3. Examples of trust-leading (a) and distrust-leading (b) areas in phishing emails

We also have to consider that each participant could explain their choices for each email in a general remarks field. Such data represent (250 × 8) = 2 000 open text fields explaining the choices. A preliminary analysis of the open text fields related to the 8 English emails highlights that participants first focused on elements leading them to distrust emails. Even if they were asked to select trust and distrust areas in emails, only 7.6% of the participants mentioned both trust and distrust elements in their written explanations, the rest of them focusing only on distrust elements. This may be a bias of the study, probably induced by the material of the study: phishing emails. The heatmaps generated by the Qualtrics solution allow to partially bypass such a bias by analysing both trust-leading and distrust-leading areas.

Participants who mentioned trust elements in their written explanation listed the presence of privacy concerns (6%) or logos (1.6%) in the emails. 44.1% of the participants mentioned the sender's address as a distrust element in their written comments. 25% of the participants mentioned the presentation of the email, *i.e.* typeface, structure, and spelling, as a distrust element, and 14% of them mentioned the pressure or emergency as a distrust element. Few of them (less than 2%) mentioned the presence of a link (particularly non-HTTPS) or an attachment to download, the occurrence of terms such as "secure", the absence of a human contact or personal data as distrust elements.

3.3 Presentation of Future Works

It is obvious that the results presented in this article are attached to the set of the study, meaning the 250 participants: students in higher education whose average age is 20.2 years. The purpose of the study is not to cover all the trust and distrust mechanisms that individuals put in place when interacting with computers or digital artefacts. The study presented in this article is still in progress and it aims to share observation statements, preliminary results, and future expectations on ways to prevent insider threats by identifying elements individuals rely on when deciding to trust or distrust phishing emails.

By understanding the ways individuals interpret, understand, trust or distrust an email, this study intents to prevent manipulation techniques hackers can use in order to influence individuals' decision to trust. Authors as [12, p. 92] stated that "in most of [the] studies no attempt was made to differentiate between the survey samples drawn from those who intentionally violate the procedures and policies and drawn from those who unintentionally violate them". The case of phishing emails is particularly interesting because existing behavioural countermeasures such as improving awareness, installing a rule-oriented organizational culture, deterrence or neutralization mechanisms show their limits on sloppiness and ignorance [12].

Currently, the research team is going deeper in the analysis of the collected data, working particularly on the overall set and not only the English emails. We aim to observe invariant elements used as trust or distrust givers by individuals. Such elements may be used by hackers as well as by security teams in organizations to adapt their formations and future actions.

As explained in Section 2.2, such a study focuses on unintentional insider threats notably caused by sloppiness or ignorance. In the future, we plan to tackle intentional insider threats, when individuals intentionally violate the information systems' security policy.

4 Conclusions and Perspectives

In this article, we focus on a particular threat for information systems' security: the unintentional and insider threat represented by individuals receiving phishing emails. Such individuals may facilitate the infiltration of an attacker despite themselves, by deciding to trust a phishing email.

In the first section, we presented the ways individuals interpret information relying on the knowledge management literature, the landscape of insider threats and their specificities, trust and distrust mechanisms involved in complex decision-making situations. In the second section, we presented a study conducted with 250 participants in order to highlight trust and distrust leading areas in emails and understand trust and distrust elements used by participants when receiving phishing emails.

The decision to trust, as well as manipulation techniques, were involved in decision-making situations well before the introduction of computers. In this article we studied the ways individuals interpret information to understand how they decide to trust or distrust phishing emails. Ethical issues should not be neglected in such a research, notably by considering the risk of dual-use of research results, as stated by [52]. The reader has to be aware that the results of this research may be used maliciously to mislead recipients.

The study presented in this article is a work-in-progress and the research team is now going deeper by analyzing the overall set of responses. In a near future it is planed to tackle another threat for information systems security: the intentional and insider threat represented by individuals deciding to intentionally violate the information systems' security policy. Such a study will be realized within industrial fields due to its potential managerial causes and implications.

References

1. Aberer, K., Despotovic, Z.: Managing trust in a peer-2-peer information system. In: Proceedings of the tenth international conference on Information and knowledge management. pp. 310–317. ACM (2001)
2. Anderson, B., Bjornn, D., Jenkins, J., Kirwan, B., Vance, A.: Improving security message adherence through improved comprehension: Neural and behavioral insights. In: Americas Conference on Information Systems (AMCIS), 2018. AIS (2018)
3. Arduin, P.E.: On the use of cognitive maps to identify meaning variance. Lecture notes in business information processing **180**, 73–80 (2014)
4. Arduin, P.E.: On the measurement of cooperative compatibility to predict meaning variance. In: Proceedings of IEEE International Conference on Computer Supported Cooperative Work in Design (CSCWD), Calabria, Italy, May 6-8. pp. 42–47 (2015)

5. Arduin, P.E.: Insider Threats. John Wiley & Sons (2018)
6. Barber, B.: The logic and limits of trust. New Brunswick, NJ: Rutgers University Press (1983)
7. Barnhoorn, J.S., Haasnoot, E., Bocanegra, B.R., van Steenbergen, H.: Qrtengine: An easy solution for running online reaction time experiments using qualtrics. Behavior Research Methods **47**(4), 918–929 (2015)
8. Bojko, A.: Informative or misleading? heatmaps deconstructed. In: Jacko, J.A. (ed.) Human-Computer Interaction. New Trends. pp. 30–39. Springer Berlin Heidelberg, Berlin, Heidelberg (2009)
9. Canohoto, A., Dibb, S., Simkin, L., Quinn, L., Analogbei, M.: Preparing for the future – how managers perceive, interpret and assess the impact of digital technologies for business. In: Proceedings of the 48th Hawaii International Conference on System Sciences, Kauai, HI, 2015 (2015)
10. Castelfranchi, C., Falcone, R.: Trust is much more than subjective probability: Mental components and sources of trust. In: Proceedings of the 33th Hawaii International Conference on System Sciences, Piscataway, NJ, 2000 (2000)
11. Costé, B., Ray, C., Coatrieux, G.: Trust assessment for the security of information systems. Advances in Knowledge Discovery and Management **8**, 1–23 (2018)
12. Crossler, R.E., Johnston, A.C., Lowry, P.B., Hu, Q., Warkentin, M., Baskerville, R.: Future directions for behavioral information security research. computers & security **32**, 90–101 (2013)
13. Deutsch, M.: Trust and suspicion. Journal of conflict resolution **2**(4), 265–279 (1958)
14. Felt, A.P., Ainslie, A., Reeder, R.W., Consolvo, S., Thyagaraja, S., Bettes, A., Harris, H., Grimes, J.: Improving ssl warnings: Comprehension and adherence. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. pp. 2893–2902. ACM (2015)
15. Guo, K., Yuan, Y., Archer, N., Connely, C.: Understanding nonmalicious security violations in the workplace: a composite behavior model. Journal of Management Information Systems **28**(2), 203–236 (2011)
16. Hancock, P.A., Billings, D.R., Schaefer, K.E., Chen, J.Y., De Visser, E.J., Parasuraman, R.: A meta-analysis of factors affecting trust in human-robot interaction. Human Factors **53**(5), 517–527 (2011)
17. Hansen, J.V., Lowry, P.B., Meservy, R.D., McDonald, D.M.: Genetic programming for prevention of cyberterrorism through dynamic and evolving intrusion detection. Decision Support Systems **43**(4), 1362–1374 (2007)
18. Hasselbring, W., Reussner, R.: Toward trustworthy software systems. Computer **39**(4), 91–92 (2006)
19. Hornung, B.: Constructing sociology from first order cybernetics: Basic concepts for a sociocybernetic analysis of information society. In: proceedings of the 4th Conference of Sociocybernetics, Corfu, Greece. (2009)
20. Hu, Q., Dinev, T., Hart, P., Cooke, D.: Managing employee compliance with information security policies: The critical role of top management and organizational culture. Decision Sciences **43**(4), 615–660 (2012)
21. Hurley, R.: The decision to trust. Harvard business review **84**, 55–62 (2006)
22. ISO/IEC: Iso/iec 27001, information security management. Technical report (2013)
23. Jones, N., Ross, H., Lynam, T., Perez, P., Leitch, A.: Mental models: An interdisciplinary synthesis of theory and methods. Ecology and Society **16**(1) (2011)
24. Kramer, R.M.: Trust and distrust in organizations: Emerging perspectives, enduring questions. Annual review of psychology **50**(1), 569–598 (1999)

25. Kuhn, T.: Criticism and the Growth of Knowledge, chap. Reflections on my critics. Cambridge University Press (1970)
26. Lamsal, P.: Understanding trust and security. Department of Computer Science, University of Helsinki, Finland (2001)
27. Lane, C., Bachmann, R., Bachmann, L.: Trust Within and Between Organizations: Conceptual Issues and Empirical Applications. Oxford University Press (1998)
28. Lavion, D.: Pwc's global economic crime and fraud survey 2018. Technical report (2018)
29. Leach, J.: Improving user security behaviour. Computers & Security **22**(8), 685–692 (2003)
30. Lee, J.D., See, K.A.: Trust in automation: Designing for appropriate reliance. Human factors **46**(1), 50–80 (2004)
31. Lewicki, R.J., Bunker, B.B.: Developing and maintaining trust in work relationships. Trust in organizations: Frontiers of theory and research **114**, 139 (1996)
32. Lewicki, R.J., Mc Allister, D.J., Bies, R.J.: Trust and distrust: New relationships and realities. Academy of management Review **23**(3), 438–458 (1998)
33. Lewis, J.D., Weigert, A.: Trust as a social reality. Social forces **63**(4), 967–985 (1985)
34. Li, X., Hess, T.J., Valacich, J.S.: Why do we trust new technology? a study of initial trust formation with organizational information systems. Journal of Strategic Information Systems **17**(1), 39–71 (2008)
35. Loch, K.D., Carr, H.H., Warkentin, M.E.: Threats to information systems: today's reality, yesterday's understanding. Mis Quarterly pp. 173–186 (1992)
36. Luhmann, N.: Trust and Power. Wiley, Chichester (1979)
37. Luhmann, N.: Familiarity, confidence, trust: Problems and alternatives. Trust: Making and breaking cooperative relations **6**, 94–107 (2000)
38. Mayer, R.C., Davis, J.H., Schoorman, F.D.: An integrative model of organizational trust. The Academy of Management Review **20**(3), 709–734 (1995)
39. Mc Knight, D.H., Carter, M., Thatcher, J.B., Clay, P.F.: Trust in a specific technology: An investigation of its components and measures. ACM Transactions on Management Information Systems (TMIS) **2**(2), 12 (2011)
40. McKnight, D.H., Chervany, N.L.: Trust and distrust definitions: One bite at a time. In: Trust in Cyber-societies, pp. 27–54. Springer (2001)
41. McKnight, D.H., Chervany, N.L.: Handbook of trust research pp. 29–51 (2006)
42. Mitnick, K., Simon, W.: The Art of Deception: Controlling the Human Element of Security. John Wiley and Sons (2003)
43. Morgan, S.: Cybercrime damages $ 6 trillion by 2021. Technical report (2016)
44. Muir, B.M.: Trust between humans and machines, and the design of decision aids. International Journal of Man-Machine Studies **27**(5-6), 527–539 (1987)
45. Muir, B.M.: Trust in automation: Part i. theoretical issues in the study of trust and human intervention in automated systems. Ergonomics **37**(11), 1905–1922 (1994)
46. Neira Ayuso, P., M. Gasca, R., Lefevre, L.: Ft-fw: A cluster-based fault-tolerant architecture for stateful firewalls **31**, 524–539 (2012)
47. Numan, J.: Knowledge-based systems as companions. Trust, human computer interaction and complex systems. Ph.D. thesis, Groningen, NL (1998)
48. Offor, P.I.: Managing risk in secure system: Antecedents to system engineers' trust assumptions decisions. In: Social Computing (SocialCom), 2013 International Conference on. pp. 478–485. IEEE (2013)
49. Polanyi, M.: Personal Knowledge: Towards a Post Critical Philosophy. Routledge (1958)

50. Polanyi, M.: Sense-giving and sense-reading. Philosophy: Journal of the Royal Institute of Philosophy **42**(162), 301–323 (1967)
51. Rajaonah, B.: A view of trust and information system security under the perspective of critical infrastructure protection. Ingénierie des Systèmes d'Information **22**(1), 109 (2017)
52. Rath, J., Ischi, M., Perkins, D.: Evolution of different dual-use concepts in international and national law and its implications on research ethics and governance. Science and Engineering Ethics **20**(3), 769–790 (2014)
53. Ruotsalainen, P., Nykänen, P., Seppälä, A., Blobel, B.: Trust-based information system architecture for personal wellness. In: MIE. pp. 136–140 (2014)
54. Sasse, M.A., Brostoff, S., Weirich, D.: Transforming the 'weakest link'—a human/computer interaction approach to usable and effective security. BT technology journal **19**(3), 122–131 (2001)
55. Schaefer, K.E., Chen, J.Y., Szalma, J.L., Hancock, P.A.: A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. Human factors **58**(3), 377–400 (2016)
56. Schneier, B.: The process of security. Information Security **3**(4), 32 (2000)
57. Schoorman, F.D., Mayer, R.C., Davis, J.H.: An integrative model of organizational trust: Past, present, and future. Academy of Management Review **32**(2), 344–354 (2007)
58. Shropshire, J.: A canonical analysis of intentional information security breaches by insiders. Information Management and Computer Security **17**(4), 221–234 (2009)
59. Stanton, J., Stam, K., Mastrangelo, P., Jolton, J.: Analysis of end user security behaviors. Computers and Security **24**(2), 124–133 (2005)
60. Swamynathan, G., Zhao, B.Y., Almeroth, K.C.: Decoupling service and feedback trust in a peer-to-peer reputation system. In: International Symposium on Parallel and Distributed Processing and Applications. pp. 82–90. Springer (2005)
61. Sztompka, P.: Trust: A Sociological Theory. Cambridge Cultural Social Studies, Cambridge University Press (1999)
62. Truong, N.B., Um, T.W., Lee, G.M.: A reputation and knowledge based trust service platform for trustworthy social internet of things. Innovations in Clouds, Internet and Networks (ICIN), Paris, France (2016)
63. Tsuchiya, S.: Improving knowledge creation ability through organizational learning. In: ISMICK 1993: Proceedings of the International Symposium on the Management of Industrial and Corporate Knowledge. pp. 87–95 (1993)
64. Vroom, C., Von Solms, R.: Towards information security behavioural compliance. Computers & Security **23**(3), 191–198 (2004)
65. Warkentin, M., Willison, R.: Behavioral and policy issues in information systems security: the insider threat. European Journal of Information Systems **18**(2), 101–105 (2009)
66. Willison, R., Warkentin, M.: Beyond deterrence: an expanded view of employee computer abuse. MIS Quartely **37**(1), 1–20 (2013)
67. Yamakawa, Y., Naito, E.: Cognitive Maps, chap. From Physical Brain to Social Brain. InTech (2010)
68. Zhi-Jun, W., Hai-Tao, Z., Ming-Hua, W., Bao-Song, P.: Msabms-based approach of detecting ldos attack. computers & security **31**(4), 402–417 (2012)