Cyber Security as Social Experiment

Wolter Pieters[*][†]
[*]University of Twente
Services, Cybersec. & Safety
Enschede,The Netherlands
w.pieters@tudelft.nl

Dina Hadžiosmanović[†]
[†]Delft University of Technology
CyberSecurity@TUDelft
Delft,The Netherlands
d.hadziosmanovic@tudelft.nl

Francien Dechesne[‡]
[‡]University of Leiden
eLaw@Leiden
Leiden,The Netherlands
f.dechesne@tudelft.nl

ABSTRACT

Lessons from previous experiences are often overlooked when deploying security-sensitive technology in the real world. At the same time, security assessments often suffer from a lack of real-world data. This appears similar to general problems in technology assessment, where knowledge about (side-)effects of a new technology often only appears when it is too late. In this context, the paradigm of new technologies as social experiments was proposed, to achieve more conscious and gradual deployment of new technologies, without losing the ability to steer the developments or make changes in designs. In this paper, we propose to apply the paradigm of new technologies as social experiments to security-sensitive technologies. This new paradigm achieves *(i)* inherent attention for the ethics of deploying security-sensitive systems in the real world, and *(ii)* more systematic extraction of real-world security data and feedback into decision making processes.

Categories and Subject Descriptors

K.6.5 [**Management of Computing and Information Systems**]: Security and Protection

General Terms

Management, Security

Keywords

adversarial experiments; cyber security; empirical security; feedback; responsible experimentation; security-by-experiment; social experiments; technology assessment

1. INTRODUCTION

Currently, there is a lot of attention for "security-by-design" as a means to secure public values in the cyber age. This would imply that important choices regarding security can already be made in the design stage, with the help of suitably chosen stakeholders. However, history has shown that for safety issues in technological developments, a design-stage approach has not been enough. The complexity of the technological infrastructure and the number of dependencies are simply too high to predict all potential undesired side-effects. Emergent properties may only appear after deployment, depending on how the technology is being used in practice (which may be totally different from theoretical predictions). We see similar issues in the security domain as well (e.g., large-scale deployments of electronic voting and smart electricity meters).

In this paper, we therefore argue for responsible piloting in cyber security, by adopting the paradigm of "new technologies as social experiments" [39] in this context. We propose to consider the *deployment of security-sensitive technologies as social experiments*, in the sense of society being subject to a test with uncertain outcome.[1] We argue that *(i)* the deployment of such technologies is *by definition* a social experiment, and that *(ii)* much can be learnt about real-world security by designing such experiments consciously. This paper is therefore not primarily about experimentation with security solutions (such as cryptographic protocols), yet with technologies in a wider sense in which security controls appear as important building blocks.

In particular, we focus on the *deployment* of such technologies as experiments (Table 1), which has also been termed "societal experiment", and might as well be called "socio-technical experiment". The phrase "cyber security as social experiment" denotes that security is not only a design feature, but also the outcome of an experiment. In this sense, this paper argues that the term "security-by-design" is too optimistic, and that it should be complemented (not replaced) by conscious piloting and gradual deployment to gain additional knowledge. We thus argue for *security-by-experiment* next to security-by-design.

In the next section, we provide a more detailed motivation for our approach. In section 3, we describe the paradigm of new technologies as social experiments in more detail. In section 4, we discuss what is needed to apply the paradigm to cyber security. In section 5, we adapt conditions for responsible experimentation from the original paradigm to cyber security. The benefits of the approach for two cases, electronic voting and smart metering, are discussed in section 6. We end with conclusions and discussion in section 7.

NSPW'14, September 15–18, 2014, Victoria, BC, Canada.
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-3062-6/14/09 ...$15.00.
http://dx.doi.org/10.1145/2683467.2683469.

[1] Of course, the term "security-sensitive technologies" implies a gradual scale, and it may not always be known upfront which technologies are security-sensitive (the same holds for safety).

Table 1: Experiments in cyber security, and the focus of the social experiments paradigm.

Target of experimentation	Experimental approach	Examples
Security controls	Experiments with controls	Cryptanalysis, ...
Security-sensitive technologies	Experiments with technologies	Penetration testing, ...
Deployment of technologies	*Cyber security as social experiment*	Monitoring, feedback, ...

2. MOTIVATION

The main motivation for considering new technologies as social experiments is what is called the Collingridge dilemma [5, 25]. This dilemma states that in the initial stages of the development of an emerging technology, the technology can easily be steered in a desirable direction, but there is a lack of knowledge on potentially undesirable side-effects. By contrast, knowledge becomes available once the technology is deployed in the real world, yet dependencies and stakeholder interests then make it almost impossible to change the development direction.

To address this dilemma, a conscious approach is needed to gradual introduction of new technologies in combination with learning. However, this condition has not been fulfilled in the context of security-sensitive cyber technologies so far. When a new security-sensitive technology is deployed, or when its deployment is scaled up, it is often unclear how to use past experiences to decide on the required level of security controls. There are many cases of security problems reported in an already deployed technology, yet the question remains how to avoid new problems next time. "Security-by-design" is seen as a solution for preventing problems, but does this really mean that we are *learning* cyber security? Doing something by design may still require building upon past experience, and this may not be happening enough.

On the one hand, this is because there is simply not enough data on cyber security risks. We don't really know how to assess what may happen to a technology when deployed. This also causes a lack of empirical validation of theoretical models [41]. The data that is available typically relates to frequencies of "random" attacks such as virus infections, not to adversary behaviour in relation to system design. Data on adversary behaviour may be "predicted" by methods such as game theory, although the results are notoriously hard to validate in real life. In addition, when learning in cyber security is discussed, it is mostly conceived as learning from incidents, not learning from defender decisions.

On the other hand, security decisions are usually conceived as point-in-time events, when a system is first designed or introduced. This complicates matters when devices have long lifespans (such as in industrial control systems). In such cases initial security decisions can become inappropriate (due to a change in security environment over time and the inability to adapt). Devices may also have dynamic properties (which emerge over time) [18].

Instantiations of these problems appear for example in the introduction of various forms of electronic voting in different countries, many of which had problems that were only acknowledged after major efforts by NGOs. New designs do not always reflect lessons that could have been learnt. Similarly, the first generation of Dutch smart electricity meters did not make it through the Senate, as privacy concerns had not been taken into account sufficiently [17]. Even though requirements were updated, new pilots still do not address security. Further back in history, lessons from the design of Roman aqueducts as critical infrastructures were only drawn 2,000 years after the fact [2].

Conceiving security as a social experiment would on the one hand enable systematic data acquisition, and on the other hand improve learning capacity with respect to security decisions. Simultaneously, the notion of social experiment points to moral considerations about responsible design of the experiments. Such interpretations are already available in, for example, nuclear energy [39] and environmental technology [42]. However, the adversarial context of cyber security brings specific challenges, and the requirements for successful application of the paradigm to cyber security are the main topic of the present paper.

3. NEW TECHNOLOGIES AS SOCIAL EXPERIMENTS

The paradigm of "new technologies as social experiments" was proposed by Van de Poel [38, 39] in the context of uncertainties around new technologies, and their potential hazards. Rather than one-off decisions on the acceptability of and conditions for new technologies, he conceives new technologies as an ongoing experiment in society, exposing society to uncertain outcomes. In the course of such an experiment, more knowledge about the effects becomes available, as well as more knowledge about the potential embedding in and adaptation to society. This could help to diminish the problems of the Collingridge dilemma. However, to prevent excessive hazards and irreversibility of the introduction, the experiments should be carefully designed, following certain conditions for responsible experimentation.

For example, in [20], the paradigm is applied to sunscreen with nano-particles. The authors argue, among other things, that certain health and environmental aspects of this technology have not been properly tested in laboratory environments, and that monitoring of effects in the social experiments is not in place. Therefore, the experiment may not be acceptable, and if it were, it could have been designed better to enable more learning.

A distinction can be drawn between new technologies as social experiments, versus piloting new technology. Essentially, the paradigm does not argue that we should organise experiments with new technology. The introduction of a new technology in society is *by definition* a social experiment; the point of the approach is that we need to design such experiments explicitly (e.g. by limited-scale pilots) in order to fulfil conditions of responsible experimentation.

Instead of trying to answer the question about the acceptability of a new technology a priori, Van de Poel argues that new (emerging) technologies need empirical data, and that we should therefore first answer the question to what extent experimentation (in the real world) is acceptable. This sounds indeed similar to the plea for more data to support (quantitative) assessments of cyber security, as well as validations of predictions made by such models [41].

Of course objections can be made to such an approach, as formulated by Peterson [30]. Peterson argues that in this approach, the acceptability question of new technologies is replaced with the acceptability of the social experiments, which is a less informative question, and not easier to answer either. For example, instead of discussing whether electronic voting is acceptable, one would answer the question whether experiments with electronic voting are acceptable. If these experiments would not yield information about the acceptability of the technology, the whole endeavour would be pointless.

In his response to the critique by Peterson, Van de Poel [40] argues that the acceptability question is merely postponed, not replaced. In the meantime, empirical data to support the decision on acceptability can be gathered. In particular, more information becomes available on the direction of development of the new technology, as "the properties and the consequences of these technologies are the emergent outcome of the co-evolution of technology and society". Unforeseen risks, but also unforeseen benefits (e.g. telephone, computers) may become more visible.

To this end, the emphasis of the approach is on an "adaptive learning process". In this process, not only the consequences of the technology can be better anticipated, but also options for increasing the resilience of the socio-technical system against disturbances (caused by the new technology) may become clear. Still, postponing the question on acceptability of the technology demands *reversibility of the experiment*: if the question is postponed but the final answer is negative, the experiment should be stopped. This requires very careful design, and may not always be possible, as some stakeholders may already have embraced the new technology.

4. CYBER SECURITY AS SOCIAL EXPERIMENT

Currently, there seems to be a lack of systematic real-world security evaluation of new information technologies. Whereas functional capabilities of technologies are tested via controlled deployments (e.g., pilot projects of smart grids), security evaluations are often not part of the analysis scope in the running pilots.[2] Therefore, we think the paradigm of new technologies as social experiments has added value for cyber security. In particular, when security aspects of the experiments are designed consciously, ethical problems may be avoided, and more data could become available for learning.

4.1 Advantages of the paradigm

There are several advantages of conceiving cyber security as a social experiment. The experimental context can provide real-world data that may be hard to get in brainstorms or simulations, for example: *(i)* new threats that emerge from the environment, such as new and unpredicted forms of use of a technology; *(ii)* security/usability tradeoffs that are decided upon by the actual users, such as choosing simple passwords; *(iii)* weaknesses in the environment, such as insecure storage of security-sensitive equipment; *(iv)* stakeholder perspectives and values based on the actual use of the technology.

In addition, new vulnerabilities often appear at the boundaries of existing classes. From a security-by-design point of view, one may search for known types of vulnerabilities, but it is hard to find the unknown ones, those that transcend existing (cultural) classification systems [4, 6, 34]. In such a context, complete design-time security may be impossible due to the limitations of human perception and imagination. Furthermore, whereas old-fashioned voting machines were designed once and then deployed, new technologies are often offered as a service [32], or have other options for real-time reconfiguration [10]. When providers may change designs on the fly, design-time assessments may simply not be sufficient.

Finally, the paradigm allow us to investigate an additional aspect: the transition between two types of technology. For example, if a new (and maybe more secure) method for online banking is introduced, this may also create incentives for cyber criminals to make use of the unfamiliarity of the customers, for example by faking the new procedures, or sending phishing e-mails referring to the new system.

4.2 The missing link

The context of cyber security brings an issue which has not been discussed in the paradigm so far: *involvement of adversaries*. Adversaries are needed in order to cause problems, and such problems may not even be visible to the ones who run the experiment. We thus need to account for specifics of security in terms of adversarial experiments[3].

Regarding existing applications of the social experiments paradigm, cyber security relates to experiments with nuclear energy. This technology may provide opportunities for adversaries as well, for example by the possibility to steal nuclear material and create dirty bombs. This is sometimes called dual use, in the sense of technologies that can be used for both beneficial and harmful purposes [14]. In the context of cyber security, cyber attacks may be conceived as dual use of the security-sensitive technology: harmful use of a technology that also provides benefits.

Still, also in the nuclear case, the existing literature does not touch upon the adversarial aspects, yet limits itself to (long-term) safety issues related to nuclear facilities and nuclear waste. Thus, even when nuclear technology is discussed [39], the emphasis is mostly on unintentional effects rather than the potential for misuse (dual use). One could argue that in the nuclear context, the health hazards make it hard to account for dual use in the social experiments (the concept of "ethical hacker" would be more difficult to implement), which could be a reason why existing work does not address this topic. Therefore, the adversarial aspects of new technologies as social experiments still need to be developed from this perspective.

4.3 Experiments in cyber security

To identify the requirements for designing adversarial aspects into social experimentation with new technologies, existing experimental approaches in cyber security may provide inspiration. Structured experiments with cyber security are becoming more common in the literature. These exper-

[2] In a different paper [9], we discuss in detail that current smart grid pilots do not systematically analyse cyber security.

[3] The term adversarial experiment is used in a different sense in technical security approaches, e.g. [15].

iments are mostly focused on user behaviour in relation to security technology, and aim at gathering quantitative (statistical) data on security-related behaviour, such as vulnerability of people to phishing attempts [13, 21]. In addition, more extensive approaches for penetration testing, including physical access and social engineering [12, 16], also point in this direction. Finally, honeypots, apart from deflecting attackers, also provide an experimental source of data on attacks on real-world systems [23].

The existing experiments in cyber security typically only focus on a small part of a technology, for example the security of password-based access control in the context of phishing attempts. In the present paper, we argue for experiments with a wider target of analysis, namely the new technology as socio-technical system, including the operational context (see Table 1). Such experiments may include penetration testing, but should also focus on higher-level properties in terms of stakeholder attitudes and potential misuse scenarios. Both external stakeholders as well as insider threats should be considered.

4.4 Dealing with adversaries

One of the main questions in cyber security experimentation is how to give shape to the adversarial aspects of the experiment. On the one hand, simply waiting for an adversary to attack the system is not efficient. On the other hand, having defenders brainstorm about possible attack does not take full advantage of the experimental context. Therefore, the adversarial roles should somehow be designed into the pilots, as is done in phishing experiments and penetration testing. Like in these approaches, some actors need to asked to behave maliciously, but now not only in relation to the technology (which could be tested in the lab), but also in relation to the deployment environment.

In this respect, existing knowledge on serious gaming in security (red/blue team exercise) could provide inspiration [26]. In this way, one could try “gaming the new technology”. Model-based approaches in security may also have a place in such games, aiming at predicting possible attack scenarios, which can then be used by the adversarial roles. Examples are adversary course of action [37], adversary view security evaluation [24], or attack navigators [35].

Thus, social experimentation for cyber security requires adversarial roles in the experiment, demanding an adaptation of the original paradigm. Designing an experiment for learning about functionality or safety is not the same as designing an experiment for learning about security. This has consequences for the criteria for responsible experimentation, as we will discuss in the next section.

5. CONDITIONS FOR RESPONSIBLE EXPERIMENTATION

In the original paradigm, Van de Poel [38] proposes conditions for responsible experimentation. In particular, he proposes four preconditions under which social experiments are justified, and four requirements for such experiments. These conditions are meant to balance the ability to learn from the experiment against potential negative effects for stakeholders and society. In the following, we re-interpret these conditions in the context of cyber security. There are several different versions of this list, and the enumeration is not necessarily exhaustive.

5.1 Preconditions

Absence of alternative testing methods.

One of the main conditions for acceptability of the social experiment is the absence (or exhaustion) of alternative testing possibilities. In particular, there needs to be a clear indication that the experimental introduction into society provides information that cannot be obtained before performing the social experiments.

In a cyber security context, the usual practices of software review, software testing and penetration testing should be applied before deployment, as well as early involvement of stakeholders (security-by-design). However, there may still be uncertainty about interactions with the threat context, social context, and other technologies, which can only be reduced by performing real-life experiments, and cannot be derived in common testbed environments (e.g., to understand aspects of alternative forms of use, or user acceptability of security policies).

Controllability.

In social experiments, it is generally infeasible to achieve the degree of control that is possible in a laboratory. Controllability refers both to control of the parameters and control of the effects. This condition requires that the effects of the social experiment should be limited to the technology that is the target of experimentation, and not cause disruptions in other vital services or functions.

In a cyber security context, controllability of parameters primarily points to controlling adversary behaviour. One would have to assign explicit adversarial roles to make this aspect controllable. In terms of controllability of effects, the experiments should ideally not affect existing services. For example, experiments with smart meters should not lead to power cut-off if a vulnerability turns out to exist in the meters. Ideally, such meters should be installed in parallel to conventional ones.

Proportionality of hazards to benefits.

Due to the uncertainty around new and emerging technologies, it is generally impossible to do a meaningful cost-benefit analysis, even a probabilistic one. The required probabilities associated with the predicted benefits, and in particular possible risks, are typically too uncertain. Available alternatives for assessment are generally pessimistic (worst-case) approaches that focus on risks only, and ignore potential benefits. According to Van de Poel, one should focus both on *expected* benefits and *credible* threats when estimating the proportionality of hazards to benefits. This gives a bit of advantage to threats over benefits (credible versus expected), but still allows accounting for both. What should be conceived as credible threats depends on the context, and fully elaborated threat scenarios will generally not be available.

In a security context, one may for example not be able to identify concrete attack scenarios against an electronic voting system, but one can be certain that there will be stakeholders who are interested in manipulating the election result. A credible hazard is therefore that the result will be tampered with. On the other hand, an expected benefit is that electronic counting may be faster and more accurate. To balance the proportionality of hazards to benefits, one could design a manual recount option, as well as a procedure to determine what would happen in case of deviations. Another question is how the costs and potential hazards of gaining new knowledge by experiment are weighed against the costs of implementing known controls. The latter strategy can be regarded both safer (because relying on known concepts and methods) and riskier (because not taking possible new threats into account). In particular, there is the question of how to optimise defender investment under uncertainty (see e.g. [3, 35]). If the defender can invest either in known controls or in reducing uncertainty (by means of pilots), which strategy is better? And what about externalities in terms of possible negative effects on society in case vulnerabilities are overlooked? There is a tradeoff here between investing in security controls and investing in experimentation.

Informed consent.

Informed consent involves acceptance of the subjects of the conditions of an experiment. To this end, the subjects are supplied with information about the goals and methods of the experiment. For example, subjects in a medical or psychological experiment will be briefed about the treatments or tests administered, and will agree to these methods before they participate. If this is not possible, for example because knowledge of the goal of the experiment would invalidate the results, additional debriefing should take place afterwards. As the introduction of new technologies is conceived as an experiment, this requirement also holds in this context. This means that those who are subject to the pilot, both users of the new technology and indirect stakeholders, should receive information and agree with conditions. It is important to provide information on a meaningful level of abstraction [33].

For social experiments with cyber security, the adversarial roles are key in informed consent. In particular, information about the role to play in the experiment should be made available to stakeholders, in combination with associated attributes, e.g. a “get-out-of-jail” card to prevent those playing adversarial roles from being prosecuted for their actions. However, not all stakeholders can have information about all roles, as this would compromise the benefits of the adversarial aspects. For example, if I already know that my neighbour is supposed to manipulate his smart meter, I may react differently, thereby interfering with meaningful observations. Therefore, debriefing may be needed in a security context, as knowledge about the existence of adversarial roles in the experiment could change behaviour [12].

5.2 Conditions for set-up

The above conditions describe under which circumstances social experiments with new technologies would be acceptable (in contrast to laboratory experiments). Next to conditions for acceptability of social experiments, Van de Poel also proposes conditions for the *design* of the experiments, if deemed acceptable. These conditions describe how the experiments should be organised to maximise learning and minimise harm.

Monitoring.

Monitoring the experiment is an essential precondition for being able to learn from the experiment. Surprises happening in the experimental context should be identified as early as possible, to determine whether action needs to be taken based on the new knowledge.

In a cyber security context, the monitoring primarily includes the use of technical tools for host and network monitoring (e.g., application logs, network event systems). The main focus of experimentation in cyber security is understanding the adversarial aspect. Therefore, the deployment should be monitored on a higher level, to understand stakeholder actions and decisions (and thus improve the knowledge about adversary behaviour). Implications for privacy of participants should be taken into account in the decision on how to monitor the experiment, as well as possible effects of awareness of monitoring on behaviour of participants.

Feedback.

Information obtained from the experiment (by monitoring) may also provide new insights in the design of the experiment. This way, when necessary, the design can be adjusted. In particular, there should be a possibility of intervention if something takes the experiment into an undesirable direction. This intervention can consist of stopping the experiment and abolishing the new technology altogether, or of adjustments to the technology and/or the experiment. In addition, effort should be spent on analysing the lessons for future experiments.

In a security context, knowledge about security weaknesses will influence the risks. By contrast, in a safety context, knowledge about a potential hazard will not increase the likelihood of occurrence. In a security context, the feedback mechanisms should therefore take this aspect of knowledge into account, *under the assumption that knowledge obtained in the experiment will also be known to adversaries.* Without this assumption, the experiment would continue under a security-by-obscurity condition.

Conscious scaling-up.

The larger the scale of the experiment, the harder it becomes to decide that it was not worth the effort. Investments by stakeholders and dependency of users on the new technology increase the dependency on the chosen path. Therefore, decisions to extend the experiment should be taken with these considerations in mind, and only after some level of knowledge about the technology has been obtained. One strategy for consciously scaling up is what is called strategic niche management [22]. In such an approach, one selects a specific niche for the new technology, such that effects of failing security will have limited impact. An example of such an approach is Internet elections only for citizens living abroad. As the number of registered expat voters is limited, the effect on the election result can never be larger than that number of votes. Of course, even within this group a specific selection of voters can be made, lowering the potential impact even further.

In security, particular attention has to be paid to the fact that larger-scale systems also become more attractive targets for attackers. In the Internet voting case, before scaling up, the question should be asked what knowledge the experiment has yielded so far about the vulnerability to attacks when the technology becomes a more attractive target.

Containment of hazards.

Containment of hazards involves adaptations to the technology itself as well as adaptations to the environment to reduce the likelihood of harm caused by the experiment. One should also be prepared for additional adaptations during the experiment, based on feedback mechanisms. Although Van de Poel argues that containment mechanisms may be seen as decreasing the possibility for learning, proper monitoring can indicate when harm would have occurred without the containment mechanisms.

In a cyber security context, back-up solutions and redundancy may help to contain hazards to some extent. For example, one may experiment with a new method of voting next to standard ballots. However, if the result obtained with the new method does not contribute to the official result, or can be falsified by a paper trail, there are no incentives for adversaries to manipulate the system. This also means that those adversaries will not be of any help in revealing security issues.

5.3 Adjusting the set-up for cyber security

Earlier we observed that the involvement of adversaries creates a different case for social experimentation. As we cannot count on real adversaries to cooperate, we introduce adversarial roles, yet at the same time we do not rule out activity of real adversaries.

In this context, there is one important aspect for discussion. When we enable learning for cyber security by means of social experiments, it may not only be the defenders who learn, *but also the adversaries*. These can be external adversaries, or participants in the experiments who have been assigned adversarial roles and then misuse what they learnt later. This is the same issue that we considered when letting students play adversarial roles [11]. In other words, aren't we providing adversaries with malicious ideas when experimenting with security-sensitive technologies, designing adversarial aspects into such experiments, and trying to learn from those? One could argue that hiding such information from adversaries would amount to security by obscurity anyway, but still the experiment may require an additional condition, stating that learning by potential adversaries should be minimised in the design of the experiment.

We propose to add an additional condition for set-up, called "responsible adversaries". This would include careful assignment of adversarial roles, and specific (ethical) training for those who play such roles.

Responsible adversaries.

When adversarial roles are designed into a social experiment, care should be taken to avoid undesirable effects of such an assignment. In particular, *(i)* a set of criteria should be established on who is eligible for adversarial roles, *(ii)* the adversaries should have anonymous means of raising questions and concerns during the experiment, and *(iii)* the adversaries should participate in an evaluation session afterwards. Finally, as it should not be assumed that the assigned roles are the only adversaries active in the experiment, a discussion on who else might have acted as an adversary and/or obtained information that would enable this in the future (NGOs, but also criminal organisations) should be part of the evaluation.

6. EXAMPLES

In the previous section, we have discussed conditions that ensure a more conscious deployment of security-sensitive technologies in society, when conceived as social experiments. In this section, we present two examples of technologies, of which the introduction *could have been conceived as a social experiment*: electronic voting and smart metering in the Netherlands. We will investigate the benefits the paradigm could have had (or could still have) for these cases. More cases that could be studied in future research can for example be found in [27].

6.1 Electronic voting

An example of different approaches to deployment of a new technology is the introduction of electronic voting in the UK and the Netherlands [19, 31, 36]. In the Netherlands, the decision on whether to adopt electronic voting machines was made by the local authorities. The voters could not choose themselves (and therefore there was no *informed consent*). In the UK, where explicit pilots were conducted, the voters could choose between different channels (e.g. paper, Internet, SMS). This provides the opportunity to gather data about user choices between different technological options, and reasons for those. Of course, this way of experimenting does not account for the lack of visibility of security to the voter (information asymmetry). Still, the experiment could provide some data on stakeholder choices between available technologies, and questions about reasons for such decisions could be included in a survey as part of the experiment.

Here, we focus on the Dutch case. In the Netherlands, electronic voting machines had been introduced since the early 1990s. After the introduction of the machines, regulations were not revisited or updated, nor was there any renewed evaluation of the risks. When a pressure group finally took up an adversarial role (white-hat) in 2006, it was relatively easy to demonstrate potential weaknesses, such as replacement of chips with the counting software, or eavesdropping on the choice of the voter. If such adversarial roles would have been explicitly designed into the experiment, the issues could have been identified much earlier.

There was no *monitoring* and *feedback* in the experiments. In particular, there was no independent recount option. Such an option (e.g. a paper trail) would allow *containment of hazards* and *controllability*: if there were any doubts about the integrity of the electronic result, there would be a backup. Such a paper trail could have been designed earlier if the introduction of the machines would have been conceived as a social experiment. In addition, systematic comparisons between paper and machine counts could have provided indications of deviations (and thereby reliability). Of course, real adversaries would have less interest in manipulating results if there is a paper backup. In terms of *proportionality*, it was unclear whether a less secure technology is acceptable in a lower threat context. For example, should Dutch voting machines also be acceptable when they would be deployed in unstable countries (an argument of the pressure group)?

Small-scale experiments with Internet voting were conducted as well, for expats and for water board (regional water authority) elections in two districts. Apart from the general acceptability of the technology, these pilots looked more like real experiments: they were small-scale, used different technologies, and had proper evaluations. However, the experiments with Internet voting suffered from the failures in the voting machine "experiment". Although there was clearly room for improvement in the technology based on the pilots, the lack of proper experimentation with the electronic voting machines gave the Internet voting experiments a bad reputation too. In addition, the plans for *scaling up* of the Internet voting system to all water boards may have come too soon. Shortly after the voting machines were abolished, the Internet voting experiments were halted as well. At a much later stage, when paper voting was the norm again, experiments with machine-countable paper ballots were conducted. However, before the evaluation reports were available, the new minister set up a commission for looking into electronic voting machine solutions once more. This was seen as a completely different option, and thereby learning opportunities were missed (lack of *feedback*).

Concerning *responsible adversaries*, the risk of eavesdropping on the vote may have been partly "created" by the pressure group. When they demonstrated that an antenna would allow you to find out somebody's vote, there was a real risk of others following up on this. In 2013, the new commission concluded that this risk was now acceptable, but without argumentation based on the earlier experiences [28]. This example points to the need for proper handling of the discovery of new threats by adversarial roles in the experiment, as well as the need for conscious *scaling up*. If the voting machines wouldn't have been used all over the country, the risk of eavesdropping would be much lower, and there would be more opportunities for controlling the risk by additional measures, or even stopping the experiment.

6.2 Smart metering

The EU Directive on energy efficiency (2006/32/EG) prescribed the installation of energy meters that provide end users with information on their actual use, so that they can contribute to energy savings. In the Netherlands, a combination of two separate legal bills was proposed in 2008, amounting to the mandatory roll-out of smart meters, which were still to be developed. These meters were to send measurements for gas (hourly), and electricity (quarter-hourly) to the *network operators*, who would forward this information to the *energy providers*, who could then inform the *consumers* about their consumption. The initial proposal also included signalling functions (to detect energy quality), switching functions (to remotely switch off in case of non-payment or disasters) and regulatory functions (to add options to the meter) for the *network operators*. In fact, some energy providers had already started to provide households with smart meters upon request (e.g. Oxxio from 2005).

After the assessment by the Dutch privacy watchdog, the proposal was amended by requiring explicit consumer consent for transferring detailed consumption data to energy suppliers (however, daily usage would be mandatorily forwarded). Also, addition of purpose specification and use limitation, data subjects' right of access, data removal after use, and suitable security measures were required according to the Dutch privacy law [7]. The October 2008 report by the Consumer Union concluded that smart meters also put pressure on the right to inviolability of the home, and the right to respect for family life [17].

On the basis of this analysis, the Dutch Senate rejected the bills in 2009, and adopted an adapted version, that allowed for conducting pilot projects (*Proeftuinen*) involving smart meters. Users, providers and operators experimented with smart grids and meters on a voluntary basis, in selected neighbourhoods. The aim was trying out incentivising users to conserve energy, and to participate in balancing the grid under the presence of sustainable energy sources whose production depends on weather conditions [7]. In the meantime, to resolve the issues raised by the Senate, a broad stakeholder collaboration came to define the so-called *Dutch Smart Meter Requirements* (DSMR4), that implement the adapted version of the bill (in particular, it specifies the last point: defining data granularity for each task). The abolition of the detailed readings was considered to take "the largest privacy sting out of the Dutch law" [7].

Interestingly, the pilot projects have, as far as we have found [8], not been exploited to explicitly experiment with the effectiveness of the new requirements with respect to the security and privacy issues that were raised when the law was rejected. The pilots are mostly focused on testing the functionality of the technology, and learning how to deal with human participation in balancing the grid. Questions about privacy and security, and associated requirements and values, were not asked to the consumers. Currently (early 2014), a new proposal for the broad smart meter roll out in the Netherlands, on a voluntary basis, is waiting for parliament approval.

User participation in the electricity net is a great paradigm shift, both for users and operators. Experience shows that wrong assumptions are easily made about tasks, responsibilities and risks with respect to (cyber) security. For example, operators are used to thinking in top-down controllable components, which made them neglect privacy issues for consumers, while users are not used to be conscious about the electricity flow, let alone to adapt their behaviour – they need incentives. The pilots that are conducted provide a good opportunity for both sides to learn in a relatively controlled environment how roles in the system may shift, and what that would mean for the risks and responsibilities with respect to cyber security. The lack of explicit attention to (cyber) security and privacy in the smart grid pilots [8] leaves room for reflection on how the pilots could have been used to learn about these aspects for smart metering, by consciously designing them as social experiments.

In the controlled environment of the pilots, the normal regulations with respect to roles and responsibilities for non-functional requirements can be taken a bit more liberally, in order to discover how regulations for the energy system should be adapted for the future system. In the current pilots, some liberty is taken with respect to privacy concerns in order to experiment with the technology (while privacy concerns are not in the focus of the pilots at all). In particular, this means that more data is collected than foreseen in the privacy requirements. It can be questioned whether it is a good idea from a *proportionality* perspective to ignore security and privacy requirements in the pilot environments.

The existing efforts seem to miss an opportunity with respect to experimenting with the non-functional properties of the system (which contrary to technological properties, has no alternative testing). The pilots could be a valuable source of *feedback* also with respect to the requirements (DSMR4), for example user experiences with the granularity of data collection. Furthermore, neither consumers nor operators know exactly what the potential benefits for adversaries are with the smart meters, nor what the damage for the consumers or the system can be. Explicit attention to this issue, for example in the form of adversarial roles, would provide both operators and consumers with valuable knowledge and understanding about potential incentives and consequences (risk) with respect to cyber security and privacy.

Table 2: An overview of conditions and examples

Condition	Adversarial aspects	E-voting	Smart metering
Absence of alternatives	Proper security testing before experimenting; lack of knowledge of real-life threats	Proper (physical) penetration tests on the equipment	Proper (physical) penetration tests on the equipment
Controllability	Controlling adversary behaviour	Assign adversarial roles instead of waiting for NGOs to intervene	
Proportionality	Account for changing threat contexts	Acceptability in different threat contexts	Make privacy rules in pilots the same as in real deployment
Informed consent	Defining roles; debriefing	Channel selection by voter	Make sure that all stakeholders learn from pilots
Monitoring	Technical monitoring; stakeholder behaviour	Statistical checks against independent second channel	
Feedback	Assume that feedback is also known to adversaries	Wait for evaluations before making decisions	Organise stakeholder security evaluations
Conscious scaling-up	Larger-scale systems more attractive to adversaries	Compare against similar technologies	Don't define scaling requirements a priori
Containment of hazards	Back-up and redundancy	Recount option	
N/A	Responsible adversaries	Social construction of eavesdropping	

Such knowledge would also improve *informed consent* for the consumers.

Finally, smart meters are rolled out in pilots first, but new houses come with smart meters by law. Given this development, it would be irresponsible to wait with assessing cyber security and privacy issues until the roll-out has reached large scale. As cyber security hazards of smart meters may increase dramatically with the coverage, one has to try to learn about vulnerabilities before they can no longer be contained. As this puts (too much) time pressure on fixing security, requiring all new houses to be equipped with smart meters violates the condition of *consciously scaling up*.

7. DISCUSSION & CONCLUSIONS

In this paper, we propose to apply the paradigm of "new technologies as social experiments" to cyber security. The use of the paradigm could lead to more systematic efforts to gather data about cyber security in new technologies, including socio-technical and human aspects, and would also enable more systematic learning from security design decisions made in the past. We have discussed its application to the cases of electronic voting and smart metering. An overview is presented in Table 2.

Rather than performing separate experimentation with cyber security technology, this paradigm conceives the introduction of new technologies itself as a social experiment. This provides a higher-level view on security experiments, integrating the adversarial roles from existing security testing approaches such as penetration testing. Based on the conditions for responsible experimentation, sufficient testing in the lab would still need to be done before piloting the technology in the real world. Nonetheless, real-world conditions would allow for gathering additional security-relevant information, such as experiences of stakeholders and the possible exploitation of weaknesses in the deployment environment. In an even broader sense, cyberspace and cyber security are themselves social experiments (which have not been very consciously designed).

Many details of how to devise the experiments have not been covered in this position paper, and are thus open for discussion and future research. Acknowledging the challenges, we foresee several interesting discussion topics: quantitative methods, cultural and institutional aspects, responsibilities in the experiment, and system interdependencies.

Most importantly, the term "experiment" seems to resonate with a quantitative social science approach, with large numbers of participants and controlled conditions. Also, longitudinal studies (observing a phenomenon over a longer period of time) might be applicable. However, the original meaning of the term within the paradigm was meant to convey exactly the opposite, namely uncertainty about effects, and the ability to learn from unexpected observations. Conditions under which pilots take place may not always be suitable for a quantitative approach, and qualitative methods may provide more in-depth knowledge. So one topic for discussion is to what extent "cyber security as social experiment" can claim to be an experiment in the sense of the social sciences, and if not, whether there is a better term to convey the meaning.

Secondly, cultural aspects may be very important in determining under which conditions new technologies as social experiments are acceptable, and under which conditions they provide benefits in terms of safeguarding values. For example, some cultures may take a more precautionary approach than others, and also public-private participation may take different shapes. In essence, the design being experimented with and being evaluated is neither a technology nor a policy alone, but a combination of those (socio-technical system). Technology acceptance models may provide inspiration for taking such aspects into account, both in terms of acceptance of the technology and in terms of acceptance of the experiments / pilots, in relation to perceived risk (see e.g. [1]). In order to make sure that conditions are consistently applied, institutions may be set up for deciding on and monitoring pilots, similarly to those focused on medical experiments. National cyber security centres may play a role here.

Thirdly, we foresee potential challenges in clarifying responsibilities. One would need to think about offering compensation of damages if people suffer negative consequences in the social experiment. For example, if one's smart meter is disabled remotely, either by real adversaries or by adversarial roles, one would like to have some form of compensation, depending on the duration. An important question is of course who pays for this. The discussion on social experiments could also be politicised in different way. For example, opponents of the technology may hijack weaknesses found in the experimental context to condemn the technology altogether. On the other hand, explicitly announcing the technologies as social experiments may also be a way of de-politicising the discussion by attempting open dialogue. In this sense, the proposed paradigm also has similarities to the open data and open government movements.

Fourthly, social experiments, even when carefully designed, cannot cover everything. Complex interdependencies in the infrastructures may cause events which are very hard to capture. For example, attackers may use unexpected attack vectors that aren't observed because they take place outside the scope of the experiment, or manage to evade monitoring efforts. Alternatively, security threats may already have acted in the supply chain rather than after deployment. If persistent changes to the infrastructure are made in these ways, it may take years before the attack is noted, if ever.

Finally, the social experiments paradigm may also be applicable to security paradigms (ways in which to evaluate the technologies / experiments from a security perspective), such as security metrics used. Next to evaluating the technologies, also the security paradigms themselves may be evaluated when deployed in social experimentation [29].

Acknowledgments

The authors wish to thank Neelke Doorn, Sean Peisert, Ibo van de Poel and Shannon Spruit for their support of initial ideas for investigating this paradigm. The anonymous reviewers were of great help in highlighting concerns and topics for discussion. The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement ICT-318003 (TRESPASS). This publication reflects only the authors' views and the Union is not liable for any use that may be made of the information contained herein. The research also received funding from the Center for Safety and Security of the Leiden, Delft and Erasmus universities.

8. REFERENCES

[1] L. AlAbdulkarim, E. Molin, Z. Lukszo, and T. Fens. Acceptance of ict-intensive socio-technical infrastructure systems: Smart metering case in the netherlands. In *Networking, Sensing and Control (ICNSC), 2014 IEEE 11th International Conference on*, pages 399–404, April 2014.

[2] M. J. Assante. Infrastructure protection in the ancient world. In *System Sciences, 2009. HICSS '09. 42nd Hawaii International Conference on*, pages 1–10, 2009.

[3] R. Böhme and T. Moore. The iterated weakest link. *Security & Privacy, IEEE*, 8(1):53–55, 2010.

[4] J. Cappos, Y. Zhuang, D. Oliveira, M. Rosenthal, and M. K.-C. Yeh. Vulnerabilities as blind spots in developer's heuristic-based decision-making processes. In *Proceedings of the 2014 New Security Paradigms Workshop*, NSPW '14, New York, NY, USA, 2014. ACM.

[5] D. Collingridge. *The social control of technology.* St Martin, New York, 1980.

[6] J. R. Crandall and D. Oliveira. Holographic vulnerability studies: Vulnerabilities as fractures in interpretation as information flows across abstraction boundaries. In *Proceedings of the 2012 New Security Paradigms Workshop*, NSPW '12, pages 141–152, New York, NY, USA, 2012. ACM.

[7] C. Cuijpers and B.-J. Koops. Smart metering and privacy in europe: Lessons from the dutch case. In S. Gutwirth, R. Leenes, P. de Hert, and Y. Poullet, editors, *European Data Protection: Coming of Age*, pages 269–293. Springer Netherlands, 2013.

[8] F. Dechesne. (Cyber)security in smart grid pilots. http://tinyurl.com/pm4a43o, December 2013.

[9] F. Dechesne, D. Hadžiosmanović, and W. Pieters. Experimenting with incentives: Security in pilots for future grids. *IEEE Security & Privacy*, 12(6), Nov.-Dec. 2014.

[10] F. Dechesne, M. Warnier, and J. Van den Hoven. Ethical requirements for reconfigurable sensor technology: a challenge for value sensitive design. *Ethics and Information Technology*, pages 1–9, 2013.

[11] T. Dimkov, W. Pieters, and P. Hartel. Training students to steal: A practical assignment in computer security education. In *Proceedings of the 42Nd ACM Technical Symposium on Computer Science Education*, SIGCSE '11, pages 21–26, New York, NY, USA, 2011. ACM.

[12] T. Dimkov, A. van Cleeff, W. Pieters, and P. Hartel. Two methodologies for physical penetration testing using social engineering. In *Proceedings of the 26th Annual Computer Security Applications Conference*, pages 399–408, New York, NY, USA, 2010. ACM.

[13] P. Finn and M. Jakobsson. Designing ethical phishing experiments. *Technology and Society Magazine, IEEE*, 26(1):46–58, Spring 2007.

[14] J. Forge. A note on the definition of "dual use". *Science and Engineering Ethics*, 16(1):111–118, 2010.

[15] P. Golle, M. Jakobsson, A. Juels, and P. Syverson. Universal re-encryption for mixnets. In T. Okamoto, editor, *Topics in Cryptology – CT-RSA 2004*, volume 2964 of *Lecture Notes in Computer Science*, pages 163–178. Springer Berlin Heidelberg, 2004.

[16] H. Hasle, Y. Kristiansen, K. Kintel, and E. Snekkenes. Measuring resistance to social engineering. In R.H. Deng, F. Bao, H. Pang, and J. Zhou, editors, *Information Security Practice and Experience*, volume 3439 of *Lecture Notes in Computer Science*, pages 132–143. Springer, 2005.

[17] R. Hoenkamp, G. B. Huitema, and A. J. C. de Moor-van Vugt. The neglected consumer: The case of the smart meter rollout in the Netherlands. *Renewable Energy Law and Policy Review*, 4:269–282, 2011.

[18] N. Husted and S. Myers. Emergent properties & security: The complexity of security as a science. In

Proceedings of the 2014 New Security Paradigms Workshop, NSPW '14, New York, NY, USA, 2014. ACM.

[19] B. Jacobs and W. Pieters. Electronic voting in the Netherlands: From early adoption to early abolishment. In A. Aldini, G. Barthe, and R. Gorrieri, editors, *Foundations of Security Analysis and Design V*, volume 5705 of *Lecture Notes in Computer Science*, pages 121–144. Springer Berlin Heidelberg, 2009.

[20] J. F. Jacobs, I. Van de Poel, and P. Osseweijer. Sunscreens with titanium dioxide (TiO2) nano-particles: A societal experiment. *NanoEthics*, 4(2):103–113, 2010.

[21] M. Jakobsson, P. Finn, and N. Johnson. Why and how to perform fraud experiments. *Security & Privacy, IEEE*, 6(2):66–68, March 2008.

[22] R. Kemp, J. Schot, and R. Hoogma. Regime shifts to sustainability through processes of niche formation: The approach of strategic niche management. *Technology Analysis & Strategic Management*, 10(2):175–198, 1998.

[23] C. Kreibich and J. Crowcroft. Honeycomb: Creating intrusion detection signatures using honeypots. *SIGCOMM Comput. Commun. Rev.*, 34(1):51–56, January 2004.

[24] E. LeMay, M. D. Ford, K. Keefe, W. H. Sanders, and C. Muehrcke. Model-based security metrics using adversary view security evaluation (ADVISE). In *Quantitative Evaluation of Systems (QEST), 2011 Eighth International Conference on*, pages 191–200, 2011.

[25] W. Liebert and J. C. Schmidt. Collingridge's dilemma and technoscience. *Poiesis & Praxis*, 7(1-2):55–71, 2010.

[26] J. Mirkovic, P. Reiher, C. Papadopoulos, A. Hussain, M. Shepard, M. Berg, and R. Jung. Testing a collaborative DDoS defense in a red team/blue team exercise. *Computers, IEEE Transactions on*, 57(8):1098–1112, Aug 2008.

[27] G. Munnichs, M. Schuijff, and M. Besters, editors. *Databases: The promises of ICT, the hunger for information, and digital autonomy.* Rathenau Instituut, The Hague, 2012.

[28] Commissie onderzoek elektronisch stemmen in het stemlokaal. Elke stem telt: Elektronisch stemmen en tellen. `http://tinyurl.com/nkg5m2s`, December 2013.

[29] S. Peisert, M. Bishop, L. Corriss, and S. J. Greenwald. Quis custodiet ipsos custodes?: A new paradigm for analyzing security paradigms with appreciation to the roman poet juvenal. In *Proceedings of the 2009 New Security Paradigms Workshop*, NSPW '09, pages 71–84, New York, NY, USA, 2009. ACM.

[30] M. B. Peterson. New technologies should not be treated as social experiments. *Ethics, Policy & Environment*, 16(3):349–351, 2013.

[31] W. Pieters. *La volonté machinale: understanding the electronic voting controversy.* PhD thesis, Radboud University Nijmegen, January 2008.

[32] W. Pieters. On thinging things and serving services: technological mediation and inseparable goods. *Ethics and Information Technology*, 15(3):195–208, 2013.

[33] W. Pieters. Explanation and trust: what to tell the user in security and AI. *Ethics and information technology*, forthcoming.

[34] W. Pieters and L. Consoli. Vulnerabilities and responsibilities: dealing with monsters in computer security. *Journal of Information, Communication and Ethics in Society*, 7(4):243–257, 2009.

[35] W. Pieters, C. W. Probst, S. Lukszo, and A. L. Montoya Morales. Cost-effectiveness of security measures: A model-based framework. In T. Tsiakis, T. Kargidis, and P. Katsaros, editors, *Approaches and Processes for Managing the Economics of Information Systems*, pages 139–156. IGI Global, Hershey, PA, USA, 2014.

[36] W. Pieters and R. van Haren. Temptations of turnout and modernisation: E-voting discourses in the UK and The Netherlands. *Journal of Information, Communication and Ethics in Society*, 5(4):276–292, 2007.

[37] E. Santos, Jr. and A. Negri. Constructing adversarial models for threat/enemy intent prediction and inferencing. In *Enabling Technologies for Simulation Science VIII*, volume 5423 of *Proc. SPIE*, pages 77–88, 2004.

[38] I. Van de Poel. The introduction of nanotechnology as a societal experiment. In S. Arnaldi, A. Lorenzet, and F. Russo, editors, *Technoscience in Progress. Managing the Uncertainty of Nanotechnology*, pages 129–142. IOS Press, 2009.

[39] I. Van de Poel. Nuclear energy as a social experiment. *Ethics, Policy & Environment*, 14(3):285–290, 2011.

[40] I. Van de Poel. Why new technologies should be conceived as social experiments. *Ethics, Policy & Environment*, 16(3):352–355, 2013.

[41] V. Verendel. Quantified security is a weak hypothesis: A critical survey of results and assumptions. In *Proceedings of the 2009 New Security Paradigms Workshop*, pages 37–50, New York, NY, USA, 2009. ACM.

[42] H. Verheul and P. J. Vergragt. Social experiments in the development of environmental technology: a bottom-up perspective. *Technology Analysis & Strategic Management*, 7(3):315–326, 1995.