This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSE.2017.2782813, IEEE Transactions on Software Engineering

The Good, the Bad and the Ugly: A Study of Security Decisions in a Cyber-Physical Systems Game

Sylvain Frey, Awais Rashid, Pauline Anthonysamy, Maria Pinto-Albuquerque, and Syed Asad Naqvi

Abstract—Stakeholders' security decisions play a fundamental role in determining security requirements, yet, little is currently understood about how different stakeholder groups within an organisation approach security and the drivers and tacit biases underpinning their decisions. We studied and contrasted the security decisions of three demographics – security experts, computer scientists and managers – when playing a tabletop game that we designed and developed. The game tasks players with managing the security of a cyber-physical environment while facing various threats. Analysis of 12 groups of players (4 groups in each of our demographics) reveals strategies that repeat in particular demographics, e.g., managers and security experts generally favoring technological solutions over personnel training, which computer scientists preferred. Surprisingly, security experts were not *ipso facto* better players – in some cases, they made very questionable decisions – yet they showed a higher level of confidence in themselves. We classified players' decision-making processes, i.e., procedure-, experience-, scenario- or intuition-driven. We identified decision patterns, both good practices and typical errors and pitfalls. Our game provides a *requirements sandbox* in which players can experiment with security risks, learn about decision-making and its consequences, and reflect on their own perception of security.

Index Terms—Security decisions; security requirements; game; decision patterns.

1 Introduction

The security of any system is a direct consequence of stakeholders' decisions regarding security requirements and their relative prioritization. Such decisions are taken with varying degrees of expertise in security. In some organizations – particularly those with resources – these are the preserve of computer (or information) security teams. In others – typically smaller organizations – the computing services team may be charged with the responsibility. Often managers have a role to play as guardians of business targets and goals. Be it common workplace practices or strategic decision making, security decisions underpin not only the initial security requirements and their prioritization but also the adaptation and evolution of these requirements as new business or security contexts arise.

However, little is currently understood about how these various demographics approach cyber security decisions and the strategies and approaches that underpin those decisions. What are the typical decision patterns, if any, the consequences of such patterns and their impact (positive or negative) on the security of the system in question? Nor is there any substantial understanding of how the strategies and decision patterns of these different groups contrast. Is security expertise necessarily an advantage when making security decisions in a given context? Answers to these questions are key to understanding the "how" and "why" behind security decision processes.

In this paper, we propose a tabletop game – Decisions and Disruptions (D-D)[1] – as a means to investigate these very questions. The game tasks a group of players with managing the security of a small utility company while facing a variety of threats. The game provides a requirements sandbox in which players can experiment with threats, learn about decision making and its consequences, and reflect on their own perception of risk. The game is intentionally kept short – 2 hours – and simple enough to be played without prior training. A cyber-physical infrastructure, depicted through a Lego® board, makes the game easy to understand and accessible to players from varying backgrounds and security expertise, without being too trivial a setting for security experts. The particular setting of a utility infrastructure is drawn from our prior experience of technical [26], [27] and non-technical investigations [28] as well as interviews with security experts, field engineers, IT users in such settings [29].

Our work complements existing work on gamification as a means to improve security awareness, education and training [9], [13]. While there is a definite educational and awareness-raising aspect to D-D (as noted consistently by players in all our subject groups), our focus in this paper is on contrasting the security decisions of the three demographics as they manifested in the game sessions. Existing work, e.g., [20], has demonstrated such use of games as an effective

- *S. Frey is with University of Southampton, Southampton, UK. E-mail: s.a.f.frey@soton.ac.uk*
- *A. Rashid is with University of Bristol, UK. E-mail: awais.rashid@bristol.ac.uk*
- *P. Anthonysamy is with Google. E-mail: anthonysp@google.com*
- *M. Pinto-Albuquerque is with Instituto Universitário de Lisboa (ISCTE-IUL), Lisboa, Portugal. E-mail: maria.albuquerque@iscte-iul.pt*
- *P. Anthonysamy and S.A. Naqvi are with Lancaster University, Lancaster, UK. E-mail: {p.anthonysamy, s.naqvi}@lancaster.ac.uk*

1. Game rules available at: http://decisions-disruptions.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSE.2017.2782813, IEEE Transactions on Software Engineering

means to study decision processes of diverse stakeholders. Specifically, the tangible, physical board enables players to manipulate security features and observe the consequences of their decisions. Recording and analysis of these discussions and interactions provides a rich data source to study their decision strategies, processes and patterns.

We report on insights gained from playing D-D with 43 players divided into homogeneous groups (group sizes of 2-6 players): 4 groups of security experts, 4 groups of non-technical managers and 4 groups of general computer scientists. Such observations should, of course, not be generalized, however, the substantial sample size enables in-depth qualitative analysis. Our analysis reveals a number of novel insights regarding security decisions of our three demographics:

- ***Strategies:*** Security experts had a strong interest in advanced technological solutions and tended to neglect intelligence gathering, to their own detriment: some security expert teams achieved poor results in the game. Managers, too, were technology-driven and focused on data protection while neglecting human factors more than other groups. Computer scientists tended to balance human factors and intelligence gathering with technical solutions, and achieved the best results of the three demographics.
- ***Decision Processes:*** Technical experience significantly changes the way players think. Teams with little technical experience had shallow, intuition-driven discussions with few concrete arguments. Technical teams, and the most experienced in particular, had much richer debates, driven by concrete scenarios, anecdotes from experience and procedural thinking. Security experts showed a high confidence in their decisions – despite some of them having bad consequences – while the other groups tended to doubt their own skills – even when they were playing good games.
- ***Patterns:*** A number of characteristic plays could be identified, some good (balance between priorities, open-mindedness and adapting strategies based on inputs that challenge one's pre-conceptions), some bad (excessive focus on particular issues, confidence in charismatic leaders), some ugly ("tunnel vision" syndrome by over-confident players). We document and discuss these patterns, showing the virtue of the positive ones, discouraging the negative ones, and inviting the readers to do their own introspection.

The rest of this paper is structured as follows. In Section 2, we situate our work with respect to the literature on security decisions and security games. Section 3 presents D-D, its game model and rules. This is followed by a description of our subject groups and the analysis approach used to study their security decisions in Section 4. Section 5 presents the strategies that drove decisions of various groups. Section 6 discusses whether groups' (and particular demographics') approaches were procedure-, experience-, scenario- or intuition-driven. Section 7 presents decision patterns – the Good, the Bad and the Ugly – i.e., the patterns that yield better results than others and the clear mistakes and pitfalls to be avoided. Section 8 discusses threats to validity and limitations of our study. Finally, we discuss the possibilities offered by D-D beyond this particular experiment (Section 9).

2 Related work

2.1 Security decisions

A key challenge faced by any organization is the need for optimal investment in security with respect to the threats it is likely to face. Consequently, balancing various factors, such as costs, against potential threats and their likelihood is a key concern. One of the initial metrics for measuring computer related risks was the Annual Loss Expectancy (ALE), which was developed by the U.S.- National Bureau of Standards in 1975 [17]. The ALE is an annual expected financial loss to an organization's information assets because of a particular threat occurring within that same calendar year. Several information security investment decision support methods have been proposed, e.g., [6], [12], some of them based on the ALE metric, e.g., [21]. Within D-D we aim to capture this *realism* of security decision-making—balancing priorities between various threats and investments is a key element in the gameplay.

Bodin et al. [5] introduced the PCR (Perceived Composite Risk) approach. They used the Analytic Hierarchy Process (AHP) to weight and combine different risk measures into a single composite metric for risk analysis. This composite metric supports decision-makers by capturing and balancing the various risk measures that apply to their organization. Baker et al. [4] proposed an event-chain risk management model in which threats are "measured as rates per year and then converted into outcomes by specifying the number or extent per year." While these works focus on providing decision support tools, the focus of our paper is to study and contrast such decisions between different demographics (with varying levels of expertise and knowledge in security).

Research has also demonstrated that a better integration is necessary between business and security perspectives. Corriss [8] has shown that management usually considers information security governance as under the jurisdiction of their information technology department, separate from corporate governance. Coles-Kemp et al. [7] have highlighted the importance of relating security and business risk. They showed that many businesses do not have the tools to relate security risks to business risks and objectives, and that the use of a facilitator can help them understand and better communicate security risks and help embed security management into business practices. Similarly, Anderson [1] notes that an important step to protect an organization's data is for managers to promote security awareness by "creating a culture where the community has the knowledge (what to do), skill (how to do it), and attitude (desire to do it)". These works demonstrate the value of improving security awareness, education and training while bridging corporate and security cultures. The insights resulting from our study are an important contribution in this direction.

2.2 Games

Games have been used as research tools in various domains. Space Fortress [10] is a video game that was developed by cognitive psychologists to understand human cognition and performance. The famous Tetris game has been used to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSE.2017.2782813, IEEE Transactions on Software Engineering

investigate the differences in the actions humans perform from a cognitive perspective [15].

Military organizations and cyber security companies have developed games to improve security awareness and education. Although the form may differ (tabletop, role playing, video game), they are based on a narrative similar to D-D: players are placed in an immersive environment where they must take decisions which require balancing business and technical constraints. Examples include CyberCIEGE [13]: a video game created by the US Naval Postgraduate School that tasks players with managing an IT organization, with the goal of maintaining user productivity while investing resources in necessary security protections against various attack scenarios. The Kaspersky Industrial Protection Simulation is another example: a board role playing game that defines itself as a "Security Monopoly" for maximizing enterprise revenue while building an industrial security capability and dealing with unexpected cyber events despite uncertain information and limited resources.

Similar to model games such as miniature war games, D-D provides players with a physical replication of the context in appreciation: players are able to visualize and manipulate elements of the infrastructure, which facilitates immersion. Lego®, in particular, has been used as a support for redesigning the organization of an industrial facility [24] and for teaching engineering principles [3].

Games have been used in education settings in numerous knowledge areas. Examples in software engineering include teaching software processes [18], [19], value-based software engineering [14], software process risk management [25], and requirements engineering good practices [23]. In [20], a jigsaw puzzle-based game is used to perform analysis and resolution of conflicts among stakeholders, showing how a game can involve players into activities usually considered boring or technical.

In the security domain, Beckers et al. [2] propose a game to capture specific security requirements – in their case pertaining to social engineering. In contrast, D-D focuses on enabling stakeholders to manipulate security features and observe the consequences of their decisions, leading to an improved understanding of both security risks and the trade-offs resulting from particular decisions. Control-Alt-Hack [9] is a tabletop card game where players take on the role of white-hat hackers. The evaluation suggests that Control-Alt-Hack represents an effective model for disseminating ideas and encouraging interest in computer security. Although the primary focus of this paper is on contrasting security decisions of various groups, D-D also has a high potential for educational purposes as discussed in Section 9.

3 D-D: THE RULES OF THE GAME

3.1 Overview

D-D is meant to be played by a team of 2 to 6 players, under the direction of a Game Master. The players act as the team in charge of cyber security in a small utility company, with the goal of minimizing security incidents. The Game Master enforces the rules and guides the players through the 4 rounds of the game. Each round – equivalent to 2 months in game time – is composed of the following steps:

1) The Game Master describes the state of the company and the different systems in the infrastructure (cf. section 3.2).
2) The players are given a budget ($100,000) and a number of possible defenses to invest in, such as firewalls, antivirus, threat assessment (cf. section 3.3).
3) The players debate which defenses are more appropriate and decide by consensus where to invest their budget.
4) The Game Master tells the players about the effects of their investments: whether their defenses deflect any attacks, and the effects of undefended attacks (cf. section 3.4). In addition to technical consequences, the share price of the company can be affected by successful attacks.

3.2 The game board

The game board represents the players' infrastructure (cf. Figures 1 and 2). It is composed of two parts: the field site (or plant) and the office. The field site is where the industrial process takes place. A couple of water turbines are controlled by a SCADA controller, operated by local technicians and engineers. A set of PCs used by local personnel and a database collecting production data sit on the field site's network. This local network is connected to the Internet in order to send strategic information to the office network, where the CEO, a part of the engineering team, and the human resources sit. The office network also hosts a number of PCs, as well as a server and a database: the company runs its own email service and website locally.

3.3 Defenses

Each round, the players are given a budget of $100,000 plus any unspent money from the previous round. Initially, the players can choose to invest among the defenses shown in Table 1. When they choose to invest in an Asset Audit, the Game Master uncovers additional defenses shown in Table 2.

It is explained upfront to the players that the costs of defenses in D-D do not reflect the actual costs of such defenses in practice. Instead, they are designed so that, with a $100,000 budget per turn and a standard price of $30,000 per defense, players can invest in 3 defenses each turn on average. Players must therefore prioritize their choices, hence enabling the study of their security decision strategies and processes. CCTV and Network Monitoring cost more ($50,000 each) to give these defenses an aura of "advanced technology", that we later use to measure how much players are attracted to sophisticated security solutions (cf. section 5). Threat Assessment and Encryption cost less ($20,000) for the sake of game balance, as they are perceived as less powerful than other defenses.

3.4 Attacks

Each round, the Game Master runs a number of attacks against the players' infrastructure, inspired by real-world threat reports [16], [22] and subsequently validated by the security experts who played the game. Attacks are carried out by three categories of attackers:

- *Script kiddies* using basic attacks (scans, DoS, phishing, server exploits) on public targets (the company web server and email addresses).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSE.2017.2782813, IEEE Transactions on Software Engineering

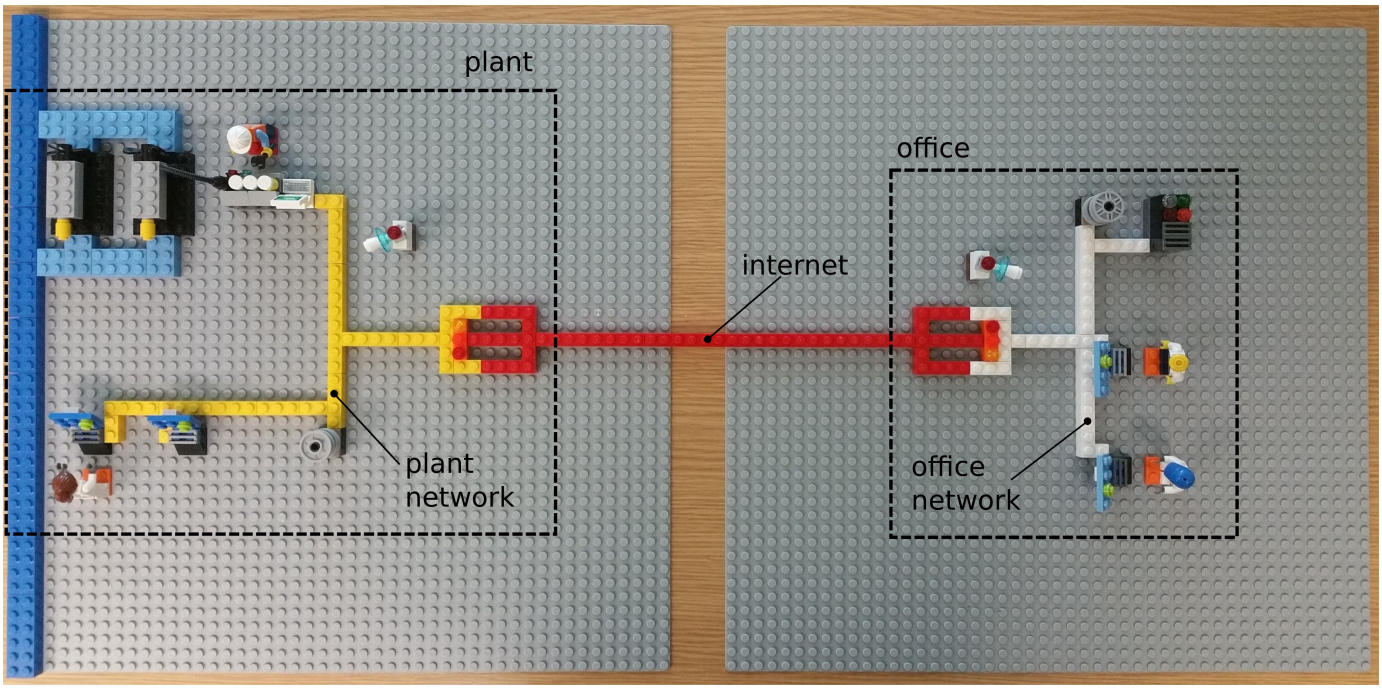

Fig. 1. Overview of the game board.

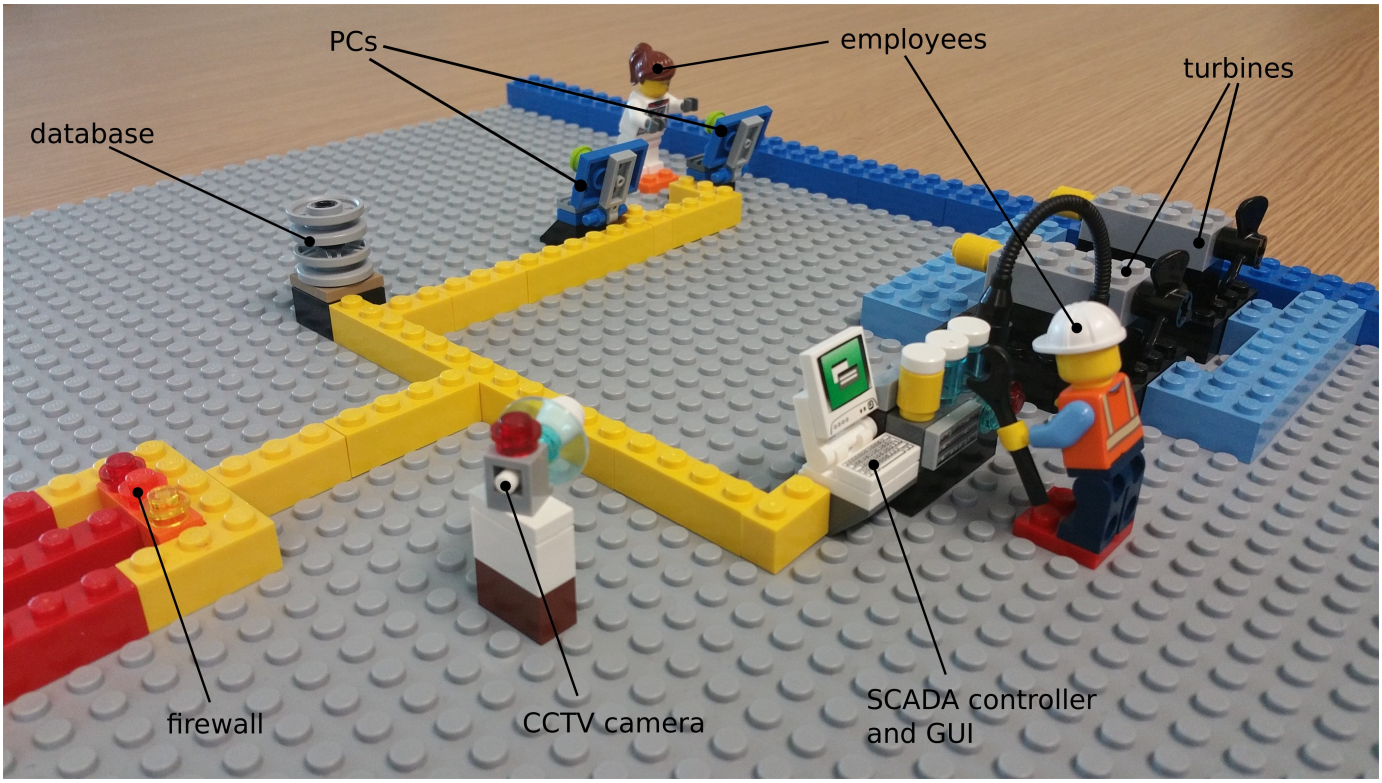

Fig. 2. The plant.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSE.2017.2782813, IEEE Transactions on Software Engineering

TABLE 1
Initial defenses available to the players.

CCTV - plant ($50,000)	Surveillance cameras and alarms that will automatically warn security guards of a physical intrusion in the plant.
CCTV - offices ($50,000)	Surveillance cameras and alarms that will automatically warn security guards of a physical intrusion in the offices.
Network monitor - plant ($50,000)	An advanced software and hardware solution that monitors all traffic on the plant network and detects ongoing attacks.
Network monitor - offices ($50,000)	An advanced software and hardware solution that monitors all traffic on the office network and detects ongoing attacks.
Firewall - plant ($30,000)	A software and hardware solution that monitors and filters unauthorized traffic coming from the Internet to the plant network.
Firewall - offices ($30,000)	A software and hardware solution that monitors and filters unauthorized traffic coming from the Internet to the office network.
Anti-virus ($30,000)	A software protection against malware for all PCs (plant and offices).
Security Training ($30,000)	Basic security hygiene for all employees (plant and offices).
Asset Audit ($30,000)	Detailed evaluation of the company's infrastructure, reveals and shuts down an open wifi network on the plant, and unlocks additional defenses (cf. Table 2).
Threat Assessment ($20,000)	Detailed information about possible threats and attacks against the company.

TABLE 2
Additional defenses available after an Asset Audit.

Patches - Controller ($30,000)	Upgrade to the firmware of the SCADA controller.
Patches - PCs ($30,000)	Upgrade to the operating system of all PCs (plant and offices).
Patches - Server & DBs ($30,000)	Upgrade to the operating system of the server and databases (plant and offices).
Encryption - PCs ($20,000)	Encryption for all PCs (plant and offices).
Encryption - databases ($20,000)	Encryption for all databases (plant and offices).

- *Organized crime* using more advanced techniques (spear phishing, infected USB drives, infiltration via an insecure wifi network) to achieve more advanced goals (data exfiltration from the offices and plant, ransom based on controller disruption).
- *Nation states* using the most advanced attacks to exfiltrate technical data from the plant and disrupt the controller.

If the players invest in a Threat Assessment, the Game Master tells them about these three types of attackers and the type of attacks and goals associated with them. Script kiddies are "100% likely" to hit the company, Organized Crime attacks are "quite likely" whereas nation states attacks are "unlikely and nearly impossible to defend against anyway". The players are, therefore, encouraged not to focus on high-profile attacks and to make sure that the organization and its infrastructure is properly secure against the most likely threats.

Each attacker follows a particular attack progression, depicted in Table 3. It is important to note that the players do not have access to this table: a Threat Assessment gives them high-level information about the attackers, their methods and the associated likelihood, but no specific timings or progressions. Most attacks are initially silent unless the players have invested in the proper type of defense: for instance, Network Scans go undetected unless Firewalls are deployed, Phishing attacks are silently successful unless Security Training has been purchased for employees. Visible attack effects hit the infrastructure when it is too late: DoS attack paralyzing an un-firewalled server, viruses disrupting PCs, or a ransom for releasing stolen data.

The attacks are designed so that low-level attacks (DoS, simple virus by Script Kiddies at round 2 and 3) hit early to assess whether the players invested in security essentials against the most common threats. More sophisticated attacks follow an Advanced Persistent Threat (APT) life-cycle. These attacks hit later (data exfiltration, controller disruption in round 4) to assess whether players can prioritize between less frequent, sophisticated attacks and frequent, low-level threats (Script Kiddies in rounds 1, 2). The effect on the company's share price is also proportional to the sophistication of the attack: small bump when hit by a Script Kiddie, significant dip when hit by Organized Crime (along with mentions by the Game Master of press articles and headlines in the news). The Nation State attacks are revealed at the end of round 4 only, when the game finishes: the Game Master then mentions to the players that they were not expected to be able to defend against them.

The possibility of adding an element of randomness to the attack scenarios was considered for its realism, but discarded as it would have biased the comparison between sessions, since groups would not have faced the same attacks.

3.5 Validation of the game model

The models of the company's infrastructure and the attacks targeting it are central elements that determine the game's realism and fairness. The cyber physical infrastructure must

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSE.2017.2782813, IEEE Transactions on Software Engineering

TABLE 3
Attacks targeting the infrastructure and the corresponding counters (defenses) noted × in the table.

Attacker	Round 1	Round 2	Round 3	Round 4
Scanning Kiddie	Scan offices × Firewall offices	Scan offices × Firewall offices	Scan offices × Firewall offices	Scan offices × Firewall offices
DoSing Kiddie		DoS offices × Firewall offices	DoS offices × Firewall offices	DoS offices × Firewall offices
Hacking Kiddie		Remote control server × Server patch	Data exfiltration server × Net. mon. offices × Encryption DB	Data exfiltration server × Net. mon. offices × Encryption DB
Phishing Kiddie	Phishing offices (trojan) × Training × Antivirus × Patches PCs	Phishing offices (trojan) × Training × Antivirus × Patches PCs	Phishing offices (trojan) × Training × Antivirus × Patches PCs	Phishing offices (trojan) × Training × Antivirus × Patches PCs
Malware Kiddie		Disruption PC offices × Training × Antivirus × Patches PCs	Disruption PC offices × Training × Antivirus × Patches PCs	Disruption PC offices × Training × Antivirus × Patches PCs
APT PC Offices	Infected USB offices × Training × Antivirus	Remote Control PC × Antivirus × Net. mon. offices	Data exfiltration PC × Antivirus × Encryption PCs × Net. mon. offices	Data exfiltration PC × Antivirus × Encryption PCs × Net. mon. offices
APT Server Offices	Phishing office credentials × Training	Remote Control Server × Net. mon. offices	Data exfiltration DB × Net. mon. offices × Encryption DB	Data exfiltration DB × Net. mon. offices × Encryption DB
APT DB Plant	Vulnerable Wi-Fi plant × Asset Audit	Remote Control DB plant × Patch server × Net. mon. plant	Data exfiltration DB plant × Net. mon. plant × Encryption DB	Data exfiltration DB plant × Net. mon. plant × Encryption DB
APT Controller	Scan plant × Firewall plant	Remote control Controller × Patch controller × Firewall plant	Disruption controller × Patch controller	Disruption controller × Patch controller
State Intelligence	Physical intrusion plant × CCTV plant	0day DB plant × Net. mon. plant	Data exfiltration DB plant × Net. mon. plant	Data exfiltration DB plant × Net. mon. plant
State Disruption	Physical intrusion plant × CCTV plant	Remote control controller (0day)		Disruption controller

include the essential elements of comparable real-life systems, despite the objective of making a simple game that non-experts can play. The attack scenarios were designed to be varied and representative of the current threat model for industrial control systems [16], [22]. Table 3 includes attacks of different natures – social engineering, cyber attacks, physical attacks – and different degrees of sophistication. This design choice favors players who are able to balance priorities between these different vectors over players focusing on a single type of threat: players are not rewarded for guessing the one particular attack they should be concerned with, but for identifying and countering as many different attacks as possible. From a study perspective, this also allows us to capture a wide variety of strategies that we can differentiate.

The infrastructure model was elaborated based on our experience with industrial control systems. It was validated by all the computer scientists and security experts who played the game. The distribution of attacks took inspiration from recent threat reports (e.g. [16], [22]) and was also validated by the security experts who played the game.

3.6 Game design discussion

The balance between *theme* and *mechanics* was a key design choice for D-D. The game was carefully designed in order not to encourage a mechanics-based play (or, in role-playing parlance, "meta-gaming"). The mechanics of the game are kept to a strict minimum from a player perspective: 4 rounds, $100,000 budget per round, 15 defense cards and the infrastructure are all the mechanics that they see. In particular, the players do not have access to the attack table, with which they would be able to understand the mechanics of the game and optimize their strategy accordingly. Instead, players must base their decisions on the thematic role of each defense and how they fit into the threat environment they are facing, which is entirely narrated.

This design choice is also enforced by the Game Master: any meta-gaming attempt is discouraged. The Game Master de-emphasizes game mechanics and asks the players to focus instead on "what they would do in the real world". The rulebook provides guidelines for Game Masters to encourage immersion and respond to players tempted by meta-gaming. For instance, a typical meta-gaming behavior would comprise second-guessing the Game Master (for instance: "I

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSE.2017.2782813, IEEE Transactions on Software Engineering

bet the GM will run exactly the attack that we won't have defended against..."). As an answer, the Game Master can emphasize that the attack scenarios are pre-determined and mimick real-world conditions, reconstructed from actual threat reports. In section 8.1 we evaluate whether players' decisions are indeed based on immersed, theme-driven thinking and to what extent do they rely on mechanics-based optimization.

4 METHODOLOGY

We played D-D with a total of 43 players divided into 12 homogeneous groups:

- 4 groups of "security experts" with a background (i.e. skills, degree and/or professional experience) in cyber-security.
- 4 groups of "computer scientists" with a background in computer science but not in cyber-security.
- 4 groups of "managers" with a management background and no skills in computer science or cyber-security.

We ensured that our participants fell into these clear categories to avoid biasing the analysis. Therefore, when referring to players, "managers" or "computer scientists" refer to the backgrounds described above instead of an actual function. "Managers" were chosen as they play a major role in decisions about security and budgets, and their background is identifiable.

The groups featured subjects from either academia or industry, the former being either academics, PhD students, postgraduate or undergraduate students in the corresponding areas. We cannot reveal affiliations for ethical reasons, but all "industry" players held a position or had previous work experience in industry. Each group is given a unique identifier shown in Table 5 (e.g., "CI2" for the second group of computer scientists from industry). For consistency, all games were run by the same Game Master with expertise in both security and games.

We advertised game sessions on our universities' mailing lists, asking for volunteers, and we reached out to a number of our industry partners. Group selection was organised on a first-come first-served basis. A £10 compensation was offered to students. An ethics agreement was signed, guaranteeing the confidentiality of the study and of all personal statements recorded during the session. The recordings are kept in a secure, encrypted location. The transcripts are anonymized and kept confidential. The project received approval from the relevant ethics committee.

This experimental sample is substantial for a qualitative analysis, much larger than existing literature on security games (e.g., [9] is based on a survey of 14 educators and observation of 11 players). We do not claim statistical significance: the quantitative values reported are to ground our qualitative insights and observations, in line with the general methodological approach taken in literature, e.g., [9], [23].

At the end of each round, the players were invited to write a short report detailing their investments and a short justification behind these. In addition to these logs of player decisions, we video-recorded all sessions with informed consent and transcribed them, then open-coded the transcripts [11]. We analyze these two data sources using several measurements as follows.

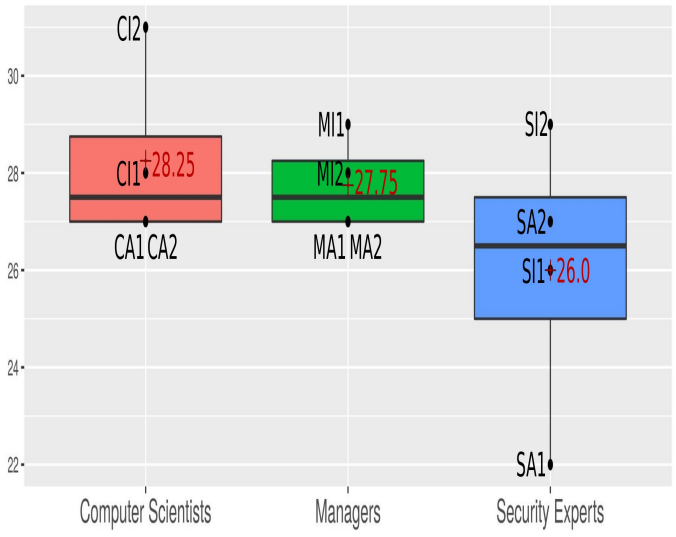

Fig. 3. Game scores (red cross = average).

4.1 Game score

To measure how well the players defended their infrastructure, we marked each game with a *Game Score* that counts how many successful attacks the players successfully defended. In itself, the game score is not an absolute measurement of the security skills of the players: it is one indicator that must be considered in the context of the other qualitative observations that this study presents.

We considered a total of 33 attacks from Script Kiddies and Organized Crime (cf. Table 3), each successful defense granting one point. State Attacks do not count as players should not be trying to defend against such high-profile threats: for instance, in round 1, defending a physical infiltration by a foreign spy with CCTV should not be considered a good play, as it comes to the detriment of essential defenses against more likely low-level threats. Figure 3 shows the game score per demographics. Computer scientists achieved the best results as a demographic, manager teams had consistently average results, whereas some teams of security experts ended up at the bottom of the score sheet – surprising results that we discuss in more detail in Sections 6 and 7.

4.2 Measuring player strategies

During the game, a record is kept of all players' investments, round per round: this record is used to measure the players' interest in different types of defenses. We assume that the earlier players invest in a defense, the more important this defense is to them. This assumption is clearly explained to the players by the Game Master: all defenses are useful and their budget is limited, therefore they are invited to prioritize the defenses they deem the most important for each round. We partition defenses into four categories and associated measurements that capture the interest of players:

1) **Data protection**: How much importance is given to protecting the company's data (e.g., in databases, PC hard drives) from being stolen?

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSE.2017.2782813, IEEE Transactions on Software Engineering

TABLE 4
Detailed investments by all teams during the game.

Team	Round 1	Round 2	Round 3	Round 4
SA1	Asset_audit Patches_controller Encryption_DBs Encryption_PCs	Firewall_plant Monitoring_office	Firewall_office Training CCTV_plant	Patches_PCs Patches_servers Antivirus Monitoring_plant
SA2	Firewall_plant Firewall_office Threat_assessment	Antivirus Monitoring_plant Encryption_DBs	Training Monitoring_office Asset_audit	Patches_servers CCTV_plant Patches_PCs
SI1	Threat_assessment Asset_audit Firewall_plant	Firewall_office Patches_PCs Patches_servers Training	Monitoring_office Monitoring_plant	Antivirus Encryption_DBs Encryption_PCs Patches_controller
SI2	Firewall_plant Asset_audit Firewall_office	Patches_servers Antivirus Encryption_DBs Training	Monitoring_plant CCTV_office	Monitoring_office CCTV_plant
CA1	Threat_assessment Firewall_office Firewall_plant	Asset_audit Antivirus Patches_PCs Training	Patches_servers Patches_controller Encryption_PCs Encryption_DBs	Monitoring_office Monitoring_plant
CA2	Threat_assessment Asset_audit Encryption_DBs Firewall_plant	Antivirus Training Firewall_office	Patches_PCs Patches_controller Monitoring_office	CCTV_plant CCTV_office
CI1	Threat_assessment Firewall_office Firewall_plant	Asset_audit Patches_PCs Patches_servers Antivirus	Training CCTV_plant Monitoring_office	Encryption_DBs Patches_controller Monitoring_plant
CI2	Threat_assessment Training Firewall_office	Asset_audit Patches_servers Firewall_plant Encryption_DBs	Patches_PCs Patches_controller Monitoring_office	CCTV_plant CCTV_office
MA1	Antivirus Firewall_office Firewall_plant	CCTV_plant Monitoring_office	Asset_audit Encryption_PCs Encryption_DBs Patches_PCs	Monitoring_plant Patches_servers Training
MA2	Threat_assessment Asset_audit Firewall_office Encryption_DBs	Firewall_plant Antivirus Encryption_PCs	Monitoring_office Patches_controller Training	Monitoring_plant CCTV_plant
MI1	Threat_assessment Antivirus Firewall_office	Asset_audit Firewall_plant Patches_PCs Patches_servers	Encryption_DBs Patches_controller Monitoring_office	CCTV_office Monitoring_plant
MI2	Threat_assessment Antivirus Firewall_office	Asset_audit Encryption_DBs Patches_PCs Patches_controller	Firewall_plant Training Encryption_PCs Patches_servers	Monitoring_office Monitoring_plant

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSE.2017.2782813, IEEE Transactions on Software Engineering

TABLE 5
Group names and player distribution.

	Academia	**Industry**
Security experts	SA1 (4 PhD students) SA2 (3 undergr. stud.)	SI1 (4 consultants) SI2 (5 consultants)
Computer scientists	CA1 (2 academics) CA2 (4 postgrad. stud.)	CI1 (6 IT engineers) CI2 (4 IT engineers)
Managers	MA1 (3 postgrad. stud.) MA2 (4 undergr. stud.)	MI1 (2 managers) MI2 (2 managers)

TABLE 6
Definition of Interest Scores in defenses.

Round the defense is played	1st	2nd	3rd	4th	never
Interest score for the defense	4	3	2	1	0

2) **Intelligence gathering**: How much importance is given to evaluating the situation (threats, assets) before investing in actual defenses?
3) **Human factors**: How much importance is given to addressing human vulnerabilities (bad security practices, social engineering)?
4) **Technological solutions**: How much the players invest in technological solutions, as opposed to the first three categories? This category is further refined into three subcategories: **physical security** (i.e. CCTV against physical intrusions), **basic cyber security** (essentials such as firewalls, antivirus, security patches) and **advanced cyber security** (highly sophisticated network monitoring and intrusion detection).

To quantify the interest of players in defenses, we associate each defense with an *Interest Score (IS)* defined in Table 6. For instance, if a team invests in an Antivirus in the first round and a Security Training in round 3, the corresponding Interest Scores for this team are:

$$IS(Antivirus) = 4$$
$$IS(Security\ Training) = 2$$

The interest of players in the four categories of defenses is then measured via the following scores:

- Data Protection Score (DPS):

$$DPS = IS(Encryption\ PCs) + IS(Encryption\ DBs)$$

- Intelligence Gathering Score (IGS):

$$IGS = IS(Asset\ Audit) + IS(Threat\ Assessment)$$

- Human Factors Score (HFS):

$$HFS = IS(Security\ Training)$$

- Physical Security Score (PSS), Basic Cyber Security Score (BCS) and Advanced Cyber Security Score (ACS):

$$PSS = IS(CCTV\ Plant) + IS(CCTV\ Office)$$

$$BCS = IS(Firewall\ Plant) + IS(Firewall\ Office) + IS(Antivirus) + IS(Patches\ PCs) + IS(Patches\ Server\ \&\ DBs) + IS(Patches\ Controller)$$

$$ACS = IS(Network\ Monitoring\ Plant) + IS(Network\ Monitoring\ Office)$$

Figure 4 presents the measurements for each of these scores.

4.3 Characterizing decision processes

After measuring player strategies, we analyzed the decision processes themselves via two indicators: the type and richness of the arguments players used and the players' confidence in their own decisions. These indicators were derived while open-coding the transcripts [11]. More precisely, each argument used by a participant is associated with one of the following categories, presented by decreasing levels of maturity:

- **Procedure**: a participant explicitly applies a methodological procedure. Example: "*We should start with an asset audit, then we can know what we are protecting and invest accordingly.*"
- **Experience**: a participant bases a decision on relevant past experience with similar situations. Examples: "*I have never seen an IT infrastructure without a firewall.*", "*Remember the news last week? They got owned by a phishing email, we should care about it.*"
- **Scenario**: a participant invents a hypothetical scenario, describing an attack or a potential situation, to illustrate a point to other players. Example: "*What if someone got access to our database? We need to encrypt it.*"
- **Intuition**: a participant provides no additional evidence but their gut instinct. In the absence of one of the former justifications, this is the default code associated with arguments. Example: "*I like the antivirus.*"

Figure 5 shows, for each team, how many arguments were used in each of these categories.

Measuring self-confidence follows a similar protocol: each mention by participants of their confidence in their decisions – either positive or negative – is coded. Examples: "*(To the Game Master:) you told us what we already knew.*" (positive); "*I don't know, I'm not sure.*" (negative). We then computed the total number of positive and negative self-evaluations. The results are shown in Figure 6.

5 Analysis of Player Strategies

In this section we discuss player strategies in terms of priorities in their investments (Figure 4) and their efficiency

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSE.2017.2782813, IEEE Transactions on Software Engineering

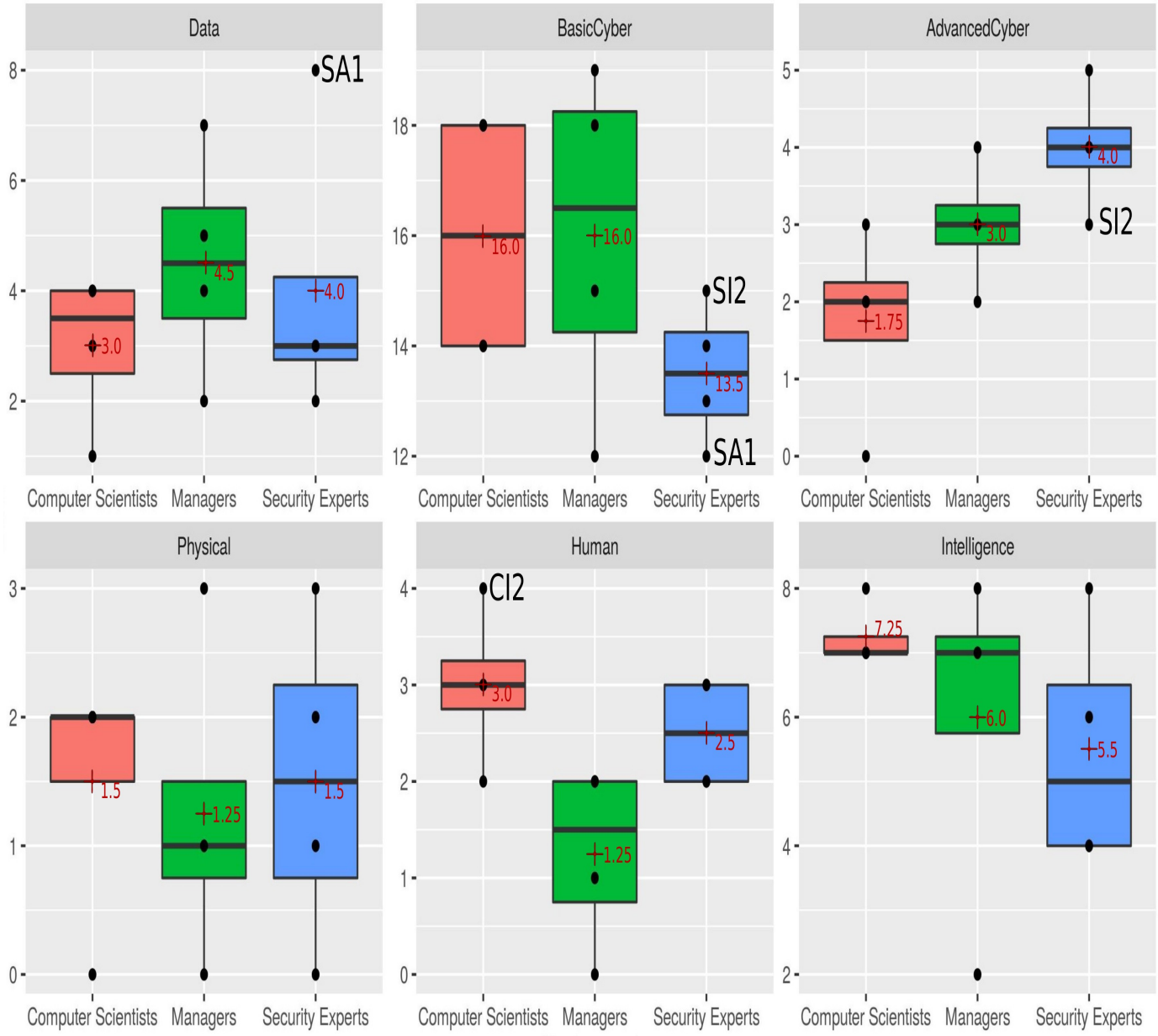

Fig. 4. Detailed scores for each demographics (for clarity, only teams mentioned in the text are labeled).

with respect to their game score (Figure 3). In terms of background, we can summarize player strategies with the following tendencies:

- Security experts were strongly attracted by Advanced Cyber Protection and neglected Basic Cyber Protection and Intelligence Gathering.
- Computer scientists favored Intelligence Gathering and Human Factors while being less interested in Advanced Cyber Protection and Data Protection.
- Managers were technology-driven (Basic and Advanced) and focused more on Data Protection than other demographics while neglecting Human Factors.

5.1 The best players are not the ones you think

Strikingly, security experts do not get better scores than the other two categories while providing the two worst performances of the panel (22 points for SA1 and 26 points for SI1 in Figure 3). Security experts were the most interested in advanced cyber security solutions to the detriment of basic protections (average ACS = 4 and average BCS = 13.5 in Figure 4). The discussions steered rapidly towards deploying "big shiny boxes" (i.e. network monitoring) in all groups of security experts. Interestingly, the most successful team of security experts (SI2, with a Game Score of 29 in Figure 3) did not follow this tendency as much as other teams: they had the lowest interest in Advanced Cyber Protection and the highest interest in Basic Cyber Protection among all teams of security experts (ACS = 3 and BCS = 15 in Figure 4).

Security experts also tended to neglect intelligence gathering and in particular to skip threat assessments (average IGS = 5.5 in Figure 4). A player from team SA1, who achieved the lowest score of the panel, stated: "*We are security experts, we don't need a threat assessment.*". Groups such as SI1 who

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSE.2017.2782813, IEEE Transactions on Software Engineering

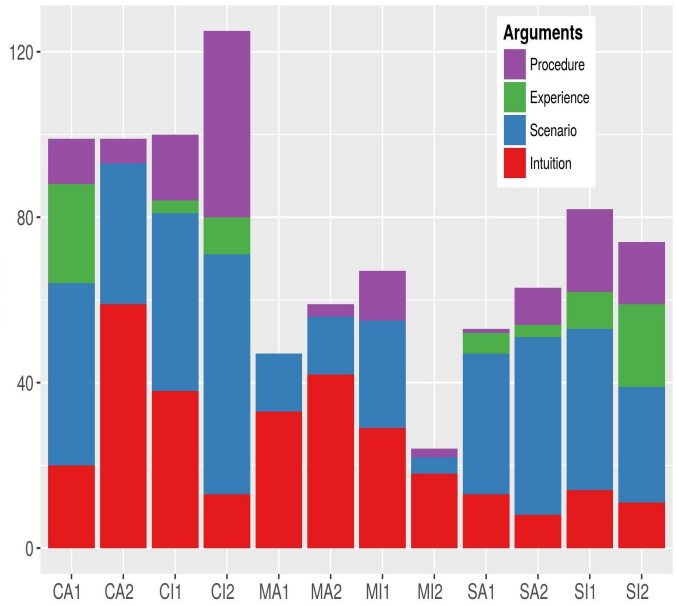

Fig. 5. Count of arguments in the transcripts.

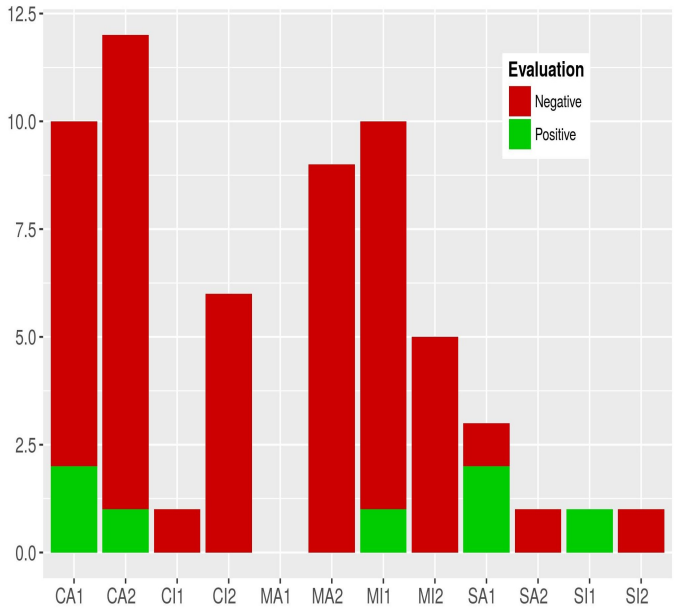

Fig. 6. Count of self-evaluation markers (no value means 0 markers were available in the transcripts).

did invest in a threat assessment noted that they had learnt little from it ("*You told us what we already knew.*"). However, the detailed analysis of their decision processes (Section 7.3) shows that they were not able to capitalize on the very threat assessment they thought was obvious.

5.2 The technology-driven

Managers were technology-driven: they were the most concerned with data protection (average DPS = 4.5 in Figure 4) while having strong interest in cyber protection, both basic and advanced (average BCS = 8 and average ACS = 3.75 respectively in Figure 4). This tendency is confirmed by in-game reflections such as the following: "*I really need to have some software that can help me choose. I need some* technical *support.*". This technology focus comes at a price: managers were the least concerned about Human Factors (average HFS = 1.25 in Figure 4)

It should be noted that, despite this surprising trust in technology over humanity, none of the teams of managers provided a bad performance. Their results were actually quite regular, all teams scoring between 27 and 29 points (Figure 3).

5.3 Balance is a key to success

Computer scientists were overall the most interested in non-technological defenses: they had the highest scores in Human Factors (average HFS = 7.5 in Figure 4) and Intelligence Gathering (average IGS = 9 in Figure 4). As summed up by a player from CI2: "*You need to see what the problems are before you try and fix them.*". Computer scientists also showed a bias against Advanced Cyber Solutions in favor of Basic Cyber Solutions (average ACS = 1.75 and average BCS = 16 respectively in Figure 4), opposite to the strategy of security experts. Finally, they were the least interested in Data Protection (average DPS = 3.75 in Figure 4).

Such balanced, and not solely technology-driven, strategies yielded good results: the best score of the 12 teams was achieved by computer scientists (31 points for CI2 in Figure 3) while the other 3 teams had average scores (27 or 28). Overall, of the three demographics, the group of computer scientists were the ones with the best results.

6 Decision Processes

Considering the way teams took their decisions, in terms of arguments and self-evaluation, we identified two broad behaviors, depending on whether teams had both experience and technical knowledge or not. Interestingly, this classification does not necessarily correlate with good results, as shown in Section 5.

6.1 The intuition-driven n00bs

Team CA2 (computer science students) and the four teams of managers (MA1 and MA2: management students, MI1 and MI2: industry managers) lacked either experience or a technical background, and sometimes both. As a result, their arguments were poor and mostly based on abstract intuitions: "*I think we should go for firewalls*", "*I like the antivirus*". This was particularly clear for MA1 and MA2 (management students) and MI2 (junior managers): they barely used any other form of argumentation, apart from a few attack scenarios. Team CA2 (computer science students) and MI1 (senior industry managers) used a higher proportion of scenarios and even some procedural thinking, such as "*we should go first for an audit, then we will know what we are protecting*". Notably, these two teams were also highly self-critical, team MI1 scoring a record number of "*I'm not sure*" and "*I don't know, what do you think?*" in the transcripts.

6.2 The l33t (or are they?)

Contrary to the previous category, security experts and experienced computer scientists mainly used rich, concrete

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSE.2017.2782813, IEEE Transactions on Software Engineering

scenarios to argue: "*Imagine if someone compromised this box...*". A clear difference can be seen between student groups (teams SA1 and SA2: security students) and teams with more experience (teams SI1 and SI2: senior security consultants, CA1: senior academics, CI1: junior IT engineers, CI2: senior IT engineers). The latter used their experience of the field much more by recalling past anecdotes. They also used procedural thinking much more often, referring explicitly to initial intelligence gathering before investing into any defenses. Procedural thinking was not a guarantee of quality however: team SI1, for instance, constantly referred to their initial threat assessment but exclusively as a justification for a budget increase, neglecting the actual content of this assessment and the threats they were facing – this "tunnel vision" syndrome is discussed in Section 7.3.

Security experts showed a high degree of confidence, maybe due to a feeling of familiarity with security issues: their self-evaluation counters are very low and feature more positive mentions than other demographics. When realizing that they had been hacked, the reaction of security experts was in general to blame the lack of budget or complain that they had been put in a very unfavorable situation. Computer scientists and managers, on the other hand, acknowledged their lack of expertise in security much more: teams CA2 and MI1, for instance, were composed of highly self-critical players. At the end of the game, they were much more likely to acknowledge their mistakes; team CI2 was actually surprised by their excellent result, as they were constantly expecting a disaster to happen until the very end.

7 The Good, the Bad, and the Ugly

The previous sections presented a number of patterns that we observed in players' decision processes: in terms of their defensive focus, argument types and self-evaluations, and the overall game scores of the different teams. This section provides more detailed analysis of the transcripts, in order to identify both positive decision patterns that yield better results and mistakes the players made. We analyze several transversal patterns (good, bad or ugly) that we illustrate with characteristic plays. We invite the readers to be inspired by the virtuous behaviors described in this catalog and to discourage bad habits (don't try this at home!).

7.1 Balance is key (good)

In D-D, finding a good balance between investments is critical to answer all threats appropriately. An inspiring example is team CI2, a group of experienced computer scientists from industry, who played a near-perfect game. The only way the team could have reached a higher game score would have been to skip the Threat Assessment (assuming the participants already had the appropriate knowledge) and to invest in an earlier Firewall, which would have deflected an inconsequential scan on the plant during the first round. A few characteristic features of this team:

- They were the only team to invest in a security training in round 1, deflecting 3 attacks at once (maximum Human Factors Score (HFS) of 4, cf. Figure 4). As one of the players said: "*You can have all the technology in the world, if people are still going to click on a dodgy link in an email...*".
- They correctly identified that the offices were more exposed than the plant, due to the public server, and prioritized their investments on this side (early office firewall and server patch in rounds 1 and 2).
- They delayed less critical investments (controller patch, CCTV) to later rounds while focusing explicitly on balancing their different defenses – technical and non-technical, plant side and office side – according to their evaluation of threats.
- The team had remarkably balanced discussions, every player expressing diverging opinions. Despite their experience, they were still self-critical (cf. self-evaluation markers on Figure 6) and were genuinely impressed by their good performance at the end of the game.

7.2 A little knowledge is a dangerous thing (bad)

Contrary to balanced approaches, excessive focus on particular threats leads to bad results. Team SA1 (security students) provided the worst performance of all teams, in terms of game score. They suffered from two major weaknesses:

- The team had an enthusiastic attraction for high-level threats: they invested straight into encryption for databases and PCs, by fear of data exfiltration, followed by a very early advanced technology – network monitoring – during round 2 (their Data Protection Score (DPS) is the highest of all team, cf. Figure 4). Meanwhile, their untrained personnel sitting in a non-firewalled office with an unpatched server were hit by multiple Script Kiddie attacks – trojan-infected email, denial of service, compromised server – at the end of round 2 (low Basic Cyber Security Score (BCS), cf. Figure 4).
- Despite their lack of experience, the team lacked self-criticism, one of the players notably stating that "*We are security experts, we don't need a threat assessment*" (their self-evaluation markers are predominantly positive, cf. Figure 6). Such a threat assessment would have precisely shifted their attention towards more likely low-level threats they finally cared about when it was too late. At the end of the game, realizing how poorly they had played, one of the players concluded: "*Ignorance was bliss.*".

7.3 The "tunnel vision" syndrome (ugly)

When a team has strong pre-conceived assumptions about security, these can drive their decisions regardless of any contradictory information or feedback collected during the game. Team SA1, described in the previous section, were clearly affected by tunnel vision: their focus on data protection and advanced cyber defenses left them vulnerable to many low-level attacks. Team SI1, a group of engineers from a cyber security consulting firm, suffered from a similar syndrome and neglected data protection altogether, to a bitter end:

- The senior engineer of the group stated, at the beginning of round 1, that "*this company's data has little value: you could publish it all*".
- After investing in a threat assessment that described a number of potential data exfiltration attacks, the players signaled to the Game Master: "*you told us what we already knew.*". Yet they kept giving a very low priority to data encryption.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSE.2017.2782813, IEEE Transactions on Software Engineering

- At the end of round 2, they were hit by a minor data exfiltration attack, and still they did not change their plans and delayed encryption investments in favor of more protection against disruptive attacks. The senior engineer explicitly said: "*I don't feel the encryption is any priority even though there has been a data breach.*".
- It took two major data exfiltrations during round 3, from both the plant and office databases, for the players to concede that encryption was important: they finally encrypted their databases during round 4, too late to stop ongoing attacks.

These two examples also show the risk of self-supported expertise: these two teams of experts had high confidence and little self-criticism. Therefore, they could not adapt to unforeseen attacks despite the relevant information being given to them. Interestingly, both of these teams expressed contempt for non-experts, using the exact same expression: "*Users are idiots.*".

7.4 Beware of the champion (for better or for worse)

Although teams take decisions collectively, individuals had a significant influence on the outcome of several games. Champions supporting their ideas with strong arguments were able to convince the rest of the team, for better or for worse. Conversely, some players failed to become effective champions for their cause and were silenced by stronger, yet wrong, arguments:

- Team CI2 (the perfect runner) was implicitly led by a senior engineer who pushed for an early security training, then directed the team's reflections according to his (correct) assessment of the risks for the infrastructure.
- One of the players in team SA1 (high-tech driven worst scorers) tried to argue in favor of investing in more basic defenses and considering human factors, yet his voice was not heard by his teammates.
- Team SI1 (second worst scorers) suffered from a tunnel vision syndrome partly because the senior engineer in the room disregarded the risks of data exfiltration.
- Team SI2 (second best scorer) played quite similarly to SI1, until one player with decades of experience in information assurance managed to convince the team to encrypt their databases, unknowingly preventing two catastrophic data exfiltration attacks – attacks that did hit SI1.

Here, it should be noted that the Game Master ensured a certain fairness during the debates by trying to balance speaking times among players. In real-world contexts, some less-vocal players would not have got the exposure they were given during the game and their influence on their team's decisions would have been even weaker.

7.5 The beginner's syndrome (good)

A lack of experience can be compensated by open-mindedness and adaptability to the inputs provided by the game. Opposite to some teams of security experts falling for their pre-conceptions, all teams of managers and computer scientists reached at least a relatively good score.

- Despite their lack of expertise, they were able to capitalize on information gathering: All non-expert teams, MA1 excepted, went for an Asset Audit and Threat Assessment during the first two rounds and interpreted it correctly.
- They did not suffer from excessive tropisms that could have put their defenses off-balance – such as an immoderate focus on data protection (SA1), a complete lack of consideration for data protection (SI1) or high interest in advanced cyber solutions (all security experts).
- They were constantly critical about themselves and their approach to the game (cf. Figure 6). Non-expert teams particularly praised the game for its educational value, which is discussed in more detail in Section 9.

8 Threats to validity

8.1 Influence of game mechanics

Players of a role-playing game such as D-D design their strategy based on two factors:

- *Mechanics*, i.e. investing in the defenses they think will optimally counter the game's attack scenario, in order to "win" the game.
- *Theme*, i.e. investing in the defenses they would invest in if they were facing the same situation in the real world.

D-D explicitly encourages theme-based play so that player decisions reflect their understanding of real-world security. In order to assess how theme-driven and how mechanics-driven the players were, we asked them to justify each of their investments with a short written sentence. We gathered 117 such justifications across the 12 games we played, and we classified it as follows:

- **109 theme-driven justifications** that unambiguously adopt an immersed, in-game perspective, for instance: "*We need to identify what we are protecting.*" to justify an Asset Audit, "*Data is the brain of the company and it shouldn't be vulnerable.*" to justify encrypting the databases.
- **5 mechanics-driven justifications** that unambiguously leverage a game mechanic, namely the round-based structure of the game (for instance: "*Can't afford all options on table - so do this first as gives benefit!! [sic]*") or the attack scenarios (for instance: "*[This defense is] more likely to stop insider threat and nation-state than CCTV.*").
- **3 ambiguous justifications** that all refer to ongoing attacks known by the players and for which we could not clearly determine whether the players were thinking in-game or meta-gaming, for instance: "*Something is going on in the office and we need to understand what.*".

Overall, 93% of the justifications relied on a theme-driven strategy, which confirms that D-D does indeed achieve its goal of immersing the players and recording their real-world perception. In the future, we plan on studying the influence of varying game mechanics, for instance, by changing the price of defenses, changing the attack scenarios, changing the infrastructure.

8.2 Influence of sample size

Although our sample size is significant for a qualitative analysis, it is not so for a quantitative analysis. We ran t-tests to assess whether there was a statistically significant difference between the distributions of the 6 Defence Scores

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSE.2017.2782813, IEEE Transactions on Software Engineering

TABLE 7
P-values from pairwise t-tests between Score distributions from different demographics.

	Data	BasicCyber	AdvancedCyber	Physical	Human	Intelligence	Game Score
Security experts vs. Computer scientists	0.545	0.121	0.029	1	0.359	0.164	0.254
Security experts vs. Managers	0.780	0.218	0.134	0.791	0.076	0.774	0.327
Computer scientists vs. Managers	0.284	1	0.155	0.767	0.033	0.427	0.660

and the Game Score for our three demographics. Results are shown in Table 7. There are no statistically significant differences, apart from two exceptions (out of 21 t-tests): security experts and computer scientists differ significantly on their AdvancedCyber score (p-value $0.029 < 0.05$), and computer scientists and managers differ significantly on their Human score (p-value $0.033 < 0.05$). These results hold after correcting for multiple testing, using Bonferroni correction. We do not claim statistical significance, and future work will focus on improving the statistical relevance of our results – namely by collecting and analysing a much larger set of games.

Another limitation of our approach is the lack of scalability of our qualitative analysis. Transcribing, coding and interpreting a single game transcript requires a significant amount of manual work from several qualified researchers. We are exploring potential ways of speeding up the analysis phase. We are currently considering automating some parts of it, for instance, via Natural Language Processing tools, while preserving the in-depth understanding of player decision processes it provides.

9 D-D BEYOND THE EXPERIMENT

Beyond its utility as a semi-controlled environment for experiments, D-D is intended to serve a number of purposes. All teams provided positive feedback after the game, although different backgrounds appreciated different aspects of D-D.

9.1 Educational training

Non-expert teams were extremely positive regarding the educational value of the game. Several management students reported to their teachers that they wished D-D had been part of their regular curriculum, as it provided them with an "informative and knowledgeable" introduction to cyber security, IT infrastructures and decision making. Before the game, players answered questions about their background and their familiarity with IT, security and industrial control systems (cf. Table 8). After the game they evaluated how much they had learnt during the session about these topics (cf. Table 9). As can be seen from the table, managers were extremely positive regarding the educational value of D-D. Making D-D a full-fledged tabletop game that can be used for educational purposes is the major objective for future work.

9.2 Corporate practice and communication

Industry participants in general were all interested in having D-D played in their organization with mixed audiences: for instance, having a CEO, a board director, a CISO and an IT engineer play the game together to discover their different cultures and build a common understanding of cyber security.

TABLE 8
Pre-session background assessment from managers (11 players in total).

Q: How would you rate your proficiency in computer science?	
No particular training	6
Some technical knowledge	3
Significant training or practice	2
Expert	0
Q: How would you rate your proficiency in cyber security?	
No particular training	8
Some technical knowledge	2
Significant training or practice	1
Expert	0
Q: How familiar are you with industrial control systems?	
Never heard the term before	5
Heard about it, but not sure what it is exactly	5
Familiar with it	1
Expert in the domain	0

TABLE 9
Post-session feedback from managers (11 players in total).

Q: What did you learn about computer science?	
Nothing	1
A few things	8
A lot	2
Q: What did you learn about cyber security?	
Nothing	0
A few things	6
A lot	5
Q: What did you learn about industrial control systems?	
Nothing	2
A few things	5
A lot	4

Participants with a governmental experience also praised the informative qualities of the game. Several teams made inquiries about future commercial versions of the game and expressed interest in the results of our study.

Industry participants also praised D-D as a game: "very enjoyable and well-constructed game" supported by a "nice design" and a "good visual design" that delivered "good fun" are some examples of the verbal and written feedback provided by players. The board and its elements in particular were explicitly appreciated by players from all backgrounds, as it provided a support for visualization and helped focusing the debates.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSE.2017.2782813, IEEE Transactions on Software Engineering

9.3 Extending D-D

In terms of the future objectives for D-D, a number of extensions to the game will be explored:

- A Game Master's guide for building new infrastructures and attack scenarios. Such an extension could increase the complexity of the game model so that companies can replicate their own infrastructure and threat environment for training purposes. This would also allow the game to be played several times by the same players in different settings (new infrastructure, new attack vectors, new attackers, new game objectives).
- A "red team vs. blue team" version where a separate team of attackers is also given a budget and objectives ("exfiltrate the HR data", "disrupt the SCADA controller"). The game then becomes an adversarial challenge between two teams that must handle partial information and anticipate their opponents' next move. This extension was particularly popular among security experts and computer scientists: several players informally asked to be registered for the (potential future) tests of this extension.
- A software version that would allow single players to play D-D individually, or several players to play without the need for a Game Master. Removing the Game Master would lead to a different experience altogether, as the Game Master has a central role in directing and interacting with the players, answering their questions, providing additional information and background, and ensuring fairness in their debates. A software version, on the other hand, sacrifices this central human dimension for the sake of portability and convenience: many players would be able to play this version of the game with minimal constraints and investments.

In order to catalyse the diffusion of the D-D and encourage the development of new versions of it, the rules of the game have been made public under a Creative Commons licence. They can be freely downloaded at:

http://decisions-disruptions.org

10 Conclusion

In today's complex organizations and connected infrastructures, little is understood about security decisions and the impact of various factors and biases behind them. It is essential to promote cultural bridges that allow different demographics to build a common understanding of the "how" and "why" of the issues at play. In this paper, we proposed Decisions & Disruptions, a game that allows participants from various backgrounds to experiment with and reflect on their approaches to security decisions. The analysis of 12 games reveals several key insights:

- **Strategic priorities and decision processes differed between demographics**: Security experts had a strong interest in advanced technological solutions. They tended to neglect intelligence gathering, due to strong self-confidence in their knowledge and expertise. However, this self-confidence was often not balanced by a willingness to reflect and critique their decisions. Managers were also technology-driven and focused on data protection while neglecting human factors. Their debates were mostly driven by intuition but did not lead to disastrous decisions. Computer scientists tended to balance human factors and intelligence gathering with technical solutions, and demonstrated a strong willingness to question their decisions. The above insights can be valuable when conducting requirements gathering or prioritization workshops with stakeholders from different backgrounds within an organization—the requirements engineer can be cognisant of these cultural biases and, when such patterns manifest, can take mitigating actions by exploring the rationale behind particular stakeholder decisions.
- **Participants with a technical background and/or experience were not necessarily better players**: Some teams of confident but over-focused experts achieved mediocre results, compared to inexperienced or non-technical players who better adapted to the various threats they were facing. Expertise was therefore not necessarily successful unless it had the willingness to question its pre-conceptions. This demonstrates the importance of incorporating so-called *lay* perspectives during security requirements engineering—non-experts possess invaluable business and operational knowledge that can contextualize the security risks and decisions pertaining to their mitigation.
- **Various characteristic patterns and their influence on security decisions manifested across the 12 games**: "balance is key" (good), "little knowledge" (bad), "tunnel vision" (ugly), "beware of the champion" (ambivalent), "beginner's syndrome" (good). Such patterns identify both good practices and typical errors and pitfalls to be avoided. D-D can, therefore, serve as a means for stakeholders to explore their perceptions and understandings of security risks, the trade-offs involved during decision-making and the impact of such decisions on the security of the overall system.

Beyond the scope of this particular experiment, we invite readers to consider their own security decisions in the light of our findings: Do any of these patterns sound familiar? On which side of the spectrum does the reader belong? Promoting good approaches and, above all, discouraging bad habits and ugly mistakes is of paramount importance in a world where cyber security is becoming a concern for everyone. Through games such as D-D, one can experiment with one's own attitude towards cyber security, reflect as to which patterns manifest in one's decisions, and hopefully end up on the good side of the spectrum.

Finally, D-D is a sandbox with clear educational value, as consistently noted by our participants. The game provides a didactic way of discovering cyber security for players with varying degrees of expertise. In corporate environments, D-D has the potential to become a strong communication and awareness-raising tool that allows players from different backgrounds – CEO, CISO, managers, IT engineers – to sit around a table, build a common understanding of security issues and bootstrap or consolidate their security requirements.

ACKNOWLEDGEMENTS.

This work is supported by UK Engineering and Physical Science Research Council Grant, Mumba: Multi-faceted metrics for ICS business risk analysis (EP/M002780/1), part of the UK Research Institute on Trustworthy Industrial Control Systems (RITICS). The authors also wish to thank Alexander whose Lego® play inspired the development of D-D.

REFERENCES

[1] A. Anderson. Effective management of information security and privacy. *Educause Quarterly*, 29(1):15, 2006.
[2] K. Beckers, S. Pape. A Serious Game for Eliciting Social Engineering Security Requirements. In *IEEE 24th International Requirements Engineering Conference, RE 2016, Beijing, China*.
[3] Auburn University. Lego lab http://ocm.auburn.edu/featured_story/lego_lab.html, 2012. Last accessed 13th May 2016.
[4] W. H. Baker and L. Wallace. Is information security under control?: Investigating quality in information security management. *Security & Privacy, IEEE*, 5(1):36–44, 2007.
[5] L. D. Bodin, L. A. Gordon, and M. P. Loeb. Information security and risk management. *Communications of the ACM*, 51(4):64–68, 2008.
[6] H. Cavusoglu, B. Mishra, and S. Raghunathan. A model for evaluating it security investments. *Communications of the ACM*, 47(7):87–92, 2004.
[7] L. Coles-Kemp and R. E. Overill. On the role of the facilitator in information security risk assessment. *Journal in Computer Virology*, 3(2):143–148, 2007.
[8] L. Corriss. Information security governance: Integrating security into the organizational culture. In *Proceedings of the 2010 Workshop on Governance of Technology, Information and Policies*, pages 35–41. ACM, 2010.
[9] T. Denning, A. Lerner, A. Shostack, and T. Kohno. Control-alt-hack: the design and evaluation of a card game for computer security awareness and education. In *2013 ACM SIGSAC Conference on Computer and Communications Security, CCS'13, Berlin, Germany, November 4-8, 2013*, pages 915–928, 2013.
[10] E. Donchin. Video games as research tools: The space fortress game. *Behavior Research Methods, Instrumentation & Computers*, 27(2):217–223, 1995.
[11] A. Strauss and J. M. Corbin. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory (2nd Edition)*. SAGE, 1998.
[12] L. A. Gordon and M. P. Loeb. The economics of information security investment. *ACM Transactions on Information and System Security (TISSEC)*, 5(4):438–457, 2002.
[13] C. E. Irvine, M. F. Thompson, and K. Allen. Cyberciege: gaming for information assurance. *IEEE Security & Privacy*, 3(3):61–64, May 2005.
[14] A. Jain and B. Boehm. Simvbse: Developing a game for value-based software engineering. In *19th Conference on Software Engineering Education & Training, April 19-21, 2006*, pages 103–114, 2006.
[15] D. Kirsh and P. Maglio. On distinguishing epistemic from pragmatic action. *Cognitive Science*, 18(4):513–549, 1994.
[16] McAfee. Threats Report http://www.mcafee.com/us/resources/reports/rp-quarterly-threats-aug-2015.pdf, 2015.
[17] National Bureau of Standards, Federal Information Processing Standards Publications (FIPS PUB) 65. Guideline for automatic data processing risk analysis, 1975.
[18] E. O. Navarro, A. Baker, and A. V. D. Hoek. Teaching software engineering using simulation games. In *International Conference on Simulation in Education, 2003*, 2003.
[19] E. O. Navarro and A. V. D. Hoek. Comprehensive evaluation of an educational software engineering simulation environment. In *20th Conference on Software Engineering Education & Training, July 3-5, 2007*, pages 195–202, 2007.
[20] M. Pinto-Albuquerque and A. Rashid. Tackling the requirements jigsaw puzzle. In *IEEE 22nd International Requirements Engineering Conference, RE 2014, Karlskrona, Sweden, August 25-29, 2014*, pages 233–242, 2014.
[21] R. K. Rainer Jr, C. A. Snyder, and H. H. Carr. Risk analysis for information technology. *Journal of Management Information Systems*, 8(1):129–147, 1991.
[22] SANS Institute. The State of Security in Control Systems Today https://www.sans.org/reading-room/whitepapers/analyst/state-security-control-systems-today-36042, 2015.
[23] R. Smith and O. Gotel. Gameplay to introduce and reinforce requirements engineering practices. In *16th IEEE International Requirements Engineering Conference, RE 2008, 8-12 September 2008, Barcelona, Catalunya, Spain*, pages 95–104, 2008.
[24] Sur-Seal. Lego plan model http://www.sur-seal.com/whoweare/journey/Lego.html, 2013. Last accessed 13th May 2016.
[25] G. Taran. Using games in software engineering education to teach risk management. In *20th Conference on Software Engineering Education & Training, July 3-5, 2007*, pages 211–220, 2007.
[26] R. Antrobus, S. Frey, B. Green, and A. Rashid. SimaticScan: Towards A Specialised Vulnerability Scanner for Industrial Control Systems. In *Proc. 4th International Symposium for ICS & SCADA Cyber Security Research 2016, ICS-CSR 2016*.
[27] W. Jardine, S. Frey, B. Breen, and A. Rashid. SENAMI: Selective Non-Invasive Active Monitoring for ICS Intrusion Detection. In *Proceedings of the 2nd ACM Workshop on Cyber-Physical Systems Security and Privacy, Vienna, Austria, October 28, 2016*, CPS-SPC'16, Pages 23–34.
[28] S. Frey, A. Rashid, A. Zanutto, J. S. Busby, and K. Szmagalska-Follis. On the role of latent design conditions in cyber-physical systems security. In *Proceedings of the 2nd International Workshop on Software Engineering for Smart Cyber-Physical Systems, Austin, TX, 2016*, SEsCPS'16, pages 43–46.
[29] A. Zanutto, B. Shreeve, K. Follis, J. S. Busby, A. Rashid. The Shadow Warriors: In the no man's land between industrial control systems and enterprise IT systems. In *Proceedings of the 3rd Workshop on Security Information Workers, Santa Clara, CA, 2017*, WSiW7.