Unknowns

CDT Threats & Risks: Session 3

Matthew Edwards

Oct. 10th, 2019

Section 1

Orientation

Structure: Fortnights

- Threat Modelling
- Unknowns
- Risk Management
- Driving Factors
- Scaling Analysis

Today's Goals

- Different categories of unknowns and uncertainties.
- Discussion of unknowns within security.
- Approaches to handling uncertainty in different areas.
- Hands-on exercise for anticipating intervention consequences.

Section 2

An Ontology of Ignorance

All Knowledge

Known Knowns	Known Unknowns
Unknown Knowns	Unknown Unknowns

The Known

Known Knowns are the things we know that we know. True statements of which we are aware, e.g., *"I am in this room."* or *"WordPress before 5.2.4 is vulnerable to stored XSS via the Customizer."* `CVE-2019-17674`. We can plan based on this knowledge.

Known Unknowns are things we are aware we don't or can't know. For example, the dates that flights will be cancelled, or whether there is a new XSS vulnerability in the latest version of Wordpress. We can plan *around* these possibilities.

What do we know we don’t know?

In pre-hoc risk management we have to deal with **static uncertainty** about our system.

- What is the likelihood that a vulnerability exists in a piece of software?
- What are the chances that an exploit can be developed successfully?
- What is the likelihood of a human being manipulated?

Dealing with static uncertainties

No good methods for understanding these values.

Typically heuristic information.

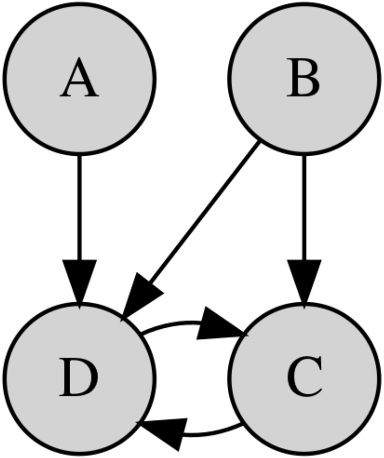
Security metrics.

Real-time and post-hoc uncertainty

This **dynamic uncertainty** mostly arises from the ‘invisibility’ of the attacker.

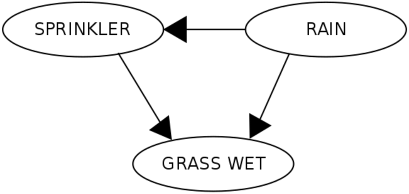
- Have we already been compromised?
- Are we currently under attack?
- What does the attacker want?
- How are they operating?
- What new threats are out there?

Dealing with dynamic uncertainties



Bayesian nets

	SPRINKLER	
RAIN	T	F
F	0.4	0.6
T	0.01	0.99



RAIN	
T	F
0.2	0.8

		GRASS WET	
SPRINKLER	RAIN	T	F
F	F	0.0	1.0
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

Unknown Knowns

Discussion

What does it mean to know something without knowing you know it? What examples can you think of?

Unknown Knowns

"[...] adjectives in English absolutely have to be in this order: opinion-size-age-shape-colour-origin-material-purpose Noun. So you can have a lovely little old rectangular green French silver whittling knife. But if you mess with that word order in the slightest you'll sound like a maniac. It's an odd thing that every English speaker uses that list, but almost none of us could write it out. As size comes before colour, green great dragons can't exist."

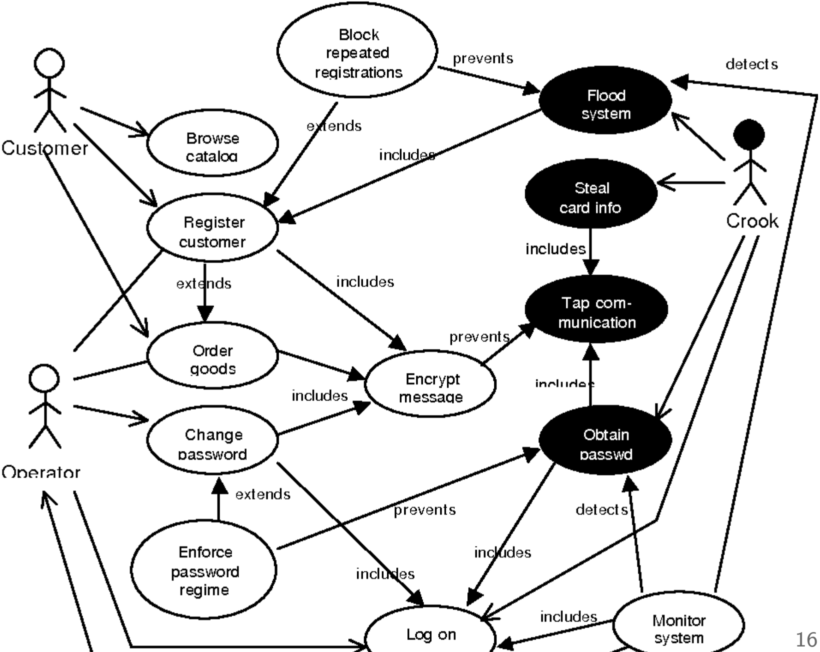
(Mark Forsythe, The Elements of Eloquence)

Unknown Knowns

In practice, handling this category involves translating *tacit* knowledge into *explicit* knowledge.

Problems like this arise in eliciting and understanding security **requirements** and **priorities**.

Security Requirements



Security Requirements

System Scope what is the system being protected?

Stakeholders who do you even need to be asking?

Understanding what are the ‘obvious’ security needs?

Incompatibility security needs in conflict with business/other security needs?

Volatility requirements will change over time.

Observation vs. Elicitation

"Rarely does an organization undertake a sophisticated or even semi-sophisticated financial analysis (i.e., cost-benefit or rate-of-return analysis) prior to making the investment or deciding on the level of investment that is needed. In fact, in many instances organizations simply react to a breach or compromise [...] and spend what it takes to solve the existing problem"[1]

[1] Rowe, Brent R., and Michael P. Gallaher. "Private sector cyber security investment strategies: An empirical analysis." In the 5th Workshop on the Economics of Information Security (WEIS06). 2006.

Observations cont.

“None of the organizations with whom we spoke felt that they had all the relevant expertise in-house to make effective cyber security investment decisions efficiently. Thus, external sources of security-related information are critically important.”[2]

[2] Rowe, Brent R., and Michael P. Gallaher. "Private sector cyber security investment strategies: An empirical analysis." In the 5th Workshop on the Economics of Information Security (WEIS06). 2006.

Observations cont.

Table: Drivers affecting organisations' cyber security investment strategy

Categories	Average percentage across organisations
Regulation-driven	30.1%
Network history/IT staff knowledge	18.9%
Client-driven	16.2%
Result of audit	12.4%
Response to current events	8.2%
Response to internal compromise	7.3%
Externally managed/determined	5.0%
Other	1.7%

Unknown Unknowns

Discussion

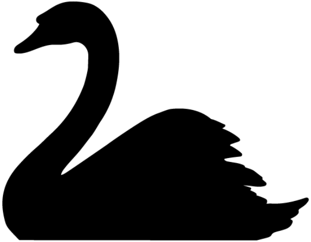
What sort of problems arise from not knowing what we don't know? What examples might we have?

Unknown Unknowns

Typically the most major problems.

New, previously unidentified classes of threat or vulnerability.

“Black swan” events.



“Black swans”

- Come as a surprise.
- Have a major effect.
- ‘Could have’ been predicted from relevant data.
- Rationalised by hindsight.

See also: *the problem of induction*.

Section 3

Unintended Consequences

Unintended Consequences

Typically, security research focuses on suggesting interventions.

Quite often, security interventions go wrong.

Sometimes they don't work.

Sometimes they make things *worse*.

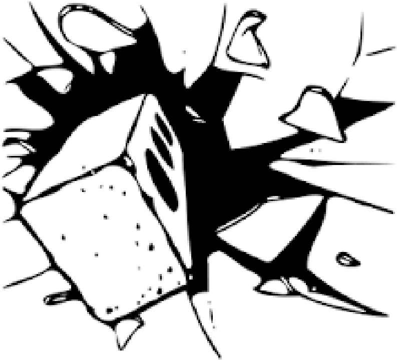
Category: Additional costs

- Financial costs (time & resources).
- Emotional & psychological burden.
- Security ‘compliance budget’

[3]

[3] Beautement, Adam, M. Angela Sasse, and Mike Wonham. ”The compliance budget: managing security behaviour in organisations.” In Proceedings of the 2008 New Security Paradigms Workshop, pp. 47-58. ACM, 2009.

Category: Misuse of Countermeasures



Countermeasures can be misused by mal-actors to cause harm.

- Reporting tools
- Victim advice
- Open classifiers

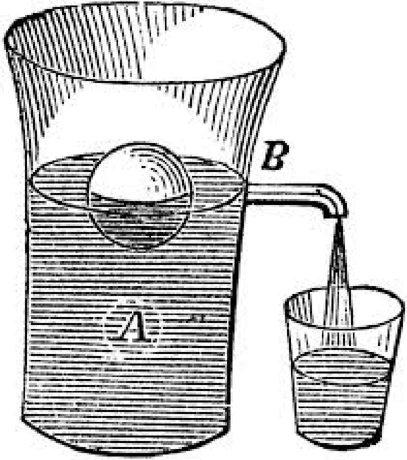
Category: False Positives

Most security decisions allocate people or behaviour into good/bad categories. Most classification systems have error. What happens to the mislabelled people?

Category: Displacement

You can 'move on' crime or misbehaviour.
This then becomes a problem for others, or could make it harder to tackle more effectively.

- Gab & Telegram
- Silk Road closure



Category: Amplification

Sometimes an intervention can cause a reaction that amplifies the thing it was attempting to prevent or discourage.

The most famous version of this is the **Streisand effect**.

Also occurs in e.g., domestic abuse.

Category: Insecure Norms

Many security countermeasures can become harmful through *over-reliance* and *dependence*, or foster other insecure behaviour (e.g., giving out identity details to websites that ask for it).

Category: Disrupting Other Countermeasures

- Removing harmful content from social media sites interferes with prosecution.
- Contradictory advice can confuse people into doing nothing.
- Forcing identification vs. managing your identity disclosure.

Categories

1. Additional costs
2. Misuse of countermeasure
3. False positives
4. Displacement
5. Amplification
6. Insecure norms
7. Disrupting other countermeasures

Exercise

Each come up with a different countermeasure/intervention for your situation, then swap and identify potential unintended harms of each countermeasure (try to get a few from each category).

Soo Yee & Manolis Phishing campaign

Robert & Priyanka Cyberbullying of/by teenagers

Hannah & Tobias (Tech-enabled) intimate partner abuse

Exercise: Part 2

Swap back, and now identify mitigations to as many unintended harms as possible while preserving your original countermeasure/ intervention.

Exercise: Part 3

Swap your lists of mitigations, and now identify potential harms for your partners' mitigations. Discuss what you're finding – how are mitigations interacting with the original goal?

Exercise: Part 4

For all the potential harms you've identified, check *for whom* these harms apply. Are particular categories of people more at risk of suffering from the intervention or mitigations?

Section 4

Next Week

Flipped Session

Within your assigned groups, decide which of you will be the **moderator**, the **prosecutor** and the **defence**. Your group will be given a paper, and you will present a ‘trial’ discussion, in which:

The Moderator will (in 5-10 minutes) first neutrally present the core of the paper’s argument as it was written.

The Prosecutor will (in 10 minutes) outline flaws with the paper’s argument, drawing on external sources as appropriate.

The Defence will (in 10 minutes) give a detailed rebuttal and highlight core strengths of the argument, drawing on sources as appropriate.

Verdict will be decided by discussion & vote of the audience.

Case 1: Knowledge & The State

Soo Yee, Manolis & Robert

Aradau, Claudia. “Assembling (non) knowledge: Security, law, and surveillance in a digital world.” *International Political Sociology* 11, no. 4 (2017): 327-342.

Case 2: Crime Reporting

Hannah, Priyanka & Tobias

Hoofnagle, Chris Jay. "Identity theft: Making the known unknowns known." *Harvard Journal of Law & Technology* 21 (2007): 97-122.

Request Session

Suggestions:

- Comparable single-paper review?
- Limitations of particular methods?
- Extended discussion session?
- ...