Flawed Mental Models Lead to Bad Cyber Security Decisions: *Let's Do a Better Job!*

Sean Smith
Department of Computer Science
Dartmouth College
sws@cs.dartmouth.edu

Ross Koppel
Department of Sociology
University of Pennsylvania
rkoppel@sas.upenn.edu

Jim Blythe
Information Sciences Institute
University of Southern California
blythe@isi.edu

Vijay Kothari
Department of Computer Science
Dartmouth College
vijayk@cs.dartmouth.edu

Visit shucs.org to learn more about the Science of Human Circumvention of Security

WHAT WE DID

Introduction

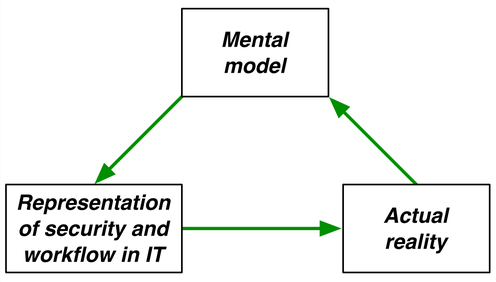
Users often work around security controls. We can pretend this doesn't happen, but it does.
In our research, we address this problem via observation and grounded theory (Bernard and Ryan, 2010; Charmaz, 2003; Pettigrew, 2000). Rather than assuming that users behave perfectly or that only bad users do bad things, we instead observe and record what really goes on compared to the various expectations. Then, after reviewing data, we develop structure and models, and bring in additional data to support, reject and refine these models.

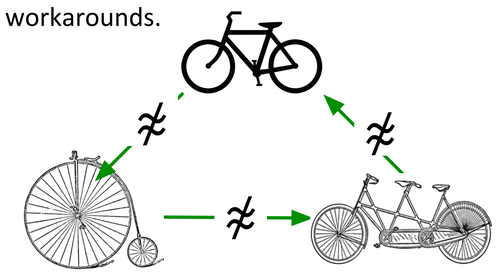


A Semiotic Model for IT Usability Trouble

In their seminal work on the meaning of language, Ogden and Richards (1927) constructed what is sometimes called the ***semiotic triad***. The vertices are the three principal objects: what the speaker (or listener/reader) ***thinks***; what ***symbol*** they use; and the actual item to which they are ***referring***.

We extend this semiotic model to examine reasons for workarounds.





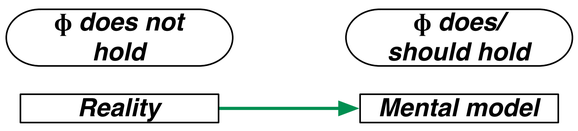
Circumvention semiotics: ***mismorphisms.***
Mappings ***fail to preserve*** structure

Loss of Properties Means Trouble

Provisioning:
- Unix sysadmins confidently creating wrong access controls.
- Users at universities, govt, and P2P accidentally make private files world readable (Maxion and Reeder, 2005).
- Investment bank employees unable to understand their own entitlements.
- Barrier to automated *role mining* is "interpretability" (Xu and Stoller, 2012)

Removing Functionality:
- smart key in Faraday foil (Paul and MacNaughton, 2014).

Adding Functionality:
- Sticky notes, shared passwords.
- US nuclear missiles had launch code "00000000"

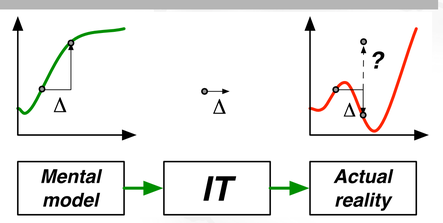


Shadow systems:
- Password-free telephone instead of online (Heckle,2011).
- Exfiltration by turning docs into images.
- Screen-scraping images into PowerPoint.
- Dropbox instead of official Sharepoint.
- Work docs sent to home email.
- Government users tunneling to university system.
- Government users working from Starbucks.

Turning Security Knobs has Unintended Consequences

Loss of Monotonicity

We implicitly have some numeric function *S* that maps a tunable parameter (e.g., password length) to the level of security achieved. The intention of the human is to tune the parameter *x* so as to maximize *S(x)*. However, if the mappings across the triad nodes fail to preserve crucial properties of this *x* vs *S(x)* curve, unfortunate things can happen.



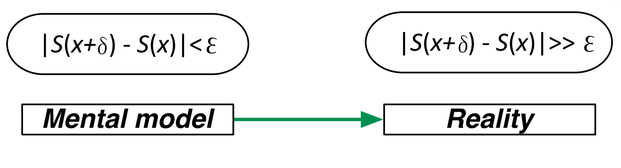
Uncanny Descent: Dialing security ***up*** can make the reality ***worse***.
- requiring strong passwords leads to writing them down or relying on security questions.
- adding S/MIME led to worse trust decisions (Masone, 2008).

Uncanny Ascent: Dialing security ***down*** can make the reality ***better***.
- eliminating unique passwords led to reduction in sharing.
- having browser remember critical site password stopped phishing.

Uncanny nop: Dialing security has almost ***no effect*** on the reality.
- passwords must be distinct from last *N*—but users knew they checked via hash.
- educating users about good behavior doesn't change behavior (e.g., Riley, 2006; Yan et al., 2005; Dhamija and Perrig, 2000; Heckle, 2011).

Loss of Continuity

Small changes in configuration can yield surprisingly big changes in security reality.

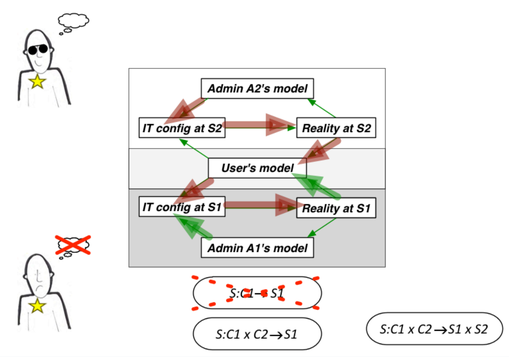




Domain and Range Trouble

Reality may have more parameters and consequences.

Example: loss of locality of control. The actual security at S1 can change because of a policy change by the admin at a different S2!
- password reuse + leak.
- training users to accept self-signed SSL certificates.
- training users to accept basic authentication.
- requiring users to change passwords.



WHAT COMES NEXT

A solution would likely have several components: effective ways to talk about aggregate security in practice, effective ways to discover and correct flaws in mental models, and effective ways to make better security decisions despite such flaws.

Measuring Aggregate Security

This requires multiple parts.
- We need to identify the scope. For example, if we change a password composition policy, we would need to know what effect the change will have on newly created passwords.
- But we may also need to consider the broader security implications (e.g., will users now be more likely to write passwords down on Post-It notes or to use the same password across many services?) or even things that may appear to extend beyond security, such as the impact on user workflow.
- We may need to also consider how our security decisions fundamentally change user behaviors, thereby having an impact on other organizations. For example, if one organization teaches employees to employ weak security practices, what is the impact on the security of other organizations? (E.g.: if Alice's employer said it was OK to accept self-signed certificates in her work application, then will she start doing that at her bank site?)
- To accurately quantify aggregate security we must also assign weights to our goals. Perhaps slightly more help-desk calls is an inconvenient, but necessary, cost that is offset by the gains of adopting a new security technology, yielding a net improvement. How do we go about quantifying this?
- Finally, given a measure of aggregate security, we will want to find ground truth values that accurately reflect the security profile. This would likely involve communicating with users by face-to-face communication and otherwise, and gathering auxiliary data, e.g., from logs, sensors, and help-desk calls.

Discovering Flawed Mental Models

Since mismorphisms are at the heart of numerous security problems, developing interpretable and meaningful representations of them are key to understanding security holes.
- A sustained and collective effort toward the development of a framework for identifying and classifying mismorphisms has the potential to dramatically increase our understanding of security problems, which in turn will ideally serve as a catalyst toward delivering scalable security solutions.
- The development of such a framework would require the collaboration of ethnographers, cognitive psychologists, and semioticians to gather ground truth data from real-world settings and build mental models from them, in conjunction with security practitioners to specify the desired goals of the models.

Tools for Making Better Decisions

We motivated this project with examples of failed security solutions because user behavior departed from the designer's model.
- Can we build frameworks to better evaluate security solutions before deployment?
- How do we incorporate these ``security'' assessments into a larger objective function that involves help desk calls, and fatigue that affects user performance on primary task, etc?

HoTSoS *Symposium and Bootcamp*
HOT TOPICS *in the* **SCIENCE OF SECURITY**
APRIL 4-5, 2017 | HANOVER, MARYLAND