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

Sean Smith, Vijay Kothari,* Jim Blythe,† Ross Koppel‡*

** Dartmouth College, † ISI, University of Southern California,, ‡ University of Pennsylvania*

Conventional computer security wisdom implicitly assumes models about humans and human organizations such as:

- Only bad people circumvent security controls. (Corollary: good users never share passwords.)
- It's actually possible for organizations to create and maintain a perfect electronic representation of the access control policies they need.

These models then translate into practices that conventional wisdom blesses as good. For just two examples:

- To make a system more secure, the security administrator should require stronger passwords and frequent password changes.
- To reduce inadvertent exposure of data from semi-public workstations, it's good to have user sessions automatically time out.

Unfortunately, fieldwork (by us[1] and many, many others) shows that these models are not necessarily true, and that the practices resulting from them do not necessarily make things better---and can in fact make things worse. E.g.:

- A colleague at a well-known IT giant tells how administrators enforced regular password changes, but only checked the hashes of the passwords. Consequently, users would merely append a number to easily-remembered password, and increment the number when forced to change---drastically reducing the benefits of requiring password changes.
- A colleague at a large hospital tells how administrators tried to reduce data exposure by adding proximity detectors to computer stations so clinician sessions would terminate, after a short timeout, if the clinician walked away. But that's not how medical exams are conducted, and the shut outs were constant. So, frustrated users covered the detectors with styrofoam cups, thus making the timeout effectively infinite, and increasing exposure.

That is, if we blindly apply conventional wisdom without validating the assumptions upon which it is based, we don't see the security gains that we might expect. This gives rise to *uncanny descents*, scenarios where we turn up security knobs with the expectation that aggregate security will improve, but we instead observe that things get worse.

We posit that these problems all result from the same underlying cause: flawed models of the interaction of humans and technology. Security policies, mechanisms, and recommendations are designed according to a human-conceived model of security, whether designed directly by the policy designer (e.g., by following tradition) or indirectly by the utilization of risk assessment or other security tools that are created by humans.

In previous work, we've characterized these causes as *mismorphisms*, ``mappings that fail to preserve structure''---especially mismatches between the security practitioner's mental model, the user's mental model, the model arising from system data, and the reality.

This situation gives rise to a grand challenge: how do we unravel this problem? Flawed models lead to bad decisions. We need a way to make better decisions.

[1] E.g., S.W. Smith, R. Koppel, J. Blythe, V. Kothari. "Mismorphism: a Semiotic Model of Computer Security Circumvention." *International Symposium on Human Aspects of Information Security and Assurance (HAISA 2015).* July 2015

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. We visit each in turn:

Measuring Aggregate Security. "Bad" and "good" implicitly require a metric. We need to construct effective, meaningful definitions of security reflecting the whole picture: multiple users, multiple sites, and what actually happens.

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. While some groups have tried to model underlying causes of security issues, 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.

Making Better Decisions. Constructing a framework to identify and classify mismorphisms is certainly a step in the right direction, but it is not enough. We need to develop practical tools and techniques that can be used to address problems induced by whole classes of mismorphisms. This would require a joint effort by researchers to predict the efficacy of security solutions, likely involving modeling experts, cognitive psychologists, and risk managers--- and ethnographers to continue to gather ground truth data and ensure deployed solutions meet their intended goals.

This paper opened 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?