

# How will humans in the loop do at detecting tangerines?

Matthew Merzbacher

/ May 11, 2016 /

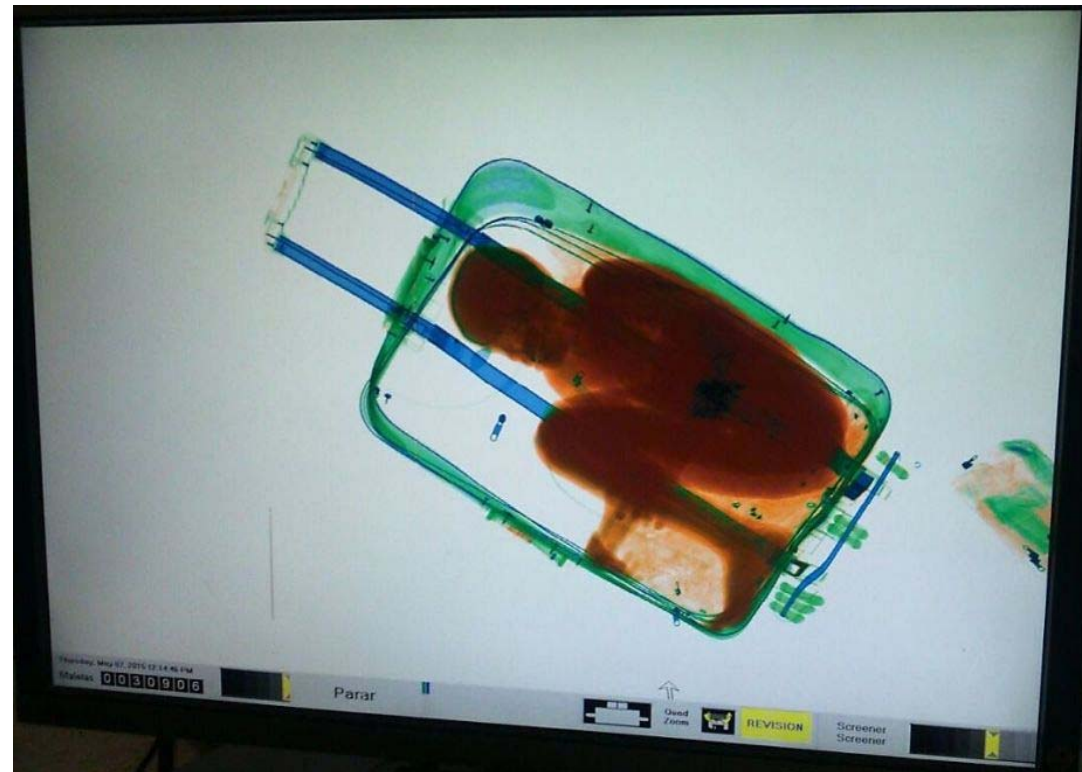
# GOAL

- **Humans (and other species) are an integral part of the security chain**
- **New threats (and presentations) can stymie humans**
- **Assess humans in the context of the overall system**
- **Without this, effectiveness is compromised**

GAO reported that... (AIT-ATR) — which display anomalies on a generic passenger outline — relied on both the technology's capability to identify potential threat items and its operators' ability to resolve them. However, GAO found that TSA did not include operators' ability in determining overall AIT-ATR system performance.

# DETECTION IN A MULTI-LEVEL WORLD

- What good is it to alarm upstream if the downstream system throws the alarm away?
- Do changes upstream cause new types of alarms that are not well-suited to downstream technologies?
- What about new alarms?
  - Tangerines!



# EXAMPLES

## → Adding a new Tangerine detecting device

- False alarms (e.g., grapefruit) hard to resolve
- Real alarms hard to identify
- Additional alarms flood overall system, causing degradation
- Device usage challenges

## → Adding Tangerine detection to existing device

- Unresolved alarms cascade downstream
- Shiny-Object Phenomenon: A change in distribution of alarm types can reduce vigilance and effectiveness towards other alarms



# DETECTION MODELING IS REALLY HARD

- **False alarm modeling can be boiled...**
- **Detection modeling is not so simple**
  - “by threat group” & “overall” detection
  - Cost of false negative = ??
  - Performance depends on correlation between levels
- **Each screening level is different**
  - Consider the whole end-to-end system
  - Human-in-the-loop challenging to model/measure
  - Specifying and measuring end-to-end performance is nearly impossible
    - And every time we change a piece, start over

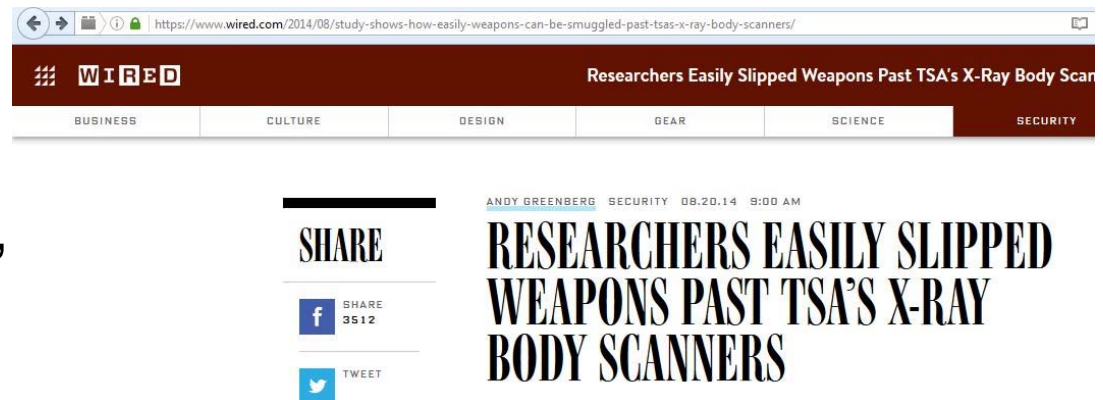
# OTHER HUMAN-IN-LOOP PROBLEMS

## → Presentation affects human performance

- How systems are deployed, installed, maintained and used is key
- May not be measured (or measurable) in lab

## → Limitations of flesh

- Training
- Ability
- Exhaustion
- Mood
- Green button syndrome
- Anecdotal behavior



# WHAT DO WE DO?

→ **How can we be fast and flexible in responding to new threats and still know what we're doing?**

→ **Test, observe, and adapt**

- Beware of capability gaps
- Operator protocols must change, but changing protocols leads to new gaps in expertise, training, and memory
- Fleshware upgrades are a challenge

*“Everything should be made as simple as possible, but not simpler”*



Roger Sessions

# PARTING QUESTIONS

- How do we specify and measure performance in a multi-level system, especially with a human-in-the-loop?
- Does the operator even need to know what a tangerine is to resolve tangerine alarms?
- How do we fuse levels of technologies (and species) and avoid the multi-modal irony?
- What good is it to detect at one level if we can't handle the alarm at the next?
  - Other than deterrent effect

*Render unto Humans, that which Humans do well;  
Render unto machines, that which machines do well*