Applying the Cognitive Continuum Theory to the Analysis of Human Phishing Email Judgments

Kylie A. Molinaro, Ph.D.[1,2] and Matthew L. Bolton, Ph.D.[1]
[1]University at Buffalo, [2]Johns Hopkins University Applied Physics Laboratory

Introduction

With the growing threat of phishing emails and the limited effectiveness of current mitigation approaches, there is an urgent need to better understand what leads to phishing victimization. Although previous research identified cognitive automaticity as a potential reason behind victimization[1], more research is needed. Prior research also has not considered the characteristics of the environment in which these judgments are made. This work aimed to fill these gaps with a novel combination of theories, analysis techniques, and measures.

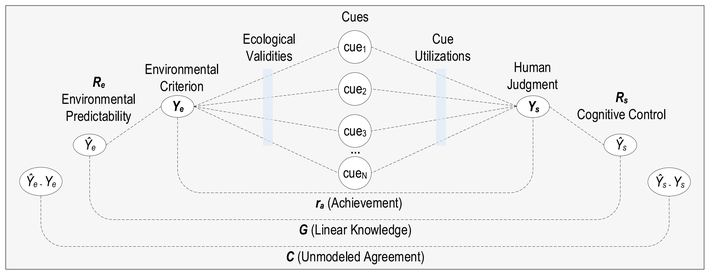
Background

PHISHING

- Malicious messages designed to appear legitimate in an attempt to get users to perform compromising actions
- Around 90% of cyber attacks begin with a phishing email

JUDGMENT ANALYSIS: THE LENS MODEL

- Technique for analyzing how people make judgments of distal criteria (the environment) using proximal cues (information in the environment) with symmetric statistical models of the environment and the judgment values made by the human[2]
- Lens model equation calculates achievement (r_a: measure of performance)
  - $r_a = GR_eR_s + C\sqrt{1-R_e^2}\sqrt{1-R_s^2}$



COGNITIVE CONTINUUM THEORY (CCT)

- Represents cognition with a continuum (versus a dichotomy) and was originally proposed by Kenneth R. Hammond[3]
- Cognitive implications of task characteristics and the human's cognition can be understood by computing task continuum index (TCI) and cognitive continuum index (CCI) scores
  - Calculated with a combination of lens model and other measures
  - Large differences between TCI and CCI scores have been associated with more judgment errors

Methodology

EXPERIMENTAL TASK AND PARTICIPANTS

- Participants sorted 40 emails (20 legitimate and 20 phishing) into "keep" or "suspicious" folders – all phishing emails were link-based attacks
  - All emails real phishing and legitimate emails
- 74 participants through Amazon Mechanical Turk
- Demographics and post task questionnaires and task instructions presented through Qualtrics
- Interacted with emails through Roundcube, a web-based email client

INDEPENDENT VARIABLES

- Dichotomous criterion: 1 for phishing, 0 for legitimate
- Dichotomous cue coding: 1 for present, 0 for absent

DEPENDENT MEASURES

- Judgment the participant made about an email:
  1 if sorted into "suspicious", 0 if sorted into "keep"
- Time to complete email sorting task
- Confidence rating (1-10) for each judgment

Data Analysis and Hypotheses

TCI SCORE CALCULATION

- **Measures:**
  - Number of cues
  - Cue redundancy
  - Standard deviation of cue weights
  - Degree of non-linearity in organizing principle
  - Degree of certainty in the task system
- All transformed to a 1-10 scale then averaged together
- **Hypothesis 1**: The task will have a TCI score oriented towards the analytical side of cognition.

CCI SCORE CALCULATION

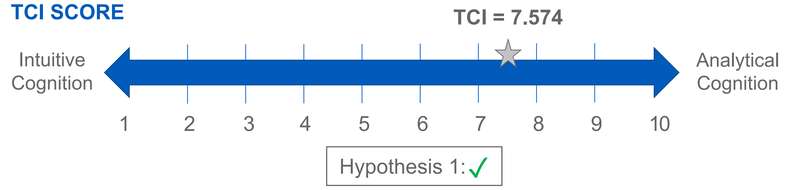
- **Measures:**
  - Cognitive control (R_s)
  - Overestimation[4]
  - Overprecision[4]
  - Degree of non-linearity in organizing principle
  - Response rate
- All transformed to a 1-10 scale then averaged together
- **Hypothesis 2**: Achievement will be positively correlated with CCI score.

DIFFERENCE BETWEEN TCI AND CCI SCORES CALCULATION

- $|CCI - TCI|$
- **Hypothesis 3**: Achievement will be negatively correlated with $|CCI - TCI|$.

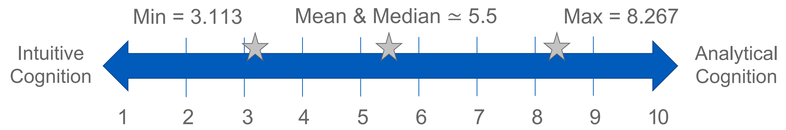
Results

TCI SCORE



Hypothesis 1: ✓

CCI SCORES



- Correlation between achievement and CCI: $r(72) = 0.744$, $p < 0.001$

Hypothesis 2: ✓

DIFFERENCE BETWEEN TCI AND CCI

	Mean	Median	Min	Max
$\lvert CCI - TCI \rvert$	2.100	2.032	0.013	4.461

- Correlation between achievement and $|CCI - TCI|$: $r(72) = -0.741$, $p < 0.001$

Hypothesis 3: ✓

Discussion

This work applied the CCT to a novel domain to understand how cognition affected phishing victimization. It was the first research to analyze the task characteristics along with user cognition in this domain. The results showed a clear relationship between cognition and performance and the task was best suited for more analytical cognition. These results have direct implications for combating phishing including: training, interface design, and user screening.

References

1. A. Vishwanath, B. Harrison, and Y. J. Ng (2016), "Suspicion, cognition, and automaticity model of phishing susceptibility".
2. E. Brunswik (1955), "Representative design and probabilistic theory in a functional psychology".
3. K.R. Hammond, R.M. Hamm, J. Grassia, and T. Pearson (1987), "Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment".
4. J. Wang, Y. Li, and H.R. Rao (2016), "Overconfidence in phishing email detection".

Acknowledgements

The authors would like to thank Dr. Anton Dahbura and Dr. Xiangyang Li from the Johns Hopkins University Information Security Institute and Dr. Nathan Bos from the Johns Hopkins University Applied Physics Laboratory for allowing them to use the data collected under the National Science Foundation Award 1544493 for the work presented here.

University at Buffalo The State University of New York
kyliemol@buffalo.edu
Department of Industrial and Systems Engineering
School of Engineering and Applied Sciences