Deliverable for WP 5.1

1. INTRODUCTION

This work package was optional and it was led by UoS. The main aim was to explore the extension of CogTool+ and cognitive modelling tools in general to cover human macro behaviours such as reasoning modes and decision-making processes.

2. LINES OF RESEARCH

UoS initiated a strand of research on users' perceived trustworthiness of sources of information in online environments (e.g., [1], [2]). By trusting unreliable sources, users can threaten the security of an online system. We thus investigated users' credibility judgements of Twitter profiles and the factors (gender of the profile owner's, visual and verbal cues) that can influence those judgements. The results showed that, overall, participants were able to detect the inauthentic nature of the profiles they were shown. Furthermore, participants showed more confidence in their credibility evaluations when competence-related trait adjectives were included in the profile summaries (please see [3] for further details). Another study we conducted used the cognitive architecture and software ACT-R to simulate human behaviour in a phishing scenario (please see [4] for further details).

Based on these preliminary studies, a future strand of research could systematically investigate credibility source and users' perceived trustworthiness of Internet-based media (e.g., websites) and data (e.g., news). Building on the eye-tracker studies conducted for work packages 1-2, future research could investigate and model users' decision-making strategies when they determine if a website is secure or not. Eye-tracking data could provide useful information in terms of attention-grabbing and attention-holding stimuli and users' visual scan paths. These implicit measures could complement users' explicit self-reports in the analysis of users' behaviours. This research could take into account variables such as users' age and expertise (e.g., trained vs untrained users) that could influence authentication strategies and error rates. For example, users' high confidence in the security of a user interface could determine their choice of a weak password for the authentication to a system they trust.

Another possible extension to cover users' macro behaviours could focus on the influence of time constraints on user authentication. Dual-process theories in psychology argue that human being can resort to two systems of thinking and reasoning (e.g., [5], [6]). System 1 is fast, automatic, heuristic, that is, people use readily available mental shortcuts to make a judgement. System 2 is slow, analytic, and rational. Time pressure (inherent in the authentication system or contingent on situational factors) could lead users to adopt heuristics (mental shortcuts), which could be optimal (fast and efficient authentication) or suboptimal (increased error rates) depending on the authentication environment.

A better understanding of users' macro behaviours could help refine CogTool+ and other cognitive modelling tools, develop better systems to study human behaviours in cyber security and wider systems, train users and protect online environments from user-based security threats.

Another extension we could look at include how human users make unintentional errors and how different designs of user interfaces can influence the human error rate. This is a less studied area in

cyber security. The project's UK PI, Shujun Li, is currently working with an UK company Corporate Risk Associates Limited (CRA) on an Innovate UK funded project on human errors in cyber security (see here for a news release from the company on the project: https://crarisk.com/cra-awarded-rd-cyber-security-grant/).A fourth extension is on learning effects. CogTool+ and most other cognitive modelling tools consider mainly skilled users. In real-world applications, it is important to consider how novice users perform and how learning impacts their performance. This is also useful for skilled users as even they have to keep learning new things related to cyber security (e.g. when the operating system and software used are upgrades to new versions with different user interfaces). This can be handled by providing different parameters to the same cognitive models (e.g., the parameters *a* and *b* in Fitts's Law for typing), but in some cases will require different models as novice users may have a different mental model when doing the same task.

At UoS we plan to apply for a new EPSRC project to investigate some of the above possible extensions, led by the project CI Patrice Rusconi and participated by the project PI Shujun Li.

3. REFERENCES

[1] C. L. Toma, "Counting on Friends: Cues to Perceived Trustworthiness in Facebook Profiles," *Proc. Eighth Int. AAAI Conf. Weblogs Soc. Media Count.*, pp. 495–504, 2014.

[2] S. Shyam Sundar, "The MAIN Model: A Heuristic Approach to Understanding Technology Effects on Credibility," *Digit. media, youth, Credibil.*, pp. 73–100, 2008.

[3] C. Sandy, P. Rusconi, and S. Li, "Can humans detect the authenticity of social media accounts? On the impact of verbal and non-verbal cues on credibility judgements of twitter profiles," in *2017 3rd IEEE International Conference on Cybernetics, CYBCONF 2017 - Proceedings*, 2017.

[4] N. Williams and S. Li, "Simulating human detection of phishing websites: An investigation into the applicability of the ACT-R cognitive behaviour architecture model," *2017 3rd IEEE Int. Conf. Cybern. CYBCONF 2017 - Proc.*, 2017.

[5] J. S. B. T. Evans, "Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition," *Annu. Rev. Psychol.*, vol. 59, no. 1, pp. 255–278, 2008.

[6] J. S. B. T. Evans, "Dual-process theories of reasoning: Contemporary issues and developmental applications," *Dev. Rev.*, vol. 31, no. 2–3, pp. 86–102, 2011.