The description-experience gap in the effect of warning reliability on user trust, reliance, and performance in a phishing context

J. Chen, A. Xiong, N. Li and R. Proctor, USA

ABSTRACT: Automation reliability is an important factor that may affect human trust in automation, which has been shown to strongly influence the way the human operator interacts with the automated system. If the trust level is too low, the human operator may not utilize the automated system as expected; if the trust level is too high, the over-trust may lead to automation biases. In these cases, the overall system performance will be undermined --- after all, the ultimate goal of human-automation collaboration is to improve performance beyond what would be achieved with either alone. Most of the past research has manipulated the automation reliability through "experience". That is, participants perform a certain task with an automated system that has a certain level of reliability (e.g., an automated warning system providing valid warnings 75% of the times). During or after the task, participants' trust and reliance on the automated system is measured, as well as the performance. However, research has shown that participants' perceived reliability usually differs from the actual reliability. In a real-world situation, it is very likely that the exact reliability can be described to the human operator (i.e., through "description"). A description-experience gap has been found robustly in human decision-making studies, according to which there are systematic differences between decisions made from description and decisions from experience. The current study examines the possible description-experience gap in the effect of automation reliability on human trust, reliance, and performance in the context of phishing. Specifically, the research investigates how the reliability of phishing warnings influences people's decisions about whether to proceed upon receiving the warning. The effect of the reliability of an automated phishing warning system is manipulated through experience with the system or through description of it. These two types of manipulations are directly compared, and the measures of interest are human trust in the warning (a subjective rating of how trustable the warning system is), human reliance on the automated system (an objective measure of whether the participants comply with the system's warnings), and performance (the overall quality of the decisions made).