Name of theory
Signal detection theory (also known as detection theory)

	Respond "absent"	Respond "present”
Stimulus present	Miss	Hit
Stimulus absent	Correct rejection	False alarm

Figure 1. Overview of the various options in a response matrix.

Description
Signal detection theory (SDT) is a method to differentiate between various actions and consequences in decision making processes. Other theories might simply consider a decision to be right or wrong, but signal detection theory differentiates those responses based on what the correct response would have been. SDT does this by creating a 2*2 matrix of responses and stimuli. If the stimulus is present and the response is “present” it is a hit, if the stimulus is absent and the response is “absent” it is a correct rejection. In both cases, a correct decision was made. If the stimulus is present, but the response is “absent”, this is a miss, as the stimulus was present but was not recognised as such. If the stimulus is absent but the response is “present”, this is a false alarm, as people would say they perceived the stimulus while in fact there was no stimulus present. If people are tested repeatedly, or a large number of people are tested once, it is possible to populate the matrix with the number of people who responded to the presence/absence of a stimulus in a presence/absence way. Using these matrices can then inform future training.

Application within the field of cybersecurity
In the case of phishing, many organisations want employees to report suspicious emails, instead of clicking on URLs in those emails or engaging with the sender in any way. Signal detection theory can be used to inform policy on the reporting behaviour of employees, and the success of a phishing awareness campaign. For example, if some after taking part in a campaign is overly cautious, reporting many emails that are in fact not phishing emails, the false alarm rate goes up and the department dealing with the reports could get overwhelmed with reported phishing emails. Likewise, an employee who believes they will never become the victim of a phishing attack, and therefore does not pay attention during the awareness campaign or any subsequent trainings, might ‘miss’ many phishing emails and not report them at all. Therefore, using the signal detection theory matrix can help organisations in deciding where they should put their effort in terms of phishing campaigns. This could be the reporting of suspicious emails in the first place, improving employees’ ability to differentiate between genuine emails and actual phishing attempts, or both.

Annotated bibliography

Canfield et al. (2015). Canfield and colleagues investigate whether SDT could be used to understand responses to phishing behaviour. They concluded that SDT is a useful framework, especially to differentiate between detecting a phishing email and acting accordingly. While participants knew what to do in case they received a phishing email, they were still not very well able to detect whether an email was a phishing attempt or a genuine email.

Martin et al. (2018). Martin and colleagues created a simulation of an email inbox where participants could interact with a range of emails, both genuine and phishing attempts (generic phishing attempts and spear phishing attempts. They investigated whether SDT could help explain responses to phishing attacks and showed that the equal-variance signal detection theory was useful, and that there was no need for an application of the unequal-variance signal detection theory to explain the findings.

Unchit et al. (2020). The authors used signal detection theory to assess high school students and staff's ability to detect phishing emails. They found age played a role in the number of reported hits and misses, but not in the number of correct rejections and false alarms. Similarly, there were no age differences in terms of confidence in being able to spot a phishing email.

References

Canfield, C., Fischhoff, B., & Davis, A. (2015). Using Signal Detection Theory to Measure Phishing Detection Ability and Behavior. Poster Abstr. *SOUPS 2015,* (3).

Martin, J., Dubé, C., & Coovert, M. D. (2018). Signal detection theory (SDT) is effective for modeling user behavior toward phishing and spear-phishing attacks. Human factors, 60(8), 1179-1191.

Unchit, P., Das, S., Kim, A., & Camp, L. J. (2020, July). Quantifying susceptibility to spear phishing in a high school environment using signal detection theory. In *International Symposium on Human Aspects of Information Security and Assurance* (pp. 109-120). Springer, Cham.