Human Error, Reliability, Resilience, and Performance, Vol. 132, 2024, 23–33
https://doi.org/10.54941/ahfe1004877

The Influence of Set Size in a Dynamic Decision-Making Task

Colleen E. Patton[1,2], Christopher D. Wickens[1], Benjamin A. Clegg[1,3], and Cap Smith[1]

[1]Colorado State University, Fort Collins, CO 80526, USA
[2]North Carolina State University, Raleigh, NC 27606, USA
[3]Montana State University, Bozeman, MT 59717, USA

ABSTRACT

In visual search, increasing the number of elements makes target identification more difficult. Identifying a target becomes even more complex in real world scenarios where operators might need to accumulate evidence across movement patterns – a form of dynamic decision making. In a simple simulation, participants moved their ownship around a water space, while observing movements of three, six, or nine vessels around them, one of which exhibited a hostile behavior. Results indicated that accuracy above chance of hostile ship detection did not differ across set sizes, however participants took more steps as the number of ships increased. Participants generally aggregated far less than the optimal amount of evidence, reflected in the overall average accuracy of 53%. This hints at overwhelming challenges to working memory in these types of dynamic decision-making situations. Implications for real-world scenarios and possible automated aids are discussed.

Keywords: Dynamic decision making, Multiple object tracking, Maritime, Decision making

INTRODUCTION

In naval contexts, the ability to identify a hostile entity is crucial. While current technologies often offer information about nearby entities, this information is not always available or accurate, especially at far distances (Dahbom & Nordlund, 2013). In maritime settings in particular, the movement patterns of ships can indicate potential hostility. For example, ships that move closer or follow your ownship's movements have been used to identify enemies (Lane et al., 2010; Liebhaber et al., 2002). This can be useful when automated technologies fail, such as in crowded ports or when the entities are at far distances. However, previous research has shown that humans are not proficient at identifying intent from movement patterns (Patton et al., 2021; 2022), and thus there is a need to elucidate what aspects of this task make it difficult.

To begin investigating this issue, Patton et al. (2021) developed a simplified simulation of an open water maritime environment. Participants executed discrete movements their own ship around a water space while monitoring the movement of several other ships, looking for one that expressed either of two hostile behaviors. The variability in movement and distance of the hostile ship were manipulated, and results indicated that the task was quite difficult, with accuracy hovering around 50% and declining with further distances and more variability. There are clear parallels between those results and multiple object tracking literature, which we now briefly review.

Multiple Object Tracking

In multiple object tracking (MOT) research, observers are asked to visually track one or more targets in continuous motion among other distractors, or non-targets. The targets may be readily identified, or the observer may need to determine which objects are targets. After a period of time with the objects moving, the observer is asked to identify the target(s) or respond to a probe about a single object. A number of variables have been identified as influencing the accuracy of MOT: the number of targets (e.g., Alvarez & Franconeri, 2007; Drew et al., 2011; Pylyshyn & Storm, 1988), the number of distractors (e.g., Bettencourt & Somers, 2009; Sears & Pylyshyn, 2000), the distance between the objects (e.g., Bettencourt & Somers, 2009; Franconeri et al., 2010), object speed (e.g., Holcombe & Chen, 2012; Meyerhoff et al., 2016; Tombu & Seiffert, 2011; see Meyerhoff et al., 2017 for review), and variability in the movement of the objects (Gao & Scholl, 2011; Patton et al., 2021).

The paradigm in Patton et al. (2021) invoked variability and distance, but the effect of set size (number of distractors) was not investigated. Although the MOT literature provides clear evidence that increasing set size leads to decreased accuracy, most of those studies lack an important element of real-world applicability: evidence accumulation, a critical process in the study of dynamic decision making.

Dynamic Decision Making

Dynamic decision-making paradigms typically involve evidence accumulation through a series of observations before a final decision or diagnosis is made (Edwards, 1962; Gonzales et al., 2017). These types of decisions often occur in real-world situations, such as doctors running tests to determine a diagnosis, firefighters determining the most likely place for a forest fire to spread next, and military personnel determining the hostility of an entity. Such tasks often result in poor performance by humans (Kersthold & Raaijmakers, 1997), due in part to the high demands on working memory to keep track of the multiple observations over time, and the mental resources required to timeshare perception of newly arriving evidence with current estimations of the state of the dynamic elements under surveillance. It would make sense then, that a task that requires both multiple object tracking and evidence accumulation would be quite difficult, and may see multiplicative effects of the number of objects to be monitored (i.e., set size). This challenge to human cognition would be detrimental in safety critical situations.

Current Study

The current study aimed to understand the impact of set size in a dynamic decision-making task. Using a modified version of the paradigm from

Patton et al. (2021), participants moved their ownship around an open water space. At the same time, varying numbers of ships moved around the same space, and participants were tasked with identifying which ship was exhibiting one of two potentially hostile behaviors: hunting or shadowing. Hunting ships moved closer to the usership in a stepwise fashion until eventually reaching it, and shadowing ships stalked the usership at a constant distance by performing the same movements as the usership. These mimic real life hostile behaviors (Lane et al., 2010; Liebhaber et al., 2002), and ensured participants did not develop strategies unique to a single form of hostile behavior. All the non-user-controlled ships on the screen exhibited slight variability in their movements, as if they were impacted by tides, channels, or weather events, as real ships might be. The simulated environment captures prototypical features of US Naval displays such as the Aegis (Smith et al., 2004). Participants were able to move their ship up to 35 times in discrete movements around the open water area before being required to make a report of which ship was hostile, and its behavior, although they could choose to report the hostile ship before 35 steps (at as few as 6 steps), thereby accumulating less evidence.

Three hypotheses were developed:

1. Results from Patton et al. (2021) will be replicated. This includes the interaction between behavior and distance on accuracy, such that as the distance of the hostile ship from the user ship increases, there is a detrimental effect on hunting, but not shadowing; the low overall accuracy around 50%; and the failure to accumulate sufficient information for accurate decision making.
2. Accuracy will decrease with set size, as seen in many multiple object tracking paradigms (e.g., Bettencourt & Somers, 2009; Sears & Pylyshyn, 2000).
3. In seeking more evidence (more steps), participants will be insufficiently calibrated to the changes in workload (number of distractors), as if they are "overconfident" in their ability to manage the higher workload situations (Wickens et al., 2021; Moore, 2020; Horrey et al., 2015). In Patton et al., (2021), participants generally used less than 20 steps out of the 35 available, and their mean accuracy was only around 50%.

METHODS

Participants

This research complied with the American Psychological Association Code of Ethics and was approved by the Institutional Review Board at Colorado State University. Informed consent was obtained from each participant. Data was collected from 35 people on Prolific, all of whom were located in the United States.

Task

Participants viewed a computer screen (see Figure 1) containing a yellow cross indicating their ship's position, which they could control, and varying numbers (3, 6 or 9) of white circles with numbers which represented other ships and were controlled by a software application. In Figure 1, there are six ships (N = 6).

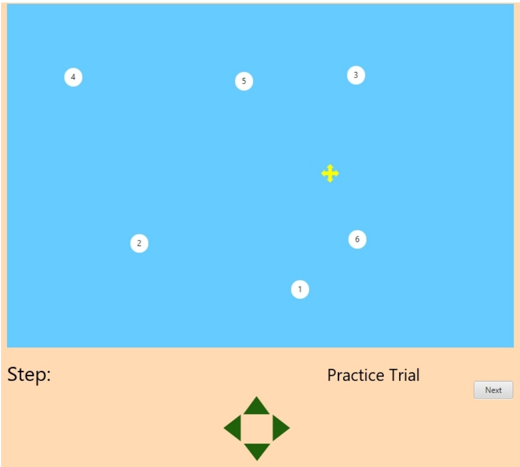

Figure 1: Screen exhibiting the experimental paradigm. Participants controlled the yellow arrow while the white circles represent computer-controlled ships.

On each trial, the starting location of all ships was randomly generated. The participant's ownship could be moved in one of 4 directions (up, down, left or right) by clicking the arrow keys at the bottom of the screen. These arrow keys could only be clicked once per second to negate the potential to create apparent motion through rapid keystrokes and also to mimic the somewhat sluggish dynamics of real vessels on the water. The movement of the participant's ship on the screen was accompanied by an update of the computer-controlled ships, although these ships were able to move diagonally. Thus, all ships moved at the same time, with at least a one second delay in between movements.

On each trial, one of the computer-controlled ships was randomly selected to act in a hostile fashion. The hostile ship's movement was contingent on the user's movements. The hostile ship would do one of two things – hunt or shadow. Hunting meant moving in a way such that it would eventually reach the user ship. An algorithm computed which directional movement produced the greatest reduction in distance between the two and moved the ship in that direction as the user ship moved. Shadowing aimed to generally keep a consistent distance from the user ship through mirroring their movements. For instance, if the user moved left, the shadowing ship also moved left. If the user ship moved towards the shadowing ship, the latter moved the same direction as the user so the distance between the ships stayed the same. These target movements occurred simultaneously with the usership movement that triggered it.

The other ships (distractors) on the display moved independently of the user's actions. The behavior of the non-hostile ships was randomly assigned other movement patterns. Some of the ships moved towards their own fixed target location, coded as an invisible point on the coordinate grid. Other ships exhibited "patrol" behaviors, where they moved in a rectangular course that covered either 1/3, 1/2 or 2/3 of the screen. The rectangular course could be oriented in any direction and the ship could start at any point on the path.

Movements of all computer-controlled ships contained 25% noise, such that, on average, every one out of four steps was not as expected for that ship's behavior. For example, if the hostile ship was shadowing, approximately every one out of four steps would not be the same as the user ship. Similarly, the movement of the non-hostile ships departed from their assigned behavior approximately once in every four movements.

Two initial practice trials demonstrated hunting behavior and shadowing behavior, with no data collected. Unlike in the experiment trials, on each practice trial the hostile ship was a different color and the hostile behavior was announced when the trial started. This allowed participants to practice working through a scenario but also illustrated the difference between hostile behaviors.

On each trial, the participant was required to make at least five steps, but no more than 35 steps, in whatever pattern they chose to probe the target's behavior, before determining which ship they believed was hostile. Once they made a decision, they clicked an "End" button. The ship display froze, and the participant indicated whether they were being hunted, shadowed, or neither. If they chose hunting or shadowing, the next question asked them to choose which ship was exhibiting that behavior by clicking the radio button that matched the ship number they believed was hostile. They then clicked "submit" and were given feedback only on the correctness of their response, but not on the correct target nor the hostile behavior exhibited on the trial.

Design

Participants completed 36 trials in three blocks of twelve trials each. Each block had either three, six or nine computer-controlled ships, and each trial would involve a randomly selected ship as the hostile target and a randomly selected hostile behavior type. As a result, each trial had one hostile ship and either two, five or eight non-hostile distractor ships. The blocks were presented in a random order for each participant.

RESULTS

Accuracy

Overall, participants correctly identified the hostile ship 53% of the time. Hunting ships were correctly identified 59% of the time, whereas shadowing ships were only correctly identified 47% of the time ($t(34) = 3.77$, $p < 0.001$, $d = 0.46$).

Number of Ships

When considering accuracy across set sizes, we use the metric of "accuracy above chance." The odds of a participant randomly guessing the correct ship and behavior is much higher at 3 ships than at 9 ships. To remedy this, chance accuracy of choosing the correct ship and behavior (16% for two distractor ships, 10% for 5 distractor ships, 8% for 8 distractor ships) was subtracted from the raw accuracy score. When this metric (accuracy above chance) was used, there was no main effect of number of ships ($F(2,68) = 1.22$, $p = 0.30$, $\eta^2 = 0.005$; Figure 2).

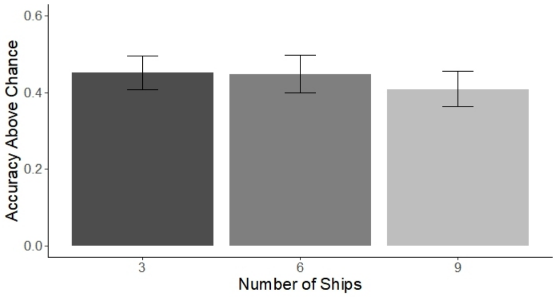

Figure 2: Accuracy above chance for each set size. Error bars represent one standard error.

Distance

Figure 3 shows the percent of correct hostile ship identifications as a function of starting distance between the hostile ship and user ship. Distance is categorized into four quartiles, with one being the closest distances and four being the farthest. Due to the randomization of starting distance, not all participants had trials with each behavior in each quartile, so a mixed linear model was run. There was a main effect of distance (estimate = −0.20, SE = 0.01, $t = -11.21$, $p < 0.001$, 95% CI[−0.24, −0.17]). There was also a main effect of behavior (estimate = −0.48, SE = 0.07, $t = -6.79$, $p < 0.001$, 95% CI[−0.62, −0.34). The interaction was significant (estimate = 0.14, SE = 0.02, $t = 5.49$, $p < 0.001$, 95% CI[0.09, 0.19]), which can be seen in the large decrease in accuracy as distance increases for hunting behaviors, but not shadowing, replicating the effects found in Patton et al. (2021; 2022), and confirming the first hypothesis.

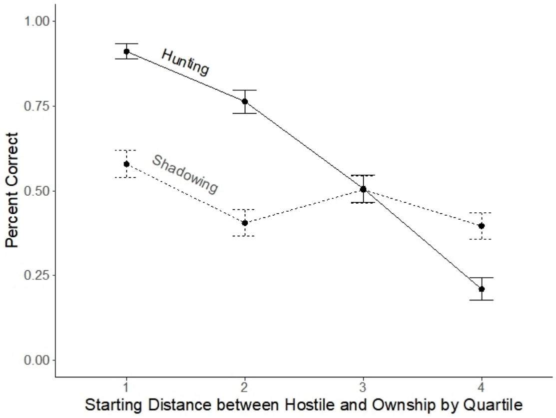

Figure 3: Accuracy of hostile ship detection by the behavior of the hostile ship and its initial distance from the usership. Error bars represent one standard error.

The joint effects on accuracy of distance and the number of distractors are shown in Figure 4. Here again, accuracy above chance was used. Because the starting location of ships was randomized, the distribution of data in the various conditions was uneven. Therefore, a mixed linear model was run. Distance significantly affected accuracy above chance (estimate = −0.13, SE = 0.02, t = −4.58, p < 0.001, 95% CI[−0.18, −.07]), with further distances leading to lower accuracy. There was no effect of the number of distractors on accuracy above chance (as seen in Figure 2; estimate = −0.01, SE = 0.01, t = −1.15, p = 0.24, 95% CI[−0.04, −0.01]). There was also no significant interaction (estimate = −0.002, SE = 0.005, t = −0.39 p = 0.69, 95% CI[−0.01, 0.008]).

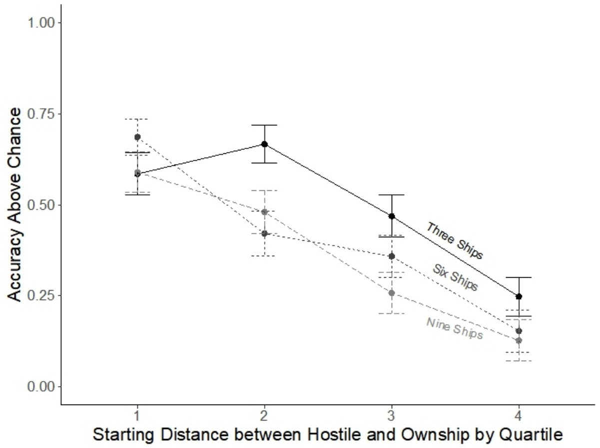

Figure 4: Accuracy above chance by set size and the starting distance of the hostile ship from the usership.

Number of Steps

The number of steps, or evidence accumulated, was compared across number of ships. With three ships, participants averaged 15.3 steps, which was significantly lower than their average of 17.5 steps with six ships (t(34) = −3.65, p < 0.001, d = 0.34), and this in turn was lower than the 19.8 steps when there were nine ships, (t(34) = −3.68, p < 0.01, d = 0.36). All three of these averages are far below the available 35 steps on each trial.

Averaged across all set sizes, the number of steps was strongly correlated with accuracy (r(33)= 0.57, p < 0.001), congruent with the idea that as participants collected more evidence, they were more accurate.

DISCUSSION

The current experiment set out to examine how changes to set size in a dynamic decision-making task impacted accuracy of identifying a target. Participants collected evidence by moving their own ship around a space while examining the movement patterns of 3, 6 or 9 other ships, one of which was a target.

The current study replicated previous findings in this paradigm (Hypothesis 1), including raw accuracy in identifying the hostile ship being higher than chance, but far from optimal. The low accuracy was seen across all set sizes, with highest accuracy at 61% - only about 50% higher than chance. This is likely due to the demands on working memory. Evidence from multiple object tracking studies suggest there are demands on working memory (Fougnie & Marois, 2009), and the current paradigm also brings in added demands of dynamic decision making (Gonzales et al., 2017) by requiring participants to accumulate evidence in the form of making movements of their ship and comparing those movements to the other ships. This additional load of remembering the movements of each ship in comparison to the usership likely contributed to the low accuracy.

Additionally, we replicated the interaction between behavior and distance. The accurate detection of shadowing ships was little impacted by the starting distance from the ownship in the way that hunting ships were, the same interaction seen in Patton et al. (2021; 2022). We infer that this difference in the distance effect is due to some fundamental properties of visual attention. Previous work has shown that when objects create a polygon – in the current case, this would be an imagined line of constant orientation and length between the usership and the suspected shadowing ship – they are easier to track (Meyerhoff et al., 2017). Furthermore, this tracking is unaffected by the distance of separation because the perception of a constant orientation is little affected by the length of the line (i.e., the vector connecting the ownship to the shadowing ship which is the distance from ownship).

Hypothesis 2 examined the effect of set size on accuracy above chance, and unexpectedly, there were no significant differences in accurate detection of the hostile ship as set size increased. This is unlike previous work in multiple object tracking, which shows decreased accuracy as set size increases (Bettencourt & Somers, 2009; Sears & Pylyshyn, 2000). It is possible that, to some extent, this null effect of set size might have resulted from the overload on working memory at even the smallest set size, imposed by the evidence accumulation requirement. However, it is also possible that to some extent, participants were able to adopt a strategy for evidence accumulation that buffered the set-size penalties, as we discuss in the following.

Hypothesis 3 proposed that participants would be insufficiently calibrated to the increase in difficulty (distractors), and this was partially supported by the data. Overall, the current data showed a clear pattern for increasing accuracy with more steps, which was also found in previous work (Patton et al., 2021; 2022). This would suggest that if participants took more steps, they may have been able to perform better. Participants did take significantly more steps as the number of distractors increased, showing some sensitivity, but their accuracy did not change to indicate calibration to the amount of evidence that was needed to be accumulated for better performance. It may be that participants were overconfident in their ability to cope with the larger set size (Wickens et al., 2021), believing the small increase in evidence was enough. It is also possible that working memory in this task is already overloaded at as few as three ships, and hence additional evidence simply could not be accumulated.

These findings warrant more research but bring important real-world considerations. Naval officers attempting to locate hostile entities in crowded waterways are unlikely to have the luxury of time. With potentially hundreds of ships around them, and their decisions being safety-critical, these results indicate that a supporting aid should be implemented to increase their speed and accuracy in these decisions. Previous work has shown that a visual aid denoting the previous location of ships (an easy to implement "history trail") can improve performance slightly (Patton et al., 2022). The increase in raw accuracy with smaller set sizes suggests that decluttering the radar screen to only denote likely hostile ships may also be a viable starting point. Yet, there is still a need to determine what type of aid could speed up the decision process and optimize accuracy, especially with larger numbers of ships.

Limitations

The current paradigm aims to answer basic research questions with real-world applications, which inherently creates limitations. The generalizability of the current paradigm is limited by the use of naïve participants without experience in maritime scenarios. Additionally, the paradigm was a simplified version of a real naval task, which eliminates some of the complexities that could impact human behavior in the real world. However, by using a controlled paradigm in a laboratory space, we were able to understand the impact of set size in a dynamic decision-making task, which can be used as a foundation for future research.

CONCLUSION

The current research provides insight into the demands on human operators when completing a multiple object tracking task in a dynamic decision-making setting. Low accuracy in identifying the hostile ship, but limited changes by set size, suggest that working memory limitations make identifying the hostile ship (target) difficult, but the ability to accumulate evidence may be able to help offset the detrimental effect of increasing set size that is typically seen in multiple object tracking scenarios. In real world contexts, where there can be dozens of objects to search through with tight time constraints, there is a need to implement an aid that supports working memory demands to encourage faster evidence accumulation and more accurate decisions.

ACKNOWLEDGMENT

The authors would like to acknowledge Dr. Jeffery Morrison at the Office of Naval Research for his support of this work under grant numbers N00014-17-1-2825 and N00014-20-1-2518.

REFERENCES

Alvarez, G. A., & Franconeri, S. L. (2007). How many objects can you track?: Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision, 7*(13), 1–10. doi: 10.1167/7.13.14.

Bettencourt, K. C., & Somers, D. C. (2009). Effects of target enhancement and distractor suppression on multiple object tracking capacity. *Journal of Vision, 9*(7), 1–11. doi: 10.1167/9.7.9.

Dahlbom, A., & Nordlund, P. J. (2013). Detection of hostile aircraft behaviors using dynamic bayesian networks. *In Proceedings of the 16th international conference on information fusion* (pp. 2033–2040). IEEE.

Drew, T., Horowitz, T. S., Wolfe, J. M., & Vogel, E. K. (2011). Delineating the neural signatures of tracking spatial position and working memory during attentive tracking. *The Journal of Neuroscience, 31*, 659–668. doi: 10.1523/JNEUROSCI.1339–10.2011.

Edwards, W. (1962). Dynamic decision theory and probabilistic information processing. *Human Factors, 4*, 59–73.

Fougnie, D., & Marois, R. (2006). Distinct capacity limits for attention and working memory evidence from attentive tracking and visual working memory paradigms. *Psychological Science, 17*, 526–534. doi: 10.1111/j.1467–9280.2006.01739.x.

Franconeri, S. L., Jonathan, S. V., & Scimeca, J. M. (2010). Tracking multiple objects is limited only by object spacing, not by speed, time, or capacity. *Psychological Science, 21*, 920–925. doi: 10.1177/0956797610373935.

Gonzales, C., Fakhari, P., & Busemeyer, J. (2017). Dynamic Decision Making: Learning Processes and New Research Directions. *Human Factors, 59*, 713–731.

Holcombe, A. O., & Chen, W. Y. (2012). Exhausting attentional tracking resources with a single fast-moving object. *Cognition, 123,* 218– 228. doi: 10.1016/j.cognition.2011.10.003.

Horrey, W. J., Lesch, M. F., Mitsopoulos-Rubens, E., & Lee, J. D. (2015). Calibration of skill and judgment in driving: Development of a conceptual framework and the implications for road safety. *Accident Analysis & Prevention, 76*, 25–33.

Kersthold, J. H., & Raaijmakers, J. G. W. (1997). Decision making in dynamic task environments. In R. Ranyard, R. W. Crozier, & O. Svenson (Eds.), *Decision making: Cognitive models and explanations* (pp. 205–217). Routledge.

Lane, R. O., Nevell, D. A., Hayward, S. D., & Beaney, T. W. (2010). Maritime anomaly detection and threat assessment [Conference session]. 2010 13th International Conference on Information Fusion. IEEE

Liebhaber, M. J., Kobus, D. A., & Feher, B. A. (2002). Studies of U. S. Navy air defense threat assessment: Cues, information order, and impact of conflicting data. *Technical Report*, SSC-1888. Meyerhoff, H., Pappenmeier, F., & Huff, M. (2017). Studying visual attention using multiple object tracking paradigm: A tutorial review. *Attention, Perception & Psychophysics, 79*, 1255–1274.

Meyerhoff, H. S., Papenmeier, F., Jahn, G., & Huff, M. (2016). Not FLEXible enough: Exploring the temporal dynamics of attentional reallocations with the multiple object tracking paradigm. *Journal of Experimental Psychology: Human Perception and Performance, 42,* 776–787. doi: 10.1037/xhp0000187.

Moore, D. A. (2020). *Perfectly confident: How to calibrate your decisions wisely.* Harper Collins.

Patton, C. E., Wickens, C. D., Smith, C. A. P., & Clegg, B. A. (2021). Detection of hostile intent by spatial movements. *Human Factors*. https://doi.org/10.1177/00187208211015022

Patton, C. E., Wickens, C. D., Clegg, B. A., Noble, K. M., & Smith, C. A. P. (2022). How history trails and set size influence detection of hostile intentions. *Cognitive Research: Principles and Implications, 7*(1), 1–12.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3*, 179–197. doi: 10.1163/156856888X00122.

Sears, C. R., & Pylyshyn, Z. W. (2000). Multiple object tracking and attentional processing. Canadian Journal of Experimental Psychology, 54, 1–14. doi: 10.1037/h0087326.

Smith, C. A. P., Johnston, J., & Paris, C. (2004). Decision support for air warfare: Detection of deceptive threats. *Group Decision and Negotiation, 13*, 129–148. https://doi.org/10.1023/ B: GRUP.0000021842. 5906.94

Tombu, M., & Seiffert, A. E. (2008). Attentional costs in multiple-object tracking. *Cognition*, 108, 1–25. doi: 10.1016/j.cognition.2007.12.014.

Wickens, C. D., Helton, W. S., Hollands, J. G., & Banbury, S. (2021). *Engineering psychology and human performance*. Routledge.