Supporting decision making in a simulated air defence activity

Chris Baber[1] and Chris Vance[2]

[1]University of Birmingham, UK, [2]MBDA, UK

ABSTRACT

A simulated air defence task is used to explore the impact of decision support on operator performance under different levels of task complexity. In this simulation, the status of autonomous air vehicles (drones) is indicated by their colour and by their threat level. Threat level is indicated by a polygon display which is automatically updated when drones fly into areas of attack or areas of risk. From two experiments, we can draw some tentative conclusions on the strategies that participants employ and the role that decision support might play in supporting or thwarting these. Contrary to instructions, participants did not always respond to the cue from the polygon display to engage (particularly when the number of drones was high, which resulted in more dynamic changes to the polygon display). Discussion with participants after the experiments suggested that some of them tried to 'read' the changes in polygon display in terms of possible paths that the drones were taking. From this, they might have attempted to anticipate when to respond and rely on their anticipation rather than the decision support. Situation awareness was rated lower when participants monitored two types of drone, and this was sufficient to lead to them performing at levels not far from chance. This is concerning in that it suggests that the decision support (in the polygon display) was not regarded as part of the situation awareness that the participants were using. It also raises the possibility of a difference between awareness of the situation and the awareness of decision options.

KEYWORDS

Air defence, decision support, polygon displays

Introduction

Air defence, particularly in busy air spaces, can be cognitively demanding (Zak et al., 2018). To support this, operator decision making can be supported, for example, through colour coding of the entities in an air space, through indication of zones in which activity might be performed, or through indication of flight paths. In this study, a simulated air defence activity was played in three conditions: one in which the participants only saw the decision support and could not see the drones; one in which participants received no decision support but could see the movement of drones; and one in which participants saw a combined view of decision support and the movement of drones.

Simulating air defence with the drones game

The drones game is a simple, one-player game that emulates some aspects of air defence. There are moving objects that need to be classified as a target and then acted upon. As the objects fly into range of a beacon, the beacon can be activated to reduce power in the drone until it crashes. In the experiments in this paper, object classification is performed automatically, in other words, objects are coloured red, yellow, blue or green to reflect threat level (from hostile and unknown to neutral and friendly); a subset of the drones are presented as yellow and then change colour to red when they enter the range of the beacon. At set up, the game divides the number of drones selected for the session in half to ‘hostile + unknown’ and ‘neutral + friendly’, and then further divides these groups to have similar numbers of each type (for odd integers, this will be one more hostile drone in the first group and one more neutral in the second group). Figure 1 shows the game screen, with objects (drones) flying over houses in ‘Los Angeles’. Houses that have a border around them are being protected from those drones that have hostile intent. These are coloured red on the screen. In the instructions to participants, ‘hostile intent’ was defined in terms of airborne cameras being flown by paparazzi to take photographs of ‘celebrities’. In order to disable a drone, a beacon can be selected (by pressing corresponding number keys on the keyboard) and activated (by pressing the space-bar) – it is also possible for the beacon to be selected using a mouse and the lightning flash used to turn the beacon on or off, but this functionality was not used in the experiments in order to control for consistency between participants. If the drone is disabled (and destroyed) in the vicinity of the protected house, then this could cause damage and so the activation on a beacon should be performed when the drone enters the beacon range (indicated by the green circle) and outside the area of risk (highlighted in orange). When the beacon is active, its range decreases (corresponding to use of energy). In order to recharge the beacon, it needs to be turned off (which requires the participant to select the beacon and use the space bar to toggle it off).

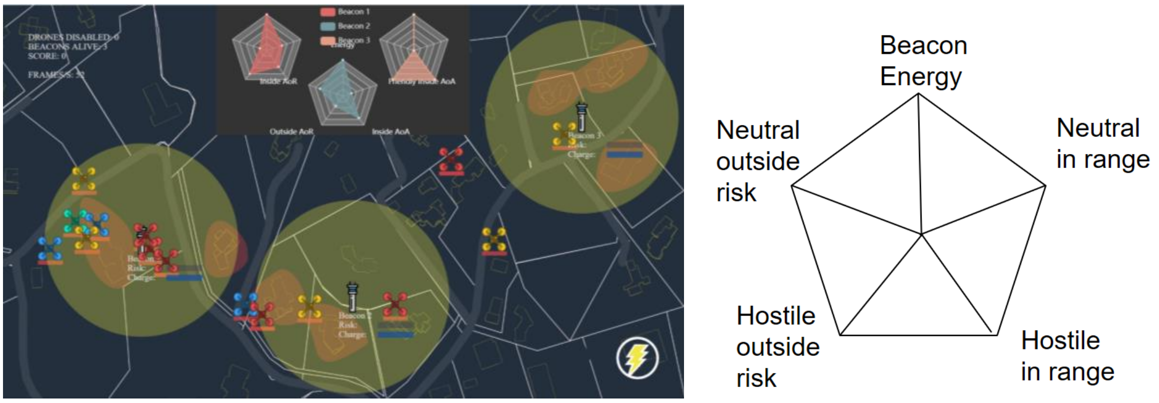

Figure 1: Drones game user interface

In the top, centre of figure 1 there are three polygon displays (one for each beacon). An example of this is shown to the right of figure 1. The polygon indicates the threat level, defined by a drone (either neutral or hostile) entering an area of risk or being in range. This indicates when a response is required. Whether the beacon can be activated is indicated by its energy level. The polygon display conveys the state of the game and indicates when it is appropriate to activate a beacon. When a hostile drone is outside the risk area and in range of the beacon, then the lower triangle of the polygon is filled. When this happens, the user will select that beacon and activate it.

Experimental design

It was expected that the use of the polygon display should simplify the task: when the bottom triangle of the polygon was filled, participants should turn on the relevant beacon.

Participants

Thirty people took part in the experiments (12 female; 18 male; average age c.24 years). Participants were recruited from the population of postgraduate students at the University of Birmingham. 28/30 participants played video games on a regular basis and all had good (self-rated) knowledge of game design. For participation, they received a £10 Amazon voucher on completion of all tasks. The same participants were involved in both experiments; the order in which they completed experiments was counter-balanced (to minimise learning or order effects). There was a break between each block of trials and the experiments were completed in an hour.

Procedure

The study received MODREC approval (914/MODREC/19) and ethics approval from the University of Birmingham (ERN_18-1988). On arrival, a demonstration of the drones game (using six drones) was given to each participant. The role of the participant was explained as a member of a company that provided security for celebrities in Los Angeles, and that they needed to deactivate drones that were threatening the celebrities while allowing the other drones to continue flying (the blue and green drones represented, for instance, delivery drones). A Latin Squares design was used to allocate participants to conditions. In both experiments, participants completed the task three times: once with the polygon display only (auto), once without automated support (self), and once with drones and polygon (both). In Experiment 1, participants responded to only the red drones. In Experiment 2, participants were required to respond to both red and yellow drones. In each experiment, half of the participants were presented with a total of nine drones and half were presented with a total of 18 drones.

Data collection, preparation and statistical testing

The drones game records the timing of actions performed by the participant, for example, beacon activation (on or off), and the state of all drones, for example, enter or leave an 'area of attack' or 'area of risk', or when a drone is 'killed', and the state of each beacon – on, off, energy level. These data are used to provide four objective measures of performance:

1) Task completion time: total time from game start until all 'hostile' drones are destroyed.
2) Time per target: the task completion time divided by the number of drones destroyed.
3) Beacon activation: the total number of times the beacons were turned on or off.
4) Signal Detection: d' = ln(((H*(1-FA))/((1-H)*FA))). For both experiments, the difference between hostile and friendly drones was used to define signal detection.

Participants were also asked to complete two subjective reports, involving rating scales:

1) NASA Task Load Index (Hart and Staveland, 1988).
2) Situation Awareness Rating Technique (Selcon et al., 1989, 1991).

For the objective measures, data were tested for normality using a Shapiro-Wilk test. Measures which were not normally distributed were subjected to Box-Cox transformation and retested. Data which followed normal distribution were analysed using Repeated Measures Analysis of Variance (with Bonferroni correction for post hoc testing). Mauchly's test was applied to determine whether there was a need for correction. Data which did not follow normal distribution were analysed with the non-parametric Friedman Analysis of Variance. All statistics were calculated using R.

Experiment one: respond to red drones only.

Table 1: Results from Experiment 1 [mean (sd); shading indicates significance]

	9 drones			18 drones		
	Auto	Both	Self	Auto	Both	Self
Task completion time (m)	0.47(0.3)	0.7(0.5)	0.7(0.5)	0.5(0.2)	1 (1)	0.8(0.4)
Time per target (s)	13(8)	20(14)	20(17)	12(6)	22(19)	18(6)
Beacon activation	5.1(2.7)	6.9(7.5)	5.8(4)	4.3(2.5)	5.5(3.8)	6.9(5)
d'	12.6	12.2	11.3	6.8	5.4	5
SART	11.8	8.8	9.9	6.4	8.4	7
NASA-TLX	35	17	21	39	35	35

Workload is rated as higher in the auto condition – when participants only had the polygon display and no sight of the moving drones, the task was rated as more demanding. With 18 drones, there are significant main effects of task completion time and time per target. The auto condition also has lower levels of beacon activation. While signal detection does not differ between conditions, the implication is that the auto guidance offers the potential to improve performance.

Experiment two: respond to both red and yellow drones.

Table 2: Results for Experiment 2 [mean (sd); shading indicates significance]

	9 drones			18 drones		
	Auto	Both	Self	Auto	Both	Self
Task completion time (m)	2.1(1.1)	1.7(0.8)	1.6(0.7)	6.4(3.1)	3.8(2.7)	2.9(1.4)
Time per target (s)	21(10)	19(9)	19(8)	30(14)	17(13)	13(7)
Beacon activation	27 (17)	16(5)	19(9)	41(10)	33(11)	29(12)
d'	3.4	5.2	5.5	1.1	1.5	1.5
SART	7.9	9.9	9.9	6.1	6.5	6.7
NASA-TLX	44	34	34	46	30	30

For 9 drones, there is a significant difference in signal detection between conditions, with auto having significantly worse d'. This means that participants were more likely to destroy friendly drones. Workload scores indicate that polygon displays were more demanding than the other conditions. As with experiment one, situation awareness for 18 drones are lower than 9 drones, and there is a tendency for situation awareness with polygon displays (and no view of the moving drones) to be lower than the other conditions. Task completion time and time per target are generally higher, and d' and SART scores are generally lower in experiment 2 than experiment 1. This implies that experiment 2 was more difficult than experiment 1. In contrast with experiment 1, the auto condition (for 18 drones) results in *higher* time per target and task completion.

Discussion

As task complexity increases (increase in type of hostile drone, between experiments one and two), so provision of automated support does not guarantee improved performance. Rather, participants seemed to apply a different strategy which was less selective and, therefore, less dependent on the information from the automated support. As Botzer et al. (2015) noted *"…if participants had just followed the cues and not employed any judgement of their own, performance in the task would have been at least as good or even better…"*.

References

Botzer, A., Meyer, J., Borowsky, A., Gdalyhau, I., Shalom, Y. B. (2015). Effects of cues on target search behavior. *Journal of Experimental Psychology : Applied*, 21:73-88.

Hart, S. G., Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology* 52:139-183). North-Holland.

Selcon, S. J., Taylor, R. M. (1989). Evaluation of the situational awareness rating technique (SART) as a tool for aircrew systems design. *Proceedings of AGARD AMP Symposium on Situational Awareness in Aerospace Operations, CP478,* Seuilly-sur-Seine: NATO AGARD.

Selcon, S. J., Taylor, R. M., Koritsas, E. (1991). Workload or situational awareness? TLX vs. SART for aerospace systems design evaluation. *Proceedings of the Human Factors Society 35th Annual Meeting,* Santa Monica, CA: HFES, 62-66.

Zak, Y., Oron-Gilad, T., Parmet, Y. (2018). Operator workload reduced in unmanned aerial vehicles: Making command and control (C2) maps more useful. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Los Angeles, CA: Sage. 1057-1061.