# To Swarm or Not to Swarm: Human-Drone Swarm Control Approaches in Maritime Search-And-Rescue

**Oscar Bjurling**
RISE Research Institutes of Sweden;
Linköping University
oscar.bjurling@ri.se

**Rego Granlund**
RISE Research Institutes of Sweden
rego.granlund@ri.se

**Edvin Granlund**
edvin.granlund@gmail.com

**Jens Alfredson**
SAAB Aeronautics
jens.alfredson@saabgroup.com

**Björn J. E. Johansson**
Linköping University
bjorn.j.johansson@liu.se

## ABSTRACT

Drone swarms are promising tools for search-and-rescue operations but there is currently little guidance regarding control input methods and interface design. We report a controlled within-subjects human-swarm interaction experiment comparing three approaches for supervising/controlling a 20-drone swarm in a simulated maritime search-and-rescue task: per-drone waypoint/route control; swarm-level task-area control using configurable templates; and a combined hybrid interface. Outcomes included self-rated mental workload and situation awareness, and log-derived performance metrics (objects found, time-to-first-detect, area covered/overlap, drone utilization, and command counts). Swarm-level task-area control yielded the best overall results: substantially lower workload, higher situation awareness, more objects found and area covered, higher fleet utilization, and far fewer user inputs than per-drone control. Hybrid control reduced workload relative to per-drone control but introduced additional “meta-control” demands and underperformed pure swarm-level control on detection. Findings highlight the advantages of macro-level tasking and the need to better support mode switching in hybrid control designs.

### Keywords

Drone swarm, human-swarm interaction, search and rescue, multi-UAV, unmanned aerial vehicles, human-autonomy teaming, simulation

## INTRODUCTION

Unmanned Aerial Vehicles (UAVs[1]) or *drones*, are increasingly being used in emergency response and crisis management (Mohd Daud et al., 2022; Restas, 2015). Several application areas for drones have been suggested, like monitoring forest fires and hotspots (Raczok et al., 2025; Sende et al., 2024), search-and-rescue (Grönbäck

---

[1] A variety of terms are used to denote unmanned flying platforms, such as unmanned aerial vehicle (UAV), drone, remotely piloted vehicles (RPV), remotely piloted aircraft systems (RPAS),or the more general unmanned aerial system (UAS), and sometimes under the umbrella term advanced air mobility (AAS), see Mohd Daud et al., (2022); or Johansson et al., (2025).

et al., 2025; Johansson et al., 2024), and pre-disaster activities such as mapping as well as post-event tasks such as damage assessment (Restas, 2015). Solutions range from remote control to highly automated (Mohd Daud et al., 2022) singular units to semi-autonomous swarms (Arnold et al., 2020; Lomonaco et al., 2018). Arguments for utilizing drone swarms[2] in search-and-rescue (SAR) comprise the need to cover large areas (Lomonaco et al., 2018) as well as the possibility to use different sensors for better detection (Khan & Amir, 2025) during the operation. Most studies conducted on swarming are simulation-based, as the technical maturity of swarms has been below practical applications until recently. Although most publications concern technical aspects of drone swarming, the human factors challenges associated with controlling swarms are increasingly recognized (Bjurling, 2025; Bjurling et al., 2020; Saffre et al., 2021). Such challenges must be addressed to fully take advantage of the potential benefits of drone swarming in emergency response and crisis management.

When a single human operator supervises dozens of semi-autonomous drones, the human's ability to command, monitor, and understand the swarm collective becomes a limiting factor (Lewis, 2013). This challenge has been highlighted in both cognitive modelling and architectural analyses of single-operator, multi-drone control (Cummings et al., 2007) and in design-focused swarm control frameworks that advocate scalable, indirect interaction mechanisms (Saffre et al., 2021).

Two broad strategies dominate user interface (UI) design for multi-drone systems: (1) direct per-agent control, in which operators set individual waypoints or teleoperate drones (individually or in groups); and (2) swarm-level control, in which operators influence the collective through high-level commands, such as virtual beacons or leader-based control. Each design strategy offers trade-offs in terms of performance and human factors such as mental workload or situation awareness. While per-agent control gives operators fine-grained authority, it scales poorly as drone numbers grow, increasing coordination demands and thus cognitive load. Swarm-level control, in contrast, promises reduced workload and scalable influence, but may undermine situation awareness if swarm intent or behavior is opaque. As a third control approach, supervisory control studies have shown that operator capacity is limited not only by interaction and wait times but also by cognitive reorientation and wait times that constrain *fan-out*; the maximum number of drones an operator can effectively control (Cummings & Mitchell, 2008).

To our knowledge, no controlled experiments have directly compared per-drone and swarm-level command-and-control (C2) modes (or their combination) under standardized conditions, which makes it hard to tell whether UI- and control-level differences translate to operator workload and task performance, particularly in modern swarm simulation environments. To address this gap, we conducted a controlled experiment comparing three modes—per-drone versus swarm-level C2 paradigms plus the combination thereof—on operator workload, situation awareness, and task performance using a novel multi-drone UI system.

## RELATED WORK

Human-Swarm Interaction (HSI) research has examined how operators can best influence autonomous agent collectives. Kolling et al. (2013) compared selection-based (e.g., direct or box selection of individual/groups of drones) and beacon-based interaction (e.g., placing areas that attract/repel drones or switch/toggle their behaviors) for large foraging swarms, finding that selection mechanisms scaled better and produced higher mission performance than indirect environmental cues, especially in cluttered environments. Kolling et al. (2012) makes the same core distinction, framing "selection" as an active, temporally persistent way to address specific robots and "beacon" control as a more passive, spatially persistent influence on nearby robots. Walker, Nunnally, et al. (2013) extended this line of work by varying the degree of human influence over swarms, showing an inverted-U relationship between the amount of intervention and task performance.

Supervisory control studies by Cummings and colleagues demonstrated the importance of allocating navigation and mission-level planning functions to automation to achieve scalable control (Cummings et al., 2007). Later work showed that human oversight can improve decentralized planner performance in complex networked missions (Cummings et al., 2012). This line of research is also where the "wait time" and "fan-out" notions originate, i.e., the insight that even if the automation can act on many vehicles, the *human* still becomes a queueing point (Cummings & Mitchell, 2008).

A complementary research stream has examined how these control modes affect human cognition and human-system interaction. Chen and Barnes (2012) showed that supervisory control with automation aids reduced operator workload and improved performance compared to manual control of multiple robots. Roldán et al. (2017) emphasized that interface features such as predictive cues and summarized swarm states are important for

---

[2] We use the term "drone swarm" in its colloquial sense to mean "many drones; multi-drone systems", without reference to any specific system, drone communication architectures, or local behaviour governance rules.

maintaining situation awareness (SA) in multi-robot contexts. Wong and Seet (2017) reviewed workload and SA implications in multi-robot supervision, noting that higher autonomy reduces micromanagement demands but can degrade SA without adequate transparency. Foundational multi-drone supervisory control studies also showed that while autonomy can reduce operator interaction demands, it may simultaneously introduce SA degradation and automation bias if transparency and feedback are insufficient (Cummings, 2015; Cummings & Mitchell, 2008). Design-focused research by Saffre and colleagues provides a complementary perspective. Saffre et al. (2021) explicitly contrast direct "micro-management" with indirect, macro-level interaction, arguing that micromanaging individual drones undermines swarm advantages.

To summarize, prior HSI and supervisory-control research has established (1) that operators can influence swarms through either active selection of specific robots or passive, spatially anchored commands (Kolling et al., 2012, 2013), (2) that automation should assume lower-level planning to preserve scalability (Cummings et al., 2012), and (3) that transparency is needed to support SA when autonomy increases. What is still under-reported is controlled, within-subjects comparisons of these influence paradigms under the same search/surveillance task, using the same interface substrate, and accounting for performance and human factors outcomes, which is the specific gap our experiment addresses.

## METHODS

The experiment employed a single-factor within-subjects design, with counterbalanced condition presentation order to mitigate potential sequence effects in a simulated maritime SAR scenario.

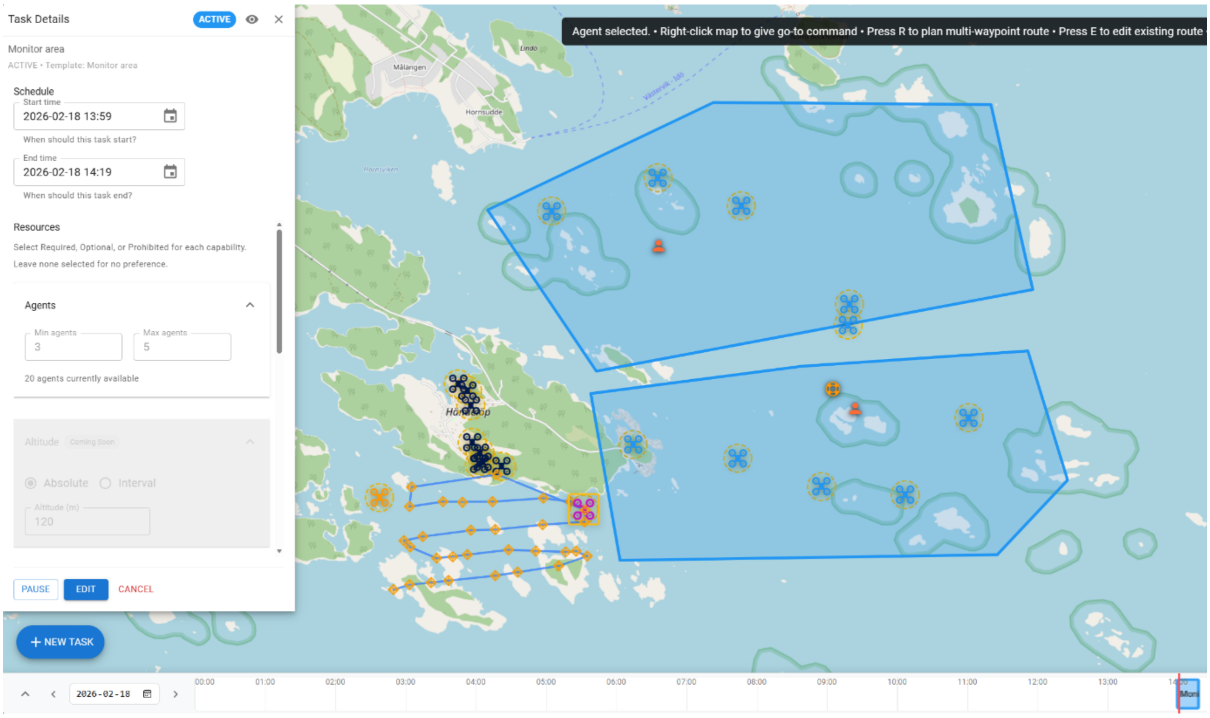
The independent variable was the control input method, which had three modes: In Condition A, drones were selected individually and manually issued go-to or route commands, corresponding to traditional multi-drone ground control station design regimes. In Condition B, the operator issued collective tasks using task templates that defined a configurable geographical area to be surveyed, and the system was delegated the responsibility of assigning drone resources and planning their routes. In Condition C, both per-drone commands and swarm-level tasking were available to the operator in the same interface. We included the combined mode explicitly because recent HSI work predicts that real mission contexts require operators to traverse several interaction strata in the same scenario, from the whole swarm, to subswarms, to individual vehicles and even to sensor/payload level, and that this multi-level traversal itself can add cognitive demands (Bjurling et al., 2020). Condition C lets us observe performance and workload when such cross-strata interaction is supported rather than forced into a single-level control scheme.

### Simulation Environment

The simulator was based on a modular concept for agent simulation and session configuration. The UI was designed with a "choir" metaphor in mind (Bjurling et al., 2021) to enable human-swarm interactions traversing swarm, subswarm, drone, and sensor/payload strata (Bjurling, 2025; Bjurling et al., 2020) following a "playbook" or mission-based approach (Giles & Giammarco, 2017; Miller et al., 2005) where configurable *task templates*—defining the geographical, temporal, and capability requirement dimensions of collective tasks—are used to issue indirect swarm-level commands. By simultaneously affording per-agent control, our design thus combines direct and indirect control mechanisms similar to the simulations developed by Saffre et al. (2022a, 2022b).

The graphic UI (Figure 1) features a map view that visualizes drones, scenario objects, and task geometries. In the map, the user can select specific drones and provide single-waypoint go-to or multi-waypoint route commands using mouse and keyboard inputs. The user can also select task templates to define (or "paint"; Saffre et al., 2021) regions of interest where drones that meet the selected task criteria should navigate and act. A task panel enables reconfiguration of task duration and agent capability requirements, like hardware (e.g., rotorcraft, fixed wing), sensors (e.g., optical or infra-red), or behavior (e.g., loiter, patrol, search patterns) requirements.

An expandable timeline display at the bottom supports operator SA by visualizing planned and ongoing tasks on a timeline. A resource dashboard displays the number of connected and available (i.e., unassigned or otherwise idle) drones, while its hover tooltip provides a collated breakdown of agent status information.



Agent selected. • Right-click map to give go-to command • Press R to plan multi-waypoint route • Press E to edit existing route

Task Details

ACTIVE

Monitor area

ACTIVE • Template: Monitor area

Schedule

Start time: 2026-02-18 13:59

When should this task start?

End time: 2026-02-18 14:19

When should this task end?

Resources

Select Required, Optional, or Prohibited for each capability.

Leave none selected for no preference.

Agents

Min agents: 3

Max agents: 5

20 agents currently available

Altitude Coming Soon

● Absolute ○ Interval

Altitude (m): 120

**Figure 1. Swarm UI (cropped). Drone icon color indicates status (e.g., idle (black), on-mission (blue), manual route (purple), returning (yellow), etc.). Blue areas indicate active collective search tasks. Task planning panel on the left configures start/end times, desired min/max number of drones, etc. Orange people/buoy icons represent located/identified objects.**

The underlying drone simulation and behavior-tree planner were identical across all conditions; only the operator control input method differed. For this pilot study, the simulation lacked computational environmental models, such as wind or sea currents, and target objects remained static throughout the trials.

## Participants

We recruited a convenience sample of 24 participants (students and the public; 13 male, 11 female), aged 19–55 years (*M* = 28.8, *SD* = 9.31). This novice sample served as an initial pilot to evaluate the feasibility of the control concepts at a foundational level. Currently, no operative drone swarms exist in maritime SAR, and authorities often avoid drone use in shared airspace with manned helicopters, meaning a pool of "expert swarm operators" is not yet available for study in Sweden. Participants reported low prior drone piloting experience (years; *M* = 0.4, *SD* = 1.47) and low-to medium Real-Time Strategy (RTS) gaming experience (5-point ordinal scale: 1=None to 5=Expert; median = 2, IQR = 1–4). Participants received no monetary compensation; light refreshments were offered. All participants provided written informed consent prior to participation.

## Procedure

All participants read and signed a written informed consent form upon their arrival for a session and filled out a separate background information form. They received the same walkthrough before starting. This included an overview and explanation of the UI elements and a general SAR briefing indicating that a rescue alert occurred somewhere north, east, and south of a starting island where the drones' home base was positioned. Participants were thus free to define their own search extent and strategy. The experimental condition order was counterbalanced. Participants were allowed to ask questions during scenarios.

The workstation featured a PC connected to a 32", 1080p resolution monitor display, and standard mouse and keyboard controls. The workstation ran a local copy of the swarm simulation platform. The participants played three scenarios with different conditions and constraints and filled out a questionnaire following each scenario. Data collection sessions lasted approximately 45 minutes.

## Scenarios

Each scenario comprised a maritime SAR task with 15 target objects (people and life rafts) statically positioned within a fixed ~22.4 km$^2$ object-placement region in an archipelagic environment setting. This region served only to bound object placement and to standardize scenario generation; it was not disclosed to participants and was not presented as a predefined search boundary.

The number of objects and the search area size were chosen to provide a problem space large enough to encourage the active use of all 20 drones, thereby testing the scalability of the control interfaces. Because the specific search space was unknown to participants and the scenario time was limited to eight minutes, participants were not expected to find all objects; thus, "time-to-first-detect" was selected as a more relevant indicator of initial search efficiency than "time-to-full-detection."

To create the three scenarios while eliminating scenario variability and minimizing learning effects, the object positions within the object-placement region were inverted (rotated horizontally or vertically while preserving overall spatial structure) to give the illusion of novel spatial object configurations for the second and third scenario encountered. Each scenario used 20 drones and ran for 8 minutes.

### Measures

The dependent variables (DVs) were mental workload (MWL), situation awareness (SA), and performance. The subjective MWL and SA DVs were recorded using a questionnaire explained in the Materials subsection. Following prior work on supervisory control and multi-drone search tasks (e.g., Cummings et al., 2012), performance was operationalized using six complementary log-derived variables capturing search effectiveness, search efficiency, coverage behavior, and interaction demand. Search effectiveness was measured as (1) objects found, i.e., the number of scenario targets detected (0-15; higher is better). Search efficiency was measured as (2) time to first detect, defined as the elapsed time (s) from scenario start until the first target was detected (lower is better). If no target was detected within the scenario duration, time-to-first detection was treated as missing. We additionally quantified spatial search behavior via (3) total area covered (km$^2$) and (4) coverage overlap, defined as the proportion (%) of covered area that was searched two or more times (lower overlap indicates less redundancy). To capture resource usage consistent with supervisory-control accounts of waiting and capacity limits, we computed (5) drone-capacity utilization as:

$$Utilization = 1 - \frac{\sum_{i=1}^{N} I_i}{NT}$$

where $I_i$ is the number of seconds drone $i$ is idle during the $T$-second scenario, and $N$ is the number of available drones. Thus, the numerator is the total idle drone-seconds across the fleet, and the denominator is the maximum possible drone-seconds (total available capacity). Higher utilization indicates that available drone resources spent more time actively executing tasks rather than waiting for new operator-issued tasks, conceptually related to "wait time" in supervisory control (Cummings & Mitchell, 2008). Finally, to quantify operator interaction demand, we computed (6) user-issued commands as the total number of discrete control inputs issued during the scenario: (#search areas created + #search areas edited) + (#route commands issued to individual drones). This measure captures the extent of manual intervention required under each control-input method.

### Materials

A self-assessment questionnaire was administered after each scenario, containing sections for each of the subjective DVs. Mental Workload was measured using the NASA TLX (Hart & Staveland, 1988) but with the Physical Demand scale removed and the Performance scale's anchor labels reversed to maintain a "low = left, high = right" response consistency. This was accounted for in post-session scoring by reverse-coding the Performance scale. The TLX items included were scored on 21-point scales which were transformed into 0-20 value scores in the analysis. Additionally, the Raw TLX (or R-TLX) scoring system was used, where the calculated mean average score of the scales represents the total MWL measure (Hart, 2006). The mean score was multiplied by five to create a total score on a 0–100 scale.

Situation awareness was measured using Braarud's SA3 instrument (Braarud, 2021), which targets the perception, comprehension, and projection levels of Endsley's original SA model (Endsley, 1995) with one 11-point Likert scale item per SA level. These items were (with scale endpoint labels in parentheses): *My observation of critical information* (Identified all needed information–Missed important information); *My understanding of what was going on* (Fully understood–Did not make sense to me); *I could look ahead, and foresee what was going to happen* (Very accurately–Could not predict). However, we opted to reverse all Likert scale anchor labels in the SA3 so that higher response values indicated increased SA. The total SA score was computed as the mean average of the three individual SA item scores.

## RESULTS

We first present the descriptive statistics for the outcome variables by condition (Table 1) and then report their mixed-effects models (Table 2) followed by post hoc comparisons (Table 3). Implications for the outcome variables are explained in subsequent paragraphs. A correlation matrix of the outcome variables is presented in Table A1 of the Appendix. Each participant completed all three scenario conditions (N = 24; 72 observations).

**Table 1. Descriptive statistics [M, (SD)] for subjective and performance outcomes by condition (N = 24 unless noted otherwise).**

| Outcome | Condition A (individual drone control) | Condition B (task area control) | Condition C (hybrid control) |
|---|---|---|---|
| MWL (RTLX) | 57.33 (16.42) | 34.50 (15.51) | 45.04 (14.32) |
| SA (SA3) | 6.00 (2.43) | 7.39 (2.01) | 7.18 (1.72) |
| Objects-Found (#) | 3.79 (1.98) | 5.50 (2.11) | 2.71 (1.94) |
| Time-To-First-Detect (s) | 198 (87.0) | 217 (76.7) | 285 (80.5)† |
| Total-Area-Coverage (km$^2$) | 7.61 (3.91) | 11.6 (3.69) | 9.57 (4.39) |
| Area-Coverage-Overlap (%) | 18.7 (7.34) | 24.7 (12.7) | 26.9 (11.8) |
| Drone-Capacity-Utilization (%) | 49.5 (23.3) | 72.0 (15.6) | 76.0 (9.31) |
| User-Control-Inputs (#) | 38.6 (35.1) | 9.33 (5.09) | 15.0 (9.96) |

Note: MWL = NASA TLX RTLX (0-100; higher = more workload). SA = SA3 mean score (1-11; higher = greater SA). Values are mean (SD). Objects found (0-15; more are better). Time-to-first-detect in raw seconds (lower is better). Total area coverage in km$^2$. Overlap in % (lower is better). Utilization is reported as percent (higher is better). User control inputs are command counts. † Condition C: N = 19 (5 missing; no objects detected during scenario).

**Table 2. Omnibus fixed effects from mixed-effect models.**

| Outcome | Condition | Presentation Order | Condition × Order |
|---|---|---|---|
| MWL (RTLX) | *F*(2, 37.9)=24.83, *p*<.001*** | *F*(2, 37.9)=5.55, *p*=.008** | *F*(4, 60.0)=2.43, *p*=.057 |
| SA (SA3) | *F*(2, 37.3)=7.69, *p*=.002** | *F*(2, 37.3)=7.57, *p*=.002** | *F*(4, 51.3)=0.51, *p*=.727 |
| Objects-Found (#) | *F*(2, 39.7)=20.60, *p*<.001*** | *F*(2, 39.7)=12.70, *p*<.001*** | *F*(4, 62.5)=1.26, *p*=.295 |
| Time-To-First-Detect (ln(s)) | *F*(2, 33.8)=7.55, *p*=.002** | *F*(2, 34.0)=0.54, *p*=.586 | *F*(4, 53.3)=0.94, *p*=.812 |
| Total-Area-Coverage (km$^2$) | *F*(2, 39.7)=9.28, *p*<.001*** | *F*(2, 39.7)=9.19, p<.001*** | *F*(4, 62.7)=0.83, *p*=.511 |
| Area-Coverage-Overlap (%) | *F*(2, 40.6)=5.94, *p*=.005** | *F*(2, 40.6)=5.30, *p*=.009** | *F*(4, 62.8)=2.74, *p*=.036* |
| Drone-Capacity-Utilization (%) | *F*(2, 40.3)=24.71, *p*<.001*** | *F*(2, 40.3)=5.10, *p*=.011* | *F*(4, 62.8)=0.45, *p*=.774 |
| User-Control-Inputs (ln(#)) | *F*(2, 39.4)=44.81, *p*<.001*** | *F*(2, 39.4)=2.12, *p*=.133 | *F*(4, 60.7)=2.75, *p*=.036* |

Note: Omnibus F-tests from LMMs include Condition and Presentation Order as categorical fixed effects (3 levels) with participant random intercepts (Satterthwaite df). Time-To-First-Detect and User-Control-Inputs were log-transformed due to right-skew. Presentation order included to account for sequence/learning effects. * p < .05, ** p < .01, *** p < .001.

**Table 3. Condition pairwise contrasts (Δ [95% CI], Holm-adjusted p).**

| Outcome | A−B | A−C | B−C |
|---|---|---|---|
| MWL (RTLX) | 22.8 [14.94, 30.72], $p$<.001*** | 12.3 [4.40, 20.18], $p$=.001** | -10.5 [-18.43, -2.65], $p$=.002** |
| SA (SA3) | -1.39 [-2.32, -0.46], $p$=.002** | -1.18 [-2.11, -0.25], $p$=.008** | 0.21 [-0.72, 1.14], $p$=.589 |
| Objects-Found (#) | -1.71 [-2.78, -0.64], $p$<.001*** | 1.08 [0.02, 2.15], $p$=.018* | 2.79 [1.72, 3.86], $p$<.001*** |
| Time-To-First-Detect (ln(s)) | -.134 [-.385, .117], $p$=.202 | -.427 [-.698, -.155], $p$=.001** | -.293 [-.564, -.021], $p$=.025* |
| Total-Area-Coverage (km$^2$) | -3.95 [-6.17, -1.72], $p$<.001*** | -1.96 [-4.19, 0.27], $p$=.072 | 1.99 [-0.24, 4.22], $p$=.072 |
| Area-Coverage-Overlap (%) | -6.00 [-12.02, 0.024], $p$=.040* | -8.25 [-14.27, -2.23], $p$=.005** | -2.25 [-8.27, 3.77], $p$=.368 |
| Drone-Capacity-Utilization (%) | -22.51 [-32.4, -12.62], $p$<.001*** | -26.49 [-36.4, -16.60], $p$<.001*** | -3.98 [-13.9, 5.91], $p$=.334 |
| User-Control-Inputs (ln(#)) | 1.367 [1.001, 1.733], $p$<.001*** | 1.026 [0.660, 1.392], $p$<.001*** | -0.341 [-0.707, 0.025], $p$=.029* |

Note: Δ is the model-estimated marginal mean difference. Brackets show 95% CIs; p-values are Holm-adjusted. Time-To-First-Detect is on the ln(seconds) scale, while User-Control-Inputs is on the ln(count) scale. Presentation Order contrasts are not shown. * $p < .05$, ** $p < .01$, *** $p < .001$.

The results indicated significant main effects of condition (i.e., control mode) on MWL and SA (Table 2). Post hoc tests identified significant differences in MWL between all conditions, where Condition A (per-drone control) induced higher MWL than both Conditions B (swarm-level control) and Condition C (hybrid control), while Condition C induced higher MWL than Condition B (Table 3). Participants also reported lower SA in Condition A than in Conditions B and C, but there were no significant differences between Conditions B and C (Table 3).

Presentation order is reported in Tables 2−3 as a categorical factor to allow for non-linear sequence effects. For brevity however, we summarize order effects in text using a linear trend parameterization (order coded 1−3 as a numeric covariate), which provides a single slope estimate consistent with an expected learning trajectory. Across subjective outcomes, significant order slopes were observed for MWL ($\beta$ = -5.40 RTLX points per position, $p$ = .003) and SA ($\beta$ = 0.74 SA3 points per position, $p$ < .001), indicating decreasing MWL and increasing SA across successive scenario positions.

Object detection count varied significantly by condition (Table 2). Post hoc contrasts showed that participants found more objects in Condition B than in Conditions A and C, and more objects in Condition A than in Condition C (Table 3). Condition further affected time-to-first detection (modeled on ln(seconds); Tables 2−3): the first detection occurred sooner in Conditions A and B than in Condition C, with no difference between Conditions A and B (Table 3).

Total area covered varied by condition and by presentation order (Table 2), where post hoc tests indicated greater area coverage in Condition B than in Condition A, whereas Condition C did not differ from the other conditions (Table 3). Complementing this, coverage overlap also differed by condition and by order (Table 2): overlap was lower in Condition A than in Conditions B and C (which did not differ) (Table 3).

Drone-capacity utilization differed by condition and order (Table 2). Post hoc contrast indicated higher utilization in Conditions B and Condition C than in Condition A, with no reliable difference between Conditions B and Condition C (Table 3).

User-issued command counts differed by condition (Table 2). Because this outcome was modeled on the ln(count) scale, we summarize multiplicative difference by exponentiating post hoc contrast estimates (ratio = exp[Δ]), Condition A required substantially more user inputs than Condition B (~3.9x) and Condition C (~2.8x), while Conditions B and Condition C differed only modestly (~0.71x; Table 3). A small Condition×Order effect was also observed (Table 2), but follow-up analyses are omitted here for brevity.

Across performance outcomes, presentation order (entered as a numeric covariate) exhibited significant linear slopes for objects found ($\beta$ = 1.10 objects per position, $p$ < .001), total area coverage ($\beta$ = 1.96 km$^2$ per position, $p$ < .001), area coverage overlap ($\beta$ = -3.01 percentage points per position, $p$ = .033), and drone capacity utilization ($\beta$ = 6.30 percentage points per position, $p$ = .002). This indicates that performance generally shifted across successive scenario positions, with more detections, broader coverage, reduced coverage redundancy, and higher utilization at later order positions. No other significant linear order effects were observed.

## CONCLUDING DISCUSSION

SAR—especially in maritime or archipelagic settings as described in this study—is a context where drone swarming is an attractive approach to quickly cover large areas. The ability to efficiently manage drone swarms is therefore of interest to the ISCRAM community as it poses both practical and research challenges. While our findings are based on a relatively small number of novice participants conducting a simulated task, which limits generalizability, it still points to several interesting observations.

The most obvious result is that the type of control mode affects both performance and human factors. First, Condition B (swarm-level control) seems to outperform Condition A (per-drone control) and Condition C (hybrid control). This supports the argument of Saffre et al. (2021) concerning micro-management versus swarm (macro) control approaches. The mental workload measures follow a similar pattern which, together with the improved performance, echo the findings by Chen and Barnes (2012).

Situational awareness (SA) measures indicated that Condition A resulted in lower self-rated SA than the other conditions. This may not be surprising as the task of controlling individual drones could have a negative impact on global SA as focus is on the individual unit rather than the area covered by the collective. This is also supported by the number of user inputs, which peaks in Condition A (Tables 1, 3), indicating a higher user involvement[3]. Additionally, others have noted a negative correlation between SA and MWL in high-workload situations (e.g., Castor, 2009) which—together with the same negative SA–MWL correlation we observed (Table A1)—supports this interpretation.

Although Condition B seems like the most promising control mode, Condition C is nevertheless the most realistic approach: it is still important to design according to the old cybernetic ethical principle of providing many options for conducting a task as this contributes to the potential variety of the system (von Foerster, 2003). From the perspective of "the law of requisite complexity" (Boisot & McKelvey, 2011), providing both per-drone and swarm-level control can be interpreted as increasing the operator's available response repertoire to cope with a messy, disturbance-rich world. This added repertoire also incurs a meta-control complexity cost (Bjurling et al., 2020; cf. Lewis, 2013): operators must detect when the current mode is insufficient and switch appropriately, so design should explicitly support mode switching (e.g., via cues or policies). This meta-control cost is reflected in the increased MWL scores for Condition C compared to Condition B (Table 3). Further, the participants had limited time to learn how to interact with the system, suggesting that the potential of the combined approach never was realised as it may demand experience to know under what circumstances it is necessary or appropriate to switch from swarm-level interaction to individual drone interaction (which may differ across domains and applications). Future research will evaluate these concepts with domain experts (e.g., SAR teams or firefighters) and incorporate higher-fidelity simulation models, including wind, sea currents, and behavioural models of missing persons' movement. Such studies will help determine which mission states—such as smaller search areas or high-certainty "drift" corridors—might prompt a preference for direct/manual control over collective tasking, or require a migration between control types.

Time-to-first-detect was significantly lower in Condition A than in Condition B or Condition C (Table 3). It is possible that more objects were positioned closer to the drones' starting positions in Condition A—and were thus more likely to be detected as drones were dispatched to other locations—than in Condition B or Condition C. A between-subjects experiment with identical object positions across conditions could control for this. Another possibility is that there were two time delays: (1) operator decision time, and (2) configuration/deployment time necessary before drones could begin their search. This corresponds to differences in interaction time and wait time (Cummings & Mitchell, 2008), which may have been shorter in Condition A.

Taking the generally improved performance of Condition B (and partially Condition C)—in terms of objects found, drone utilization, and total area coverage (Tables 1, 3)—compared to Condition A plus the lower time-to-first-detect into account, we hypothesize a "performance ramp-up" for swarm-level control compared to individual drone control: individual drone control is easy to understand, initiate, and control but it does not scale well with increasing drone numbers, while swarm-level control requires more initial mental work to understand (UI design and controls, configuring task area geometry, etc.). Once a swarm-level mission is deployed, its performance over time exceeds that of individual drone control. Moreover, this ramp-up is (at least) two-dimensional: (1) in the

---

[3] We did not observe a relation between objects found (task performance) and user control input counts (Table A1; cf. Walker et al., 2013), although both metrics differed across all control mode conditions.

simulation itself (planning tasks, dispatching drone, etc. as discussed), but also (2) in terms of pre-trial instructions and training (i.e., experience) that affects SA. The observed training effects suggest that subjective ratings and performance were still stabilizing as participants learned the UI concepts, not just the task. This aligns with findings that higher-level automation initially increase SA wait time but reduces delays once operators internalize the control mode (Cummings & Mitchell, 2008). Future evaluations should include training sessions or measure time-to-proficiency to separate UI learning from sustained performance.

Additionally, we note that Condition B outperformed Condition A in subjective human factors and objective performance metrics despite its swarm route planner not being fully tuned to optimize overlap in search passes. Observed performance differences could therefore be conservative. However, the reduced drone utilization in Condition A (compared to Condition B or C) could also contribute to the differences in area coverage overlap.

In conclusion, our study compared three approaches to multi-drone C2 and their effects on human factors and performance outcomes in a simulated maritime SAR mission context. Our results suggest that swarm-level control generally offers lower cognitive demands (MWL, SA) and better task performance compared to individual drone control when drone numbers far exceed typical "fan-out" limits despite requiring more configuration to deploy. Hybrid control regimes also outperform individual drone control, but its task performance is more sensitive to operator expertise and task domain requirements. It also carries an added workload cost compared to purely swarm-level control, but it likely remains the most realistic control approach for future drone swarm applications, including SAR.

## ACKNOWLEDGMENTS

This work was supported by the Swedish Defence Material Administration and NFFP (National Aviation Research Programme), which is funded by VINNOVA (Swedish Governmental Agency for Innovation Systems, 2024-03158), the Swedish Armed Forces, and the Swedish Defence Material Administration. The authors would like to thank: Charlie Simonsson for assisting with data collection; the participants for contributing their time to our research; and the reviewers whose feedback helped to improve the quality of this paper.

## REFERENCES

Arnold, R., Jablonski, J., Abruzzo, B., & Mezzacappa, E. (2020). Heterogeneous UAV Multi-Role Swarming Behaviors for Search and Rescue. *2020 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA)*, 122–128. https://doi.org/10.1109/CogSIMA49017.2020.9215994

Bjurling, O. (2025). *Designing Human-Swarm Interaction Systems* [Doctoral dissertation, Linköping University]. https://doi.org/10.3384/9789180759595

Bjurling, O., Arvola, M., & Ziemke, T. (2021). Swarms, Teams, or Choirs? Metaphors in Multi-UAV Systems Design. In M. Zallio, C. Raymundo Ibañez, & J. H. Hernandez (Eds.), *Advances in Human Factors in Robots, Unmanned Systems and Cybersecurity* (Vol. 268, pp. 10–15). Springer International Publishing. https://doi.org/10.1007/978-3-030-79997-7_2

Bjurling, O., Granlund, R., Alfredson, J., Arvola, M., & Ziemke, T. (2020). Drone Swarms in Forest Firefighting: A Local Development Case Study of Multi-Level Human-Swarm Interaction. *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, 1–7. https://doi.org/10.1145/3419249.3421239

Boisot, M., & McKelvey, B. (2011). Complexity and Organization-Environment Relations: Revisiting Ashby's Law of Requisite Variety. In P. Allen, S. Maguire, & B. McKelvey (Eds.), *The SAGE Handbook of Complexity and Management* (pp. 279–298). SAGE Publications.

Braarud, P. Ø. (2021). Investigating the validity of subjective workload rating (NASA TLX) and subjective situation awareness rating (SART) for cognitively complex human–machine work. *International Journal of Industrial Ergonomics*, *86*, 103233. https://doi.org/10.1016/j.ergon.2021.103233

Castor, M. (2009). *The use of structural equation modeling to describe the effect of operator functional state on air-to-air engagement outcomes* [PhD dissertation, Linköping University]. https://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-17224

Chen, J. Y. C., & Barnes, M. J. (2012). Supervisory control of multiple robots in dynamic tasking environments. *Ergonomics*, *55*(9), 1043–1058. https://doi.org/10.1080/00140139.2012.689013

Cummings, M. L. (2015). Operator Interaction with Centralized Versus Decentralized UAV Architectures. In K. P. Valavanis & G. J. Vachtsevanos (Eds.), *Handbook of Unmanned Aerial Vehicles* (pp. 977–992). Springer. https://doi.org/10.1007/978-90-481-9707-1_117

Cummings, M. L., Bruni, S., Mercier, S., & Mitchell, P. J. (2007). Automation Architecture for Single Operator, Multiple UAV Command and Control. *International C2 Journal*, *1*(2), 1–24.

Cummings, M. L., How, J. P., Whitten, A., & Toupet, O. (2012). The Impact of Human–Automation Collaboration in Decentralized Multiple Unmanned Vehicle Control. *Proceedings of the IEEE*, *100*(3), 660–671. https://doi.org/10.1109/JPROC.2011.2174104

Cummings, M. L., & Mitchell, P. J. (2008). Predicting Controller Capacity in Supervisory Control of Multiple UAVs. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *38*(2), 451–460. https://doi.org/10.1109/TSMCA.2007.914757

Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors*, *37*(1), 32–64. https://doi.org/10.1518/001872095779049543

Giles, K., & Giammarco, K. (2017). Mission-based Architecture for Swarm Composability (MASC). *Procedia Computer Science*, *114*, 57–64. https://doi.org/10.1016/j.procs.2017.09.005

Grönbäck, A.-M., Granberg, T. A., & Matinrad, N. (2025). Evaluating Response Times of UAS as Search and Rescue Resources Dispatched from Commercial Maritime Traffic. *Proceedings of the 22nd International ISCRAM Conference*. Information Systems for Crisis Response and Management (ISCRAM'25). https://doi.org/10.59297/jscvvf86

Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(9), 904–908. https://doi.org/10.1177/154193120605000909

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock & N. Meshkati (Eds.), *Advances in Psychology* (Vol. 52, pp. 139–183). North-Holland. https://doi.org/10.1016/S0166-4115(08)62386-9

Johansson, B., Tordenlid, J., Lundberg, J., & Alfredson, J. (2024). One System to Connect them All—A Core System for Realizing Integrated Command and Control Research. *Proceedings of the 21st International ISCRAM Conference*. Information Systems for Crisis Response and Management (ISCRAM'24). https://doi.org/10.59297/r96sya40

Khan, A., & Amir, M. (2025). Harnessing Drone Swarms for Enhanced Search and Rescue Operations: Efficiency, Resilience, and Future Directions. *Frontiers in Computational Spatial Intelligence*, *3*(3), 141–153.

Kolling, A., Nunnally, S., & Lewis, M. (2012). Towards human control of robot swarms. *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI '12*. The seventh annual ACM/IEEE international conference. https://doi.org/10.1145/2157689.2157704

Kolling, A., Sycara, K., Nunnally, S., & Lewis, M. (2013). Human Swarm Interaction: An Experimental Study of Two Types of Interaction with Foraging Swarms. *Journal of Human-Robot Interaction*, *2*(2), 103–128. https://doi.org/10.5898/JHRI.2.2.Kolling

Lewis, M. (2013). Human Interaction with Multiple Remote Robots. *Reviews of Human Factors and Ergonomics*, *9*(1), 131–174. https://doi.org/10.1177/1557234X13506688

Lomonaco, V., Trotta, A., Ziosi, M., Ávila, J. de D. Y., & Díaz-Rodríguez, N. (2018). *Intelligent Drone Swarm for Search and Rescue Operations at Sea* (arXiv Preprint 1811.05291). https://doi.org/10.48550/arXiv.1811.05291

Miller, C., Funk, H., Wu, P., Goldman, R., Meisner, J., & Chapman, M. (2005). The Playbook™ Approach to Adaptive Automation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *49*(1), 15–19. https://doi.org/10.1177/154193120504900105

Mohd Daud, S. M. S., Mohd Yusof, M. Y. P., Heo, C. C., Khoo, L. S., Chainchel Singh, M. K., Mahmood, M. S., & Nawawi, H. (2022). Applications of drone in disaster management: A scoping review. *Science & Justice*, *62*(1), 30–42. https://doi.org/10.1016/j.scijus.2021.11.002

Raczok, T., Ivens, S.-N., Seidel, L., & Gehringer, S. (2025). Wildfire Detection and Monitoring: A Drone-Based Approach and Comparative Analysis. *Proceedings of the 22nd International ISCRAM Conference*. Information Systems for Crisis Response and Management (ISCRAM'25). https://doi.org/10.59297/2qqdvq51

Restas, A. (2015). Drone Applications for Supporting Disaster Management. *World Journal of Engineering and Technology*, *3*(3C), 316–321. https://doi.org/10.4236/wjet.2015.33C047

Roldán, J. J., Peña-Tapia, E., Martín-Barrio, A., Olivares-Méndez, M. A., Del Cerro, J., & Barrientos, A. (2017). Multi-Robot Interfaces and Operator Situational Awareness: Study of the Impact of Immersion and Prediction. *Sensors*, *17*(8), 1720. https://doi.org/10.3390/s17081720

Saffre, F., Hildmann, H., & Karvonen, H. (2021). The Design Challenges of Drone Swarm Control. In D. Harris

& W.-C. Li (Eds.), *Engineering Psychology and Cognitive Ergonomics* (Vol. 12767, pp. 408–426). Springer International Publishing. https://link.springer.com/10.1007/978-3-030-77932-0_32

Saffre, F., Hildmann, H., Karvonen, H., & Lind, T. (2022a). Monitoring and Cordoning Wildfires with an Autonomous Swarm of Unmanned Aerial Vehicles. *Drones*, *6*(10), 301. https://doi.org/10.3390/drones6100301

Saffre, F., Hildmann, H., Karvonen, H., & Lind, T. (2022b). Self-Swarming for Multi-Robot Systems Deployed for Situational Awareness. In T. Lipping, P. Linna, & N. Narra (Eds.), *New Developments and Environmental Applications of Drones* (pp. 51–72). Springer. https://doi.org/10.1007/978-3-030-77860-6_3

Sende, M., Raffelsberger, C., Hayat, S., Köfler, A., & Almer, A. (2024). Drone Swarm for Post-Wildfire Hot Spot Detection: Technology Assessment and PoC Demonstrator. *Proceedings of the 21st International ISCRAM Conference*. Information Systems for Crisis Response and Management (ISCRAM'24). https://doi.org/10.59297/512raf78

von Foerster, H. (2003). *Understanding Understanding: Essays on Cybernetics and Cognition*. Springer. https://doi.org/10.1007/b97451

Walker, P., Nunnally, S., Lewis, M., Chakraborty, N., & Sycara, K. (2013). Levels of Automation for Human Influence of Robot Swarms. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *57*(1), 429–433. https://doi.org/10.1177/1541931213571093

Wong, C. Y., & Seet, G. (2017). Workload, awareness and automation in multiple-robot supervision. *International Journal of Advanced Robotic Systems*, *14*(3), 1–16. https://doi.org/10.1177/1729881417710463

## APPENDIX

**Table A1. Descriptive Pearson correlations among subjective and performance outcomes (trial-level).**

| | MWL | SA | Obj. Found | First Det. (ln) | TotAre Cov | AreCov Overlap | Drone Util. |
|---|---|---|---|---|---|---|---|
| SA | -0.357, *p*=.002 ** | | | | | | |
| Obj. Found | -0.390, *p*<.001 *** | 0.126, *p*=.293 | | | | | |
| FirstDet. (ln) | -0.082, *p*=.508 | -0.071, *p*=.570 | -0.214, *p*=.082 | | | | |
| TotAreaCov | -0.439, *p*<.001 *** | 0.230, *p*=.052 | 0.692, *p*<.001 *** | -0.091, *p*=.464 | | | |
| AreCovOverlap | 0.227, *p*=.055 | 0.094, *p*=.433 | -0.418, *p*<.001 *** | 0.199, *p*=.106 | -0.315, *p*=.007 ** | | |
| Drone Util. | -0.400, *p*<.001 *** | 0.272, *p*=.021 * | 0.322, *p*=.006 ** | -0.069, *p*=.580 | 0.693, *p*<.001 *** | 0.018, *p*=.878 | |
| UserContInp (ln) | 0.424, *p*<.001 *** | -0.092, *p*=.441 | -0.101, *p*=.400 | -0.202, *p*=.101 | -0.100, *p*=.405 | -0.061, *p*=.612 | -0.160, *p*=.179 |

Note: Values are Pearson's *r* computed across trial-level observations (72 rows; pairwise deletion for missing values). Because observations are repeated within participants, p-values should be interpreted cautiously; the matrix is provided for descriptive purposes. Time-to-first-detection and user-control inputs are ln-transformed. * p < .05, ** p < .01, *** p < .001.