Human Factors in Robots, Drones and Unmanned Systems, Vol. 218, 2026, 39–49
https://doi.org/10.54941/ahfe1007676

Effects of Swarm Size Variability on Operator Workload

William Hunt[1], Aleksandra Landowska[2], Horia A. Maior[2], Sarvapali D. Ramchurn[1], and Mohammad Soorati[1]

[1]University of Southampton, UK
[2]University of Nottingham, UK

ABSTRACT

Real-world deployments of human–swarm teams depend on balancing operator workload to leverage human strengths without inducing overload. A key challenge is that swarm size is often dynamic: robots may join or leave the mission due to failures or redeployment, causing abrupt workload fluctuations. Understanding how such changes affect human workload and performance is critical for robust human–swarm interaction design. This paper investigates how the magnitude and direction of changes in swarm size influence operator workload. Drawing on the concept of workload history, we test three hypotheses: (1) workload remains elevated following decreases in swarm size, (2) small increases are more manageable than large jumps, and (3) sufficiently large changes override these effects by inducing a cognitive reset. We conducted two studies (N = 34) using a monitoring task with simulated drone swarms of varying sizes. By varying the swarm size between episodes, we measured perceived workload relative to swarm size changes. Results show that objective performance is largely unaffected by small changes in swarm size, while subjective workload is sensitive to both change direction and magnitude. Small increases preserve lower workload, whereas small decreases leave workload elevated, indicating workload residue; large changes in either direction attenuate these effects, suggesting a reset response. These findings offer actionable guidance for managing swarm-size transitions to support operator workload in dynamic human–swarm systems.

Keywords: Human-swarm interaction, Workload

INTRODUCTION

Decentralised robot swarms offer robustness and scalability in high-stakes environments; such safety-critical applications typically require human oversight for supervision, ethical judgement, and high-level control (Lyons et al., 2025). Human–Swarm Interaction (HSI) is the discipline of designing interfaces and interaction mechanisms that best utilise the human within the human–swarm team. This often involves balancing operator workload: the human must be engaged without being overloaded. Optimising workload in human–swarm teams therefore requires a combination of psychological insight and interface design, influencing how the swarm is controlled, commanded, and visualised (Kolling et al., 2015).

Workload history is the concept whereby the effects of a recent period of high demand persist after workload is reduced, often impacting performance

Received February 6, 2026; Revised April 7, 2026; Accepted April 21, 2026; Available online July 20, 2026

(Fuenzalida, 2007). These hysteresis effects, or those where the relationship is dependant on the history or direction of the change, have been observed for mental workload (Jansen et al., 2016). Such effects may arise in swarm operations, where the number of agents an operator must monitor can change dynamically. Some tasks benefit from gradual transitions in workload rather than abrupt shifts, although for drone control, varying workload over time has been shown to outperform maintaining a constant level, potentially by regulating limited cognitive resources (Devlin et al., 2020). The relationship between swarm size and cognitive load more generally depends on interaction modality and task structure, with plateaus or inflection points often emerging. Sudden changes that induce shock or surprise are well studied in aerospace contexts, where pilots are trained to manage them (Duchevet et al., 2025); in swarm settings, such events may be more frequent and less predictable due to the difficulty of monitoring many distributed assets simultaneously.

We propose the following hypotheses: **H1** — Following a period of increased cognitive demand (larger swarms), operators' workload remains higher when the swarm size is reduced, compared to episodes at the same swarm size without a prior overload; **H2** — For a given swarm size, operators retain lower workload when that swarm size is reached via gradual increases rather than via a single abrupt increase; **H3** — Larger jumps in swarm size override the directional effects of H1 and H2, producing a cognitive reset that returns workload toward the no-change baseline regardless of change direction.

BACKGROUND

HSI often pursues "scale invariance", whereby an operator can effectively manage a swarm regardless of its size by abstracting away per-agent detail (Meyer et al., 2022). Closely related is the notion of "fan-out", which captures how many robots a human can reasonably supervise at a required level of performance (Humann and Pollard, 2019). Importantly, this limit is not fixed; it varies widely depending on task demands, interaction modality, and swarm behaviour, making it impractical to define a universal swarm size that can be managed (Chandarana et al., 2018). A range of interaction paradigms have been explored to support scalable HSI, most commonly through GUI-based control that allows operators to task individual robots or influence collective behaviour indirectly (Kolling et al., 2013). More abstract representations, such as heatmaps or aggregate visualisations, can reduce workload by suppressing low-level detail, but are not universally preferred; operators often still require access to individual agents for tasks such as fault diagnosis (Divband Soorati et al., 2021).

Central to human–swarm teaming is the management of mental workload. Overload can impair performance when operators are unable to keep up with task demands, while underload may be detrimental to engagement or vigilance (Adams et al., 2023). Mental workload is often assessed using subjective measures such as the NASA Task Load Index (TLX) (Hart and Staveland, 1988), which is widely accepted but necessarily retrospective and unsuitable for real-time adaptation. Consequently, physiological measures such as electroencephalography (Singh, 2025), heart rate variability (Marois et al., 2024), and pupilometry have been explored in HSI (St-Onge et al., 2019). The effect of swarm size on operator workload has been explored to a limited extent, but existing evidence suggests that larger swarms tend to increase perceived difficulty, arousal, and cognitive demand (Kaduk et al., 2023). As swarm size increases, operators are exposed to a greater number of low-level movements and interactions, which can elevate workload even in relatively simple monitoring tasks (Harriott et al., 2014), with similar effects observed when managing multiple subswarms (Kaduk et al., 2024).

Most experimental studies assume a fixed swarm size for the duration of a task. In practice, however, swarm size is often dynamic: robots may fail, be redeployed, or join the mission as demands change, leading to abrupt workload fluctuations (Ramchurn et al., 2016). Scalability is a defining property of swarms, but in HSI this scalability must extend to the human operator as well (Pendleton and Goodrich, 2013). Prior work suggests that performance may improve with increasing swarm size up to a point before declining as operators fatigue over time (Morrow and Zawodniok, 2024). While adaptive approaches have been proposed to monitor or respond to workload changes at runtime (Abioye et al., 2024). The effect of these swarm-size changes on workload is still not well understood.

METHODOLOGY

We design a simple experiment that measures an operator's ability to monitor a swarm for changes. A swarm of N drones flocks across the map; partway through, one drone disappears, and the user must report the vanished drone's colour. This is not intended to model real drone failure (which could often be flagged automatically), rather serving as a measure for the operator's capacity to keep track of many entities within the swarm. We repeat this task across varying values of N, measuring both accuracy and self-reported mental workload. We define an episode $E(N)$ as a 7-second clip of N drones flocking across the map using the boids algorithm (Reynolds, 1987). Drones are divided between four colours (red, blue, green, yellow). At a random time between 2–4 seconds into the episode, one random drone disappears; at the end, the user reports its colour. Figures 1 and 2 shows an example episode. All episodes were recorded in the Human And Robot Interactive Swarm (HARIS) simulator (Soorati et al., 2024). After each episode, participants also report their mental workload on a 1–7 scale.

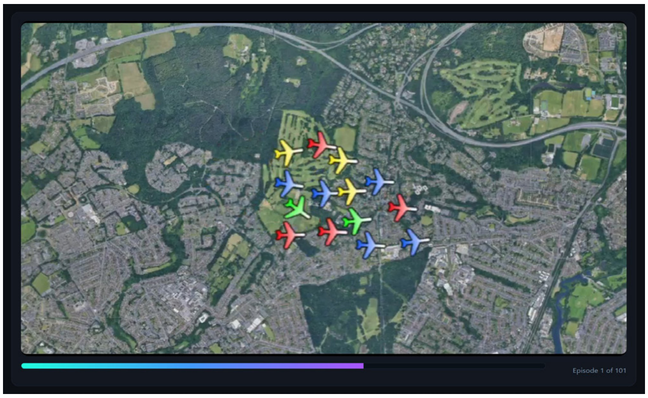

Figure 1: Video shown to the user. Drones flock from left to right, and one disappears.

Also of interest is the effect of the previous episode. In line with hypotheses **H1** and **H2**, we treat accuracy and subjective workload as being reported partly relative to the preceding swarm size. We describe an episode as $E(P, N)$, where P and N are the number of agents in the *previous* and *current* episodes respectively; The change size is $\Delta = N - P$ (Δ may be negative). We assume *first-order dependence*: performance and workload in an episode depend on N and the immediately prior P, but not on earlier episodes. Episodes are grouped into "decks" which are shown sequentially to participants; deck construction differs between studies, as described below.

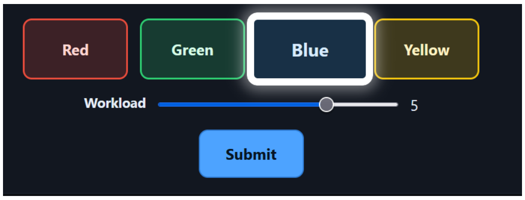

Figure 2: Review panel. After the video ends, the user reports which colour drone disappeared and their workload.

EXPERIMENT DESIGN

We split the experiment into two studies, one testing **H1** and **H2** using small step changes in swarm size, and one testing **H3** using larger jumps to examine potential shock effects. Both studies were conducted on the same platform. Ethics approval was obtained through the University of Southampton ethics committee (ERGO-108904). Participants were recruited via Prolific, with inclusion criteria of age 18–65, residence in the UK or USA, and first-language English; colour-blindness was the only exclusion criterion. Participants completed the task on the Gorilla platform, first passing an Ishihara colour-vision test, then watching a 5-minute instructional video and completing a short practice tutorial. The tutorial also served as an attention check: participants had to score at least 3/7 and provide varied workload reports to proceed. Participants then answered brief demographic questions, completed the experiment, and finished with a post-task survey. Our primary metrics were accuracy in identifying the disappeared drone's colour and self-reported workload after each episode. We additionally collected post-task subjective data to help contextualise the results.

Study 1 – Directional Effects of Small Swarm Size Changes

Study 1 tests workload residue under decreasing swarm size (**H1**) and easing-in under increasing swarm size (**H2**). We used 9 swarm sizes ($N \in \{4, 6, 8, 10, 12, 14, 16, 18, 20\}$) with transitions $\Delta \in \{0, -2, +2\}$, each repeated $r = 4$ times. To balance transitions, we generated episode decks as an Eulerian cycle over nodes N, where edges correspond to episodes $E(a, b)$. Using a modified Hierholzer algorithm[1], each possible transition appears exactly

[1]The original algorithm is deterministic. We choose a random start node and shuffle outgoing edges at each step, producing a different valid cycle per run.

four times, excluding the initial "warm-up" episode. Two complementary decks were generated to reduce positional bias. The post-task survey focused on participants' perceptions of swarm size and size changes. Participants rated how larger swarms affected workload (5- point Likert), then provided open-ended reflections and Likert ratings on whether increases or decreases felt easier or harder. To capture emotional response, particularly surprise, we included a short subset of the PANAS-X questionnaire (Watson and Clark, 1994), consisting of 13 items spanning Attentiveness, Fatigue, Serenity, and Surprise, plus an "other" option, administered for both increase and decrease conditions. Finally, participants reported their task strategy and perceived relationships between swarm size, workload, and accuracy. We recruited 16 participants.

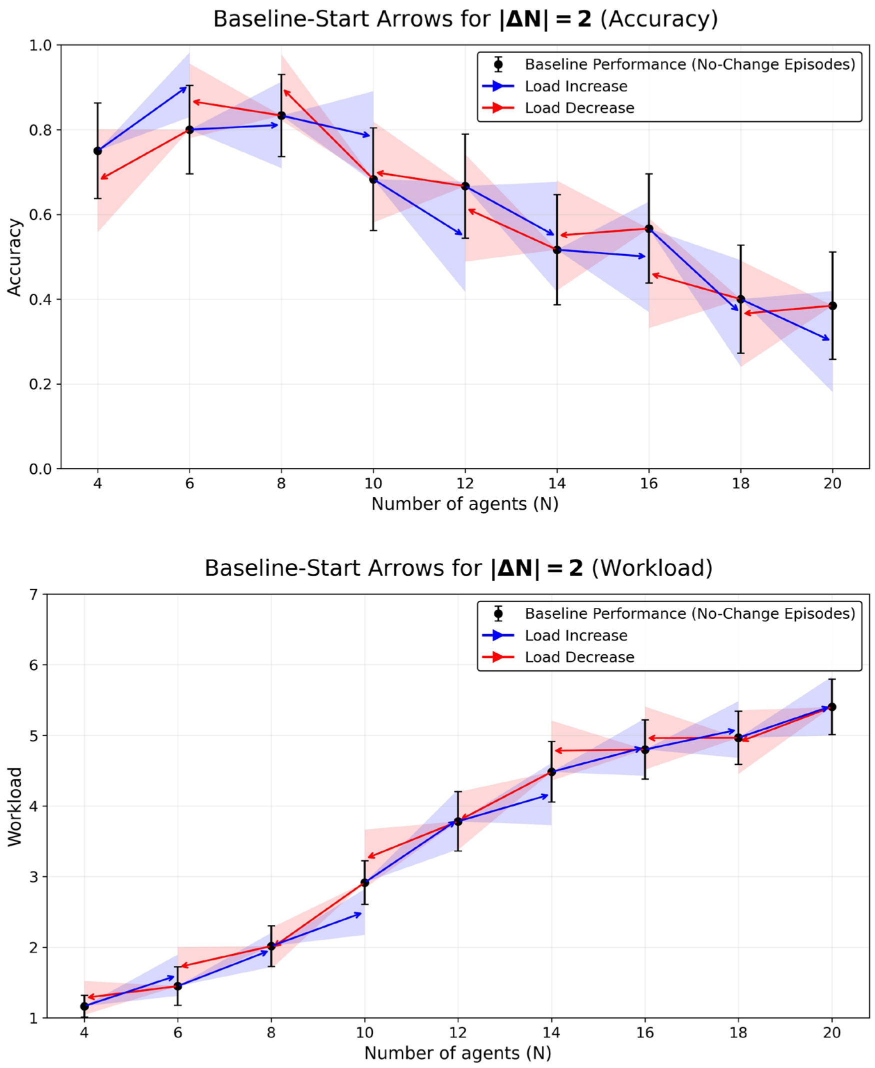

Figure 3: Directional arrows showing changes vs baseline (black dots). E.g. A blue arrow from 8 → 10 agents represents *E*(8, 10). Shaded areas around arrows and bars on baselines denote standard error of the mean.

Study 2 – Directional Effects of Large Swarm Size Changes

Study 2 followed the same structure as Study 1 but targeted larger shifts in swarm size to test **H3**. We balanced colour distributions to maintain engagement throughout each episode, then used 5 swarm sizes ($N \in \{8, 12, 16, 20, 24\}$) with transitions $\Delta \in \{0, \pm 4, \pm 8, \pm 12, \pm 16\}$ over $r = 3$ repetitions. Four complementary decks were generated to reduce positional bias. The post-task survey additionally included the NASA Task Load Index (TLX) (Hart and Staveland, 1988) to capture overall workload beyond per-episode ratings. We recruited 18 participants.

RESULTS

We present the results of each study separately in the context of its focus.

Study 1

The results for small changes in swarm size were modest but promising. Shown in figure 3 is the baseline (all cases where the number of drones did not change), along with arrows showing the accuracy or workload resulting from each transition. Accuracy results are variable, with the direction of change having minimal impact on performance outside. Regarding workload, increases generally resulted in equal or lessened workload, while decreases resulted in equal or higher workload. This pattern is consistent with **H1** and **H2**: on the increase, operators appear to be "eased in", while on the decrease there exists a "workload residue".

Study 2

Figure 4 shows arrow graphs for all adjacent ($\Delta = 4$) changes. No clear pattern is seen for accuracy, however the workload results replicate those in Study 1: increases in the number of drones resulted in equal or lessened workload, while decreases tended to be associated with higher workload. This repetition strengthens the case for both **H2** and to a lesser extent **H1**.

Figure 4 shows arrow graphs for all medium ($\Delta = 8$) changes; these changes are four-times larger than those in Study 1. A clearer pattern emerges at this jump size: increases in swarm size decrease accuracy, and decreases improve accuracy in all cases. Performance effects which were minimal at lower deltas may now be emerging; large increases in swarm size cause a drop in performance relative to baseline and large decreases cause a slight improvement. Although not directly related to our hypotheses, these results may be of some import. Workload differences begin collapsing at this magnitude of change; the deviations from baseline concordant with **H1** and **H2** are gone, supporting **H3** - that a large enough change "shocks" the user, invalidating the other two effects. Figure 4 shows similar results. The differences from baseline, especially in the workload case, get smaller as Δ increases, however the accuracy pattern seems to persist, and the workload differences are completely gone. This further supports that larger changes degrade the effects of **H1** and **H2**, supporting **H3**. Results for $\Delta = 16$ are only a single pair of arrows so are not shown.

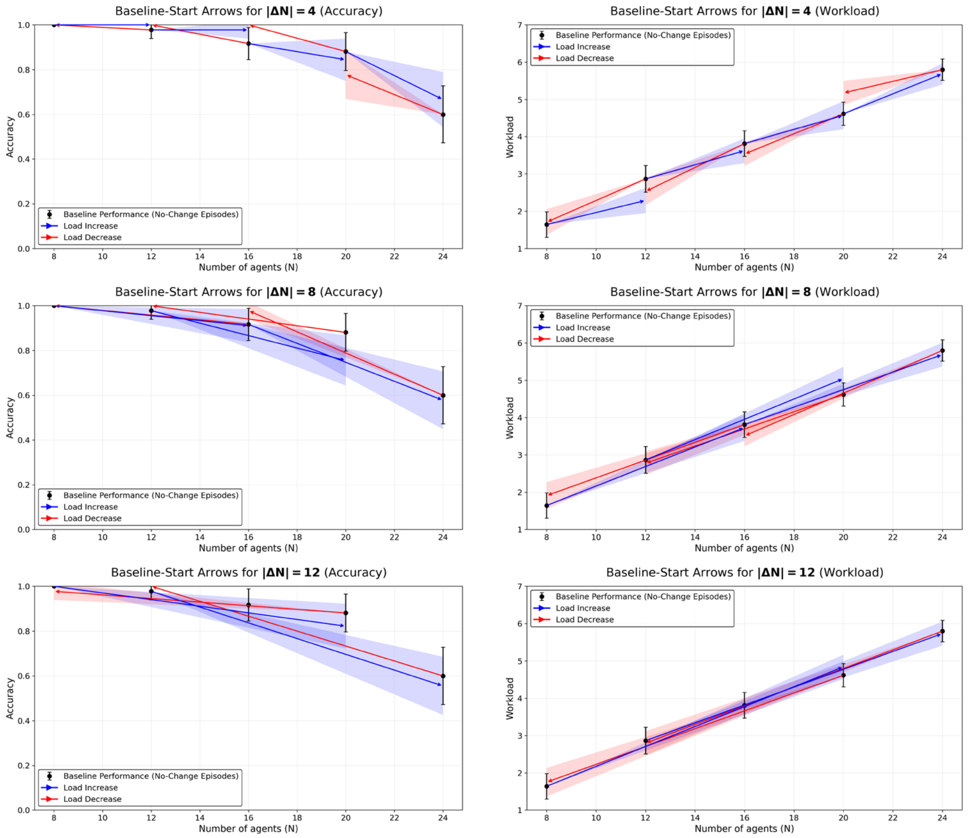

Figure 4: Study 2 directional changes relative to baseline (black dots), grouped by swarm-size change magnitude. For each Δ, accuracy (left) and workload (right) are shown. Shaded areas denote standard error of the mean.

Subjective Responses

A consistent theme across Study 1 (where only small changes were tested) and Study 2 was that reductions in swarm size feel easier yet do not fully return workload to baseline. Several participants described immediate relief but also acknowledged lingering cognitive effort after a decrease. For example, P13 noted that it was "a bit of a relief when the number dropped down" because it gave their brain "a second to relax after trying so hard on the larger number of drones," implying that the prior load persisted despite the decrease. P15 similarly stated that "when the number of drones increased I felt more stressed and panicked; when the number decreased I felt relaxed," yet did not suggest a full reset to baseline. Another participant (P4) described feeling "panicked" by changes and "relieved when I saw less drones," yet said the change "was throwing me off," indicating lingering disruption. These accounts collectively support **H1**: small decreases ease workload but leave a residual cognitive burden. Participants' reflections on small increases were more muted; there were indications that gradual growth in swarm size was manageable, for example P1 explained that they "always prepared for higher numbers" and were "pleasantly surprised" when the increase was smaller, implying that anticipation and gradual changes allowed them to adapt. Despite this, participants were generally less supportive of **H2** overall, seldom mentioning small increases being preferable to larger ones.

Subjective responses most strongly endorsed the shock-severity hypothesis. Respondents frequently described large jumps in swarm size, especially increases, as surprising, disorienting, or panic-inducing. Participant 10 observed that when a trial with fewer drones was followed by one with many more, "it increased my stress and decreased performance." Participants also reported slightly higher surprise for increases (mean ≈ 2.3 on a 1–5 PANAS-X scale) than for decreases (mean ≈ 1.9), indicating that sudden increases do indeed elicit a startle response. Together, these findings suggest that large, abrupt changes produce a qualitative shift in cognitive state, "invalidating the effects of workload history and returning the user closer to baseline, supporting **H3**".

DISCUSSION

Our first hypothesis (**H1**) concerned *workload residue*, the idea that following overload, operators would perform worse or feel more burdened when the swarm size was subsequently reduced. Study 1 showed decreases in swarm size typically to resulted in similar or higher workload, although this started to break down at $\Delta = 4$ and was not at all present for larger deltas. Subjective responses tended to support this idea. Overall, **H1** is fairly well supported at low deltas only.

Our second hypothesis (**H2**) concerned *easing in*; that gradually increasing the swarm size up to N agents would result in lower workload than an abrupt jump to the same size. Study 1 supported the easing-in effect, with increases generally yielding equal or lower workload than baseline. Study 2 replicated this for $\Delta = 4$, but it broke down for larger changes, which again is to be expected. Similarly to **H1**, **H2** is well supported, and our quantitative results suggested that it perhaps extends slightly further than **H1**, being present at $\Delta = 4$, whereas **H1** only held strongly at $\Delta = 2$. This being said, the subjective experience of participants did not strongly support **H2**, although it is possible that our survey simply did not prompt the participants sufficiently well to reveal this.

Our third hypothesis (**H3**) concerned the *shocking effect*. This proposed that larger jumps in swarm size would mentally reset the user, returning their workload closer to the no-change baseline. This effect, if present, would compete with **H1** and **H2**; these two concepts may hold up to some jump size, after which the shocking effect overcomes the two small-change effects. While small jumps showed directional differences consistent with **H1** and **H2**, these differences *collapsed* as deltas increased. Sufficiently large changes in swarm size caused any effects of **H1** and **H2** to be overcome by the magnitude of the shift, which we suspect is explainable through the mental resetting of the user. Subjective responses quite clearly supported startle and surprise effects, and multiple participants explicitly confirmed their own belief in **H3**. Overall, **H3** is supported - there exists a shocking effect which overcomes **H1** and **H2**.

CONCLUSION

Robotic swarms in real-world deployments are likely to be dynamic in size, and a single operator may have rapidly varying numbers of robots to monitor. We hypothesised that managing these size changes may be beneficial to human-swarm operator workload. **H1** - That there exists a workload residue" when the swarm size slowly decreases, appears to suggest that although performance is mostly unaffected, operator workload does remain elevated during these slow reductions. **H2** - That a user could be "eased-in" slowly to a larger swarm size, holds for workload effects, although also has minimal performance impact. Finally **H3** - That larger changes "shock" the operator and invalidate the other two effects, seems to be correct. Our findings are promising, although they focus only on the operator's ability to monitor the swarm; in practice, most monitoring can be automated and instead it is tasking and attending to novel problems that the operator would spend more time on. Future work should explore these more complex human-swarm teaming missions. Regardless, it would behove swarm designers to be aware of changeable swarm sizes in real-world deployments, and consider balancing out increases and decreases where possible to "ease-in" the operators with small increases, and utilise larger changes to "shock" the user, overcoming the "workload residue" effect.

ACKNOWLEDGMENTS

This work was supported by the Trustworthy Autonomous Systems Hub [EP/V00784X/1]. The lead author was also supported by UK Research and Innovation [EP/S024298/1].

REFERENCES

Abioye, A. O., Landowska, A., Hunt, W., Maior, H., Ramchurn, S. D., Naiseh, M., Banks, A., and Soorati, M. D. (2024). Adaptive human-swarm interaction based on workload measurement using functional near-infrared spectroscopy.

Adams, J., Hamell, J., and Walker, P. (2023). Can a single human supervise a swarm of 100 heterogeneous robots? *Field Robotics*.

Chandarana, M., Lewis, M., Sycara, K., and Scherer, S. (2018). Determining effective swarm sizes for multi-job type missions. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 4848–4853.

Devlin, S. P., Moacdieh, N. M., Wickens, C. D., and Riggs, S. L. (2020). Transitions between low and high levels of mental workload can improve multitasking performance. *IISE transactions on occupational ergonomics and human factors*, 8(2):72–87.

Divband Soorati, M., Clark, J., Ghofrani, J., Tarapore, D., and Ramchurn, S. D. (2021). Designing a user-centered interaction interface for human–swarm teaming. *Drones*, 5(4):131.

Duchevet, A., Imbert, J.-P., Garcia, J., Lamirault, B., and Causse, M. (2025). Investigating the independent and combined effects of startle and surprise in a simulated flight task. *Human Factors*, 67(11):1170–1187. PMID: 40373188.

Fuenzalida, E. (2007). Effect of workload history on task performance. *Human factors*, 49:277–91.

Harriott, C. E., Seiffert, A. E., Hayes, S. T., and Adams, J. A. (2014). Biologically-inspired human-swarm interaction metrics. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Volume 58, 1471–1475.

Hart, S. G. and Staveland, L. E. (1988). Development of nasa-tlx (task load index): Results of empirical and theoretical research. In Hancock, P. A. and Meshkati, N., editors, *Human Mental Workload*, 139–183. North-Holland.

Humann, J. and Pollard, K. A. (2019). Human factors in the scalability of multirobot operation: A review and simulation. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 700–707.

Jansen, R. J., Sawyer, B. D., Van Egmond, R., De Ridder, H., and Hancock, P. A. (2016). Hysteresis in mental workload and task performance: The influence of demand transitions and task prioritization. *Human factors*, 58(8):1143–1157.

Kaduk, J., Cavdan, M., Drewing, K., Vatakis, A., and Hamann, H. (2023). Effects of human-swarm interaction on subjective time perception: Swarm size and speed. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, 456–465.

Kaduk, J., Cavdan, M., Drewing, K., Vatakis, A., and Hamann, H. (2024). From one to many: How active robot swarm sizes influence human cognitive processes. In *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*.

Kolling, A., Sycara, K., Nunnally, S., and Lewis, M. (2013). Human swarm interaction: An experimental study of two types of interaction with foraging swarms. *Journal of Human-Robot Interaction*, 2(2):103–129.

Kolling, A., Walker, P., Chakraborty, N., Sycara, K., and Lewis, M. (2015). Human interaction with robot swarms: A survey. *IEEE Transactions on Human-Machine Systems*, 46(1):9–26.

Lyons, J. B., Capiola, A., Adams, J. A., Mator, J. D., Cherry, E., and Barrera, K. (2025). Examining the human-centred challenges of human–swarm interaction. *Philosophical Transactions A*, 383(2289):20240140.

Marois, A., Mouratille, D., Pratviel, Y., Chamberland, C., and Tremblay, S. (2024). Using cardiac and electrodermal activity as cognitive markers for interruptions and distraction in a surveillance simulation. In *Neuroergonomics and Cognitive Engineering (AHFE Conference Proceedings)*.

Meyer, J., Pinosky, A., Trzpit, T., Colgate, E., and Murphey, T. D. (2022). A game benchmark for real-time human-swarm control. In *2022 IEEE 18th International Conference on Automation Science and Engineering (CASE)*, 743–750.

Morrow, J. and Zawodniok, M. (2024). Evaluation of the human-robot-interaction dynamic under mental fatigue constraints in search and rescue operations. In *2024 International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*, 1–7. IEEE.

Pendleton, B. and Goodrich, M. (2013). Scalable human interaction with robotic swarms. In *AIAA Infotech@Aerospace (I@A) Conference*.

Ramchurn, S. D., Huynh, T. D., Wu, F., Ikuno, Y., Flann, J., Moreau, L., Fischer, J. E., Jiang, W., Rodden, T., Simpson, E., et al. (2016). A disaster response system based on human-agent collectives. *Journal of Artificial Intelligence Research*, 57:661–708.

Reynolds, C. W. (1987). Flocks, herds and schools: A distributed behavioral model. In *Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, 25–34.

Singh, S. (2025). Optimizing human-machine interfaces for neuroergonomics: Cognitive workload and performance in suas operations. In *Human-Computer Interaction & Emerging Technologies (AHFE Conference Proceedings)*.

Soorati, M. D., Naiseh, M., Hunt, W., Parnell, K., Clark, J., and Ramchurn, S. D. (2024). Enabling trustworthiness in human-swarm systems through a digital twin. In *Putting AI in the Critical Loop*, 93–125. Elsevier.

St-Onge, D., Kaufmann, M., Panerati, J., Ramtoula, B., Cao, Y., Coffey, E. B., and Beltrame, G. (2019). Planetary exploration with robot teams: Implementing higher autonomy with swarm intelligence. *IEEE Robotics & Automation Magazine*, 27(2):159–168.

Watson, D. and Clark, L. A. (1994). The panas-x: Manual for the positive and negative affect schedule-expanded form.