Hunt, E. R., Cullen, C. B., & Hauert, S. (2021). Value at Risk strategies for robot swarms in hazardous environments. In H. G. Nguyen, P. L. Muench, & B. K. Skibba (Eds.), *Proc. SPIE 11758: Unmanned Systems Technology XXIII* [117580M] (Proceedings of SPIE - The International Society for Optical Engineering; Vol. 11758). Society of Photo-Optical Instrumentation Engineers (SPIE).
<https://doi.org/10.1117/12.2585760>

Peer reviewed version

Link to published version (if available):
[10.1117/12.2585760](https://doi.org/10.1117/12.2585760)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via SPIE at <https://doi.org/10.1117/12.2585760> . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Value at Risk strategies for robot swarms in hazardous environments

Edmund R. Hunt^{a*}, Conor B. Cullen^a, and Sabine Hauert^a

^aDepartment of Engineering Mathematics, University of Bristol, Ada Lovelace Building,
University Walk, Bristol, BS8 1TW

ABSTRACT

A promising application of swarm robotics is environmental monitoring, whereby large numbers of self-organizing agents are deployed concurrently to gather information on a short- or long-term basis. The greater resilience and attritability of a swarm system may be particularly valuable in hazardous or adversarial environments where a proportion of agents are likely to be damaged or destroyed. In such contexts the profitable information gathered by sensor payloads on robots is associated with material risk. Relatively little work has been done to consider how to manage this risk autonomously and effectively at the individual and system level. The field of financial risk management provides pre-existing tools and frameworks to get a head-start on this challenge. The method of Value at Risk (VaR) allows the easy quantification of prospective losses over a defined time period and confidence interval. Here, we consider VaR in a multi-agent context where the environment is intrinsically risky, for example containing damaging radiation sources. In agent-based simulations, individuals calculate VaR in real time and broadcast a self-triggered alert to their neighbors when their VaR limit is breached, helping them to avoid the area. This minimal communication system is effective at decreasing overall swarm exposure to hazards, while permitting agents to make risk-weighted explorations of hazardous areas. We further discuss the opportunity for finance-inspired risk management frameworks to be developed in the multi-agent systems context.

Keywords: Swarm Robotics, Risk Management, Value at Risk, Environmental Monitoring, Hazards

1. INTRODUCTION

Robot swarms have the key distinguishing feature that they rely on ‘bottom-up’ self-organization to coordinate;¹ this is inspired by how biological collectives are organized.² One advantage of using a collective group of robots as opposed to a singular robot is that the loss of one robot to material damage will not render the overall system inoperable. This allows the serious consideration of an ‘attritable’³ or expendable⁴ design, trading off characteristics of reliability and maintainability in favor of lower cost and greater numbers. The potential applications of robot swarms are diverse, but here we consider more dangerous deployment scenarios that are most likely to compromise a proportion of the swarm. These extreme environments are too hazardous for humans due to radiation, high temperatures, chemical spillage etc. Nevertheless, exploration and sensing in such places is very valuable, in relation to the actionable information it can put into users’ hands. Swarms could be utilized in these areas to efficiently explore and/or cover the area, gathering information ranging from the location of survivors in a disaster area to the sources of dangerously high radiation levels (following a nuclear reactor disaster, for example).

Key to making effective use of a swarm’s (partial) expendability will be the assessment of the level of risk and reward present in the environment. If there are high hazard levels with little reward present (e.g. the area has already been recently explored), it makes little sense to put a robot in harm’s way. On the other hand, the risk of sending a robot into a danger zone might be justified by an offsetting prospect of reward. Here, we note the potential of Value at Risk (VaR) methods to make a computationally low-cost estimate of the risk level present in a locality, that should be relevant in the context of relatively simple (expendable) robots. In a swarm context, risk can be managed not only at the level of the individual, but the whole swarm. This paper complements our recent work on how individual risk limits could be adapted online.⁵ Risk-aware decision-making will be a vital step forward for real-world usability, making swarms safe⁶ and effective.

*Correspondence to E.R. Hunt. E-mail: edmund.hunt@bristol.ac.uk

2. METHODS

2.1 Value at Risk

We begin by providing some brief background on Value at Risk (VaR). This metric is widely used in the finance industry to estimate and manage risk.^{7,8} A general mathematical definition of VaR is as follows. X is a profit and loss distribution (loss negative, profit positive). The VaR at level $\alpha \in (0, 1)$ is the smallest number y such that the probability that $Y := -X$ does not exceed y is at least $1 - \alpha$. $\text{VaR}_\alpha(X)$ is the $(1 - \alpha)$ quantile of Y , i.e.,

$$\text{VaR}_\alpha(X) = -\inf\{x \in \mathbb{R} : F_X(x) > \alpha\} = F_Y^{-1}(1 - \alpha)$$

where F_X is the cumulative distribution function of X .⁹ One can use this definition for calculating VaR if it is assumed that X follows some parametric distribution. Otherwise, the so-called Historical Method of calculating VaR simply involves reorganizing the past N profits and losses up to the current time t from smallest to largest, and finding the $(100 - c)^{\text{th}}$ percentile value, where c is the confidence level. For example, if $c = 95\%$, then the Historical VaR is the value q such that the area under the empirical cumulative distribution function from $-\infty$ to q sums to $1 - c$, which in this case is 0.05. In other words, this percentile contains the worst 5% of returns in a rolling window of size N . The value $1 - c$ can also be referred to as the probability of VaR failure, or p_{VF} .⁷ One advantage of using this method is that it makes no assumptions about the underlying distribution of the returns.⁸ VaR provides a reference point for the magnitude of losses that can occur in a given time period. If it is calculated based on one-day returns, for example, it provides an estimate of the biggest loss expected in the forthcoming one-day period, up to a given confidence interval. A key desirable feature of VaR is that it compresses the values of changing market variables into a single quantifiable measure/metric that is intuitive to understand.

Figure 1. Histogram showing the frequency of all daily FTSE100 returns from a ten-year period along with the cumulative frequency and 95% Historical VaR.

One of the best known investment indicators is the FTSE100, a share index of 100 companies listed on the London Stock Exchange with the largest market capitalization. Taking daily closing price data from the last 10 years,¹⁰ the daily returns time series, r_t , can be calculated as the relative difference of prices between days,

$$r_t = \frac{p_t - p_{t-1}}{p_{t-1}}, \quad (1)$$

and here p_t is the closing price for any day $t \in \{01/01/2009, 30/10/2019\}$ and p_{t-1} is the closing price for the day before, excluding the first day in the time series. Fig. 1 shows a histogram of the daily returns over the collected

data. The Historical VaR at the 95% confidence level is indicated as the 5th percentile. The one-day Historical VaR metric at this level is -1.58%, meaning the entirety of the returns to the left of this point are considered the worst 5% of the set of returns, so any new data points falling in this region result in a VaR failure.

In the field of robotics, the relevance of VaR as a risk metric has recently been noted in the area of motion planning and control, in the form of conditional value-at-risk (CVaR).^{11–15} CVaR, also known as expected shortfall (ES), is a more conservative measure than VaR which better accounts for tail risk (rare but dangerous events). The CVaR of a random loss represents the conditional expectation of the loss within the $(1 - \alpha)$ worst-case quantile of the loss distribution, where $\alpha \in (0, 1)$.¹⁶ For a random variable X , CVaR and VaR are related as follows: $\text{CVaR}_\alpha(X) = \mathbb{E}[X|X \geq \text{VaR}_\alpha(X)]$. CVaR has been preferred in recent research^{11–15} because it is noted to be a ‘coherent’ risk measure, in respect of certain formal characteristics,⁹ which are arguably important for risk assessment rationality.¹² Practically speaking, because CVaR gives more conservative risk estimates, here we generally use VaR for our risk metric because we investigate deployment of ‘expendable’ robot swarms where some amount of tail risk is tolerated by the user. See though section 3.3 where we employ CVaR as a metric for risk-averse ‘shy’ agents. Nevertheless, for the foregoing reasons CVaR might be the preferable choice for future research, with behavioral thresholds suitably adjusted.

2.2 Agent-based simulation

This section will explain the basic setup for modeling VaR in a virtual simulation of a single robot with random movement. Dimensions of these simulated ‘agents’ were originally imagined to be similar to those of Kilobots, which are used for conceptual swarm robotics research in the laboratory.¹⁷ They are about the size of a coin, move with slip-stick locomotion via two small vibrating motors, communicate via infrared within a 10 cm radius, have an ambient light sensor, and have a low cost per unit of around \$14 allowing large swarm sizes to be tested. In the absence of external cues (e.g. a beacon robot or guiding light) they essentially perform random walks through space, hence the simple movement models implemented here. Nevertheless, the simulation dimensions are arbitrary and should be relevant also to larger robots: the centimeter length scale could be read as meters. After a simulation has been set up, strategies utilizing VaR will be defined and compared by testing a performance metric in multiple simulation environments.

The baseline simulation includes only one agent moving in a random walk through different environments. The simple algorithms described here were implemented in MATLAB 2019a. For an agent a , the position $\mathbf{x}_a(t) = (x_t, y_t)$ is initialized to the center of a two-dimensional, square grid world of area $L \times L$ for a chosen length L , with $x_t, y_t \in \mathbb{Z}$. We choose a dimension of $L = 100$, imagined initially to be cm (Kilobot length scale) but could also be meters, thus $\mathbf{x}_a(0) = (50, 50)$. The agent navigates the grid world via a random walk process with 8 possible directions. At each time step t , the agent moves by making a random action $\mathbf{a} \in \{N, NE, E, SE, S, SW, W, NW\}$, so that $\text{prob}(N) = \text{prob}(NE) = \dots = \text{prob}(NW) = 1/8$. Boundaries are reflective, so if the agent is at a boundary then it will choose its next move from its adjacent cells within the grid world (either 3 or 5 options). This creates a closed state space where limits act as ‘walls’, bounded by $x_{\max} = y_{\max} = 100$ and $x_{\min} = y_{\min} = 0$. Hüttenrauch et al.¹⁸ consider a periodic toroidal state space in which agents that go over boundaries appear on the opposite side of the grid world, but this is not used here. A small subsection of the grid world is shown in Fig. 2.

We consider a scenario where a robot swarm is deployed in a risky (hazardous) environment to gather information. To bridge the gap between financial risk management and robot swarm behavior, it is necessary to consider what corresponds to ‘profit’ and ‘loss’ and whether and how they can be put on the same scale. In the simple agent-based model we develop here, loss corresponds to exposure to damaging radiation and profit to new information (c.f. ‘food’ in foraging models). While radiation exposure can be objectively measured as a physical quantity, the value of information is inherently subjective and depends on the user. Consider for example that 100 arbitrary units of radiation could be ‘just’ tolerated in exchange for 1 unit of ‘information’, and thus a scaling factor of 0.01 could be used to rescale the ‘loss’ onto the same profit and loss (P&L) scale. In this case, a -1 loss would be balanced by a $+1$ gain, to result in a ‘break even’ where the user is indifferent to the outcome, because the damage to the robots is not outweighed by what has been learned. In another example, swarm that can gather $+2$ information in a time period at the expense of 0.01×-200 exposure would

Figure 2. A 10×10 subsection of the grid world, showing the single agent in red initially in the green cell (50, 50) at $t = 0$. The directions of all possible moves are shown for $t = 1$. Each grid cell has area 1 cm^2 , inspired from arenas in robotics applications.¹ Actual positions are in the midpoints of the axes labels.

achieve a ‘profit’ of +1. Putting profit and loss on the same scale in such a way therefore depends on the user establishing what is just tolerable.

The grid world environment is simulated to mimic the real world scenario of a room filled with radioactive materials. The space is populated by positive ‘rewards’ in each cell, with some cells more valuable than others. Agents are very minimal: they have no prior knowledge of the world, do not localize their position, and do not store a map as they move. It is assumed that the agent has a sensor that measures the information collected (i.e. reward) and the damage received in one cell. Information becomes zero once the agent has collected the information so that repeated visits to one cell do not permit a reward. At the same moment as gathering the reward associated with a grid cell, the agent simultaneously experiences the overlaid ‘damage’ grid to simulate the negative effect the radiation will have on the agent. If the accumulated damage exceeds a specified threshold, then the agent is ‘dead’ and stops moving. Reward and damage matrices ($\mathbf{R} \in \mathbb{R}^{100 \times 100}$ and $\mathbf{D} \in \mathbb{R}^{100 \times 100}$ respectively) have the same dimensions as the grid world and are generated using the probability density function (PDF) of a gamma distribution. This is chosen because of its shape and scale parameters, which can be adjusted to simulate a low amount of regions with density towards the higher end of the distribution, representing occasional areas of high radiation/information. The elements of $\mathbf{D}$ are chosen to be equal for each simulation by using the gamma PDF (with shape parameter and scale parameter both fixed at 1) to generate a ‘random noise’ environment (Fig. 3(a)). The elements of $\mathbf{D}$ can also be altered for the purpose of testing how much exploration can occur in different extreme scenarios. One of these is the ‘correlated’ environment (Fig. 3(b)), where elements of $\mathbf{D}$ in the right tail of the gamma PDF ($d \in \mathbf{D}$ where $d > 6$, i.e. $P(d > 6) = 0.14$, making up 14% of the environment) are expanded into circular regions of radius 5 cm with the same damage in all points of the circle. This effectively creates larger hazardous zones/environmental fluctuations in which the agent will gain severe damage when entering. The same method is used to generate elements of $\mathbf{R}$ so there are also areas of valuable information that occur sparsely in high density (Fig. 3(d)). The damage map is even more extreme/volatile in a ‘variable’ environment (Fig. 3(c)), where the radius is randomly assigned between 5 and 10 cm for hazardous regions.

2.3 Avoidance of Risk Using Adaptive Thresholds

One simple way for the agent to avoid severe damage is to increase speed when it senses a high level of radiation. This could be done relative to the immediate reward in a *Return Damage Avoidance* strategy. This can be done by setting a threshold r_{Thr} , such that the step size increases to 2 if the value of the agent’s return at time t is

Figure 3. Simulated environments in which the agent navigates. The grid world consists of one of the three damage maps (a, b or c) overlapping with the correlated reward map in d. The ‘correlated’ damage environment (b) is so-called because damage is ‘patchy’, while the ‘variable’ environment (c) is so-called because the patches of damage have variable size.

less than r_{Thr} ($r_t < r_{\text{Thr}}$). The reward at each time step is calculated as $\kappa_t \in \mathbf{R}$ and the damage is $d_t \in \mathbf{D}$, each within a rolling time window of size N . The return used for VaR calculations, r_t , is thus defined as

$$r_t = \kappa_t - d_t, \quad (2)$$

relative to r_{t-1} , hence a profit would correspond to more rewards collected than damage and a loss would correspond to more damage taken than rewards. If the randomly chosen direction was E , for instance, the agent would move two cells in a time step in that direction instead of one, given that $r_t < r_{\text{Thr}}$. This behavior ensures that the agent escapes an area of high damage or avoids staying there for too long, as a relatively low return would correspond to a loss. Neither the reward nor damage is received in the intermediate step during a movement of two steps so that increasing velocity would not result in more rewards.

Instead of a behavioral threshold based on the instantaneous return, the agent could use VaR to possibly improve the behavior of speed doubling in a *VaR Damage Avoidance* strategy. The VaR at time t , VaR_t , in the specified time horizon, N , is calculated using the historical method. Now a threshold VaR_{Thr} is set such that the step size increases to 2 if the value of the agent’s VaR at time t is more than VaR_{Thr} ($\text{VaR}_t > \text{VaR}_{\text{Thr}}$). This strategy is expected to be less reflexively reactive to entering high damage areas, because it is backward looking in terms of the input data window. Depending on the circumstances, this could be an asset or a drawback of

the approach. For example, if the damage increase is only transitory, it may not be appropriate to speed up and away from an area with high information rewards.

Both the threshold values r_{Thr} and VaR_{Thr} are adapted from the agent’s initial experience of the environment. Conceptually, this could be compared to the phenomenon of ‘phenotypic plasticity’ in biology, where animals’ ecological conditions during development shape their behavioral traits.¹⁹ For a time window of N , the threshold value is set to the 5th percentile of returns from $t = 0$ to $t = N$, as this is the same significance level VaR is calculated at ($\alpha=0.05$). This adaptive behavior gives the agent a chance to explore its environment in a random walk with constant step size, and the threshold will be set based on how much risk it associates with the environment, as this quantile represents the lowest 5% of returns. VaR_{Thr} is set to $5 \cdot |r_{\text{Thr}}|$ because VaR is on a positive scale as opposed to the negative scale of returns. An example of the thresholds set in the same simulation are shown in Fig. 4.

(a) Time series of the returns recorded by the agent (r_t) and the calculated adaptive threshold r_{Thr} .

(b) Time series of the Historical VaR (VaR_t) and the calculated adaptive threshold VaR_{Thr} .

Figure 4. Example simulation from $t = 0$ to $t = 150$ with a rolling time window of $N = 50$. The Historical VaR at the 95% confidence level is used for the calculation of thresholds.

A more risky environment will have a more positive VaR threshold, VaR_{Thr} , and so any area that has more risk associated with it (i.e. more volatile) than the initial explored area will result in a step size increase. In Fig. 4(a), the first 50 time steps shows little volatility in returns and the threshold is set relatively high. However, after this period there is more uncertainty (i.e. more variation in returns), but the returns remain above the threshold for some of this time so the agent does not increase its step size. The VaR plot in Fig. 4(b) shows that the agent’s VaR exceeds the threshold from $t = 50$ to $t = 95$, so is likely to move faster through any highly damaging areas. This may or may not be an advantage of using the *VaR Damage Avoidance* strategy as opposed to the *Return Damage Avoidance* strategy. For brevity, we focus here on the *VaR Damage Avoidance* strategy.

2.4 Modeling Communication

It is important to acknowledge that most swarm robot systems are built with limited communication range and individual memory constraints.¹ This problem is considered by Cheng & Dasgupta²⁰ where swarm robots (agents) are modeled with these limitations as well as susceptibility to sensor noise and are subject to permanent failures. The model so far has already addressed the latter issue by implementing the maximum damage capacity ($d_{\text{max}} = 250$ or 1000 in the single agent simulations).

The objective of the multi-agent system implementation is to maximize exploration of valuable areas in the environment before all agents fail due to radiation exposure. The rewards gained must be kept track of to reveal the location of these areas to the system controller/user, characterized by the reward map in Fig. 3(d). It is

possible for agents to move to the same position in the same time step, but only one of them receives the reward with equal probability between them. The set of parameters in the updated model are defined as:

- $\mathcal{A}$, the set of all agents $\{1, \dots, P\}$, where P is the population size so that $|\mathcal{A}| = P$
- $\mathbf{m}_{\text{com}}^a(t)$, the communication array of agent $a \in \mathcal{A}$ at time t containing neighboring agent positions
- m_{com} , the radius for the communication array of all agents (i.e. communication range)
- $\mathbf{x}_a(t) = (x_t^a, y_t^a)$, the position of agent $a \in \mathcal{A}$ at time t

The communication and coverage regions are local to each agent and updated at each time step. The communication array $\mathbf{m}_{\text{com}}^a(t)$ for agent a contains the positions of agents within a radius of m_{com} . That is, $\mathbf{m}_{\text{com}}^a(t) = \mathbf{x}_b(t)$ where $b \in \mathcal{A} - \{a\}$. Agents may not be capable of recording this information in the form of memory, so it is assumed that the user has access to this information as soon as the agents transmit it. That is, local information gain is separate to the mapping of the environment, which is made visible to the system user by cells in which rewards have been collected.

To measure the effect of using different assumptions of communication capabilities, there are two different observation scenarios for agents, as used by Hüttenrauch et al.¹⁸ Formally, observability of a system is defined as how much knowledge of the internal states of the agents can be inferred from knowledge of the external outputs.²¹ In this case, each agent observes spatial positions of other agents which may be subject to constraints. One scenario is the global observability case, in which there are no communication constraints because all agents are considered to be in range of each other and are therefore neighbors, so

$$\mathcal{N}_{\mathcal{G}}(i) = \{j \in \{1, \dots, P\} | i \neq j\} \quad (3)$$

represents the neighborhood set of agent $i \in \mathcal{A}$ in a graph $\mathcal{G} = (V, E)$, where V is the set of nodes representing the agents, $V = \{v_1, \dots, v_P\}$, and $E \subset V \times V$ are the edges which represent ties between agents ($\{v_i, v_j\}$ indicates that agents i and j are neighbors).¹⁸ If there are communication constraints (i.e. considering communication arrays $\mathbf{m}_{\text{com}}^a(t)$), an edge in the network only connects to an agent if its position is within the communication array of a different agent or agents. Thus, the neighborhood set of agent $i \in \mathcal{A}$ at time t becomes

$$\mathcal{N}_{\mathcal{G}}(i) = \{j \in \{1, \dots, P\} | i \neq j, \mathbf{x}_j(t) \in \mathbf{m}_{\text{com}}^i(t)\}, \quad (4)$$

such that the distance between agents i and j , $\sqrt{(x_t^i - x_t^j)^2 + (y_t^i - y_t^j)^2}$, is less than or equal to m_{com} . $m_{\text{com}} = 10$ is chosen after insight into Kilobot constraints (10 cm sensor range)¹⁷ and the chosen arena scale. To permit the coordination of neighbors to reduce risk, if a connection is made and any of the agents exceed their individual return threshold, then all other agents in proximity will double their step size for as long as this condition is satisfied; when out of range, the agent's velocity returns to normal. Instead of one singular threshold used in the one agent case, each agent has its own threshold characteristic of the VaR of its return time series. The motivation behind this is to attempt to make neighboring agents escape risky areas more quickly if one agent has already experienced a breach in their return threshold. This is an example of self-triggered communication which can be effective at significantly reducing communication bandwidth.²² An example of the strategy with local observability in simulation is shown in Fig. 5. To assess the usefulness of this sensor network (broadcasting included), performance measurements will also be taken where agents do not exchange this information (no broadcasting). The latter is a network with no broadcasting so each node (agent) only makes a decision in regards to its own behavior. In summary, the strategies that are proposed as extensions to the *VaR Damage Avoidance* strategy are:

1. *Local Observability (No Broadcasting)*: agents set unique static VaR thresholds at $t = 50$ for the remainder of the simulation, which only the individual has access to.
2. *Local Observability (Broadcasting)*: agents set unique static VaR thresholds at $t = 50$ for the remainder of the simulation, but an agent will broadcast the command to increase speed to neighboring agents if a threshold breach occurs for the agent in question (see Fig. 5).

3. *Swarm Level (Global Observability)*: all agents have access to the returns calculated by other agents, so a static shared ‘global’ VaR threshold is initialized at $t = 50$ for the remainder of the simulation. Individual agents increase their step size if their individual return series breaches this threshold (see Fig. 6).

(a) Agents have proximity ‘discs’ of radius 10. Those highlighted in red have an increased step size. The agents within range have overlapping disks, so one of them has exceeded their individual VaR limit, r_{Thr} , which is then broadcast to the other agent, increasing its step size as a result.

(b) Time series plot of the returns of each agent in a population of 4 agents. The four return series each have an individual VaR limit, r_{Thr} , shown by the bold horizontal line in its unique corresponding colour.

Figure 5. Local Observability strategy in which agents have a radius of communication of 10cm and can only share information when in range.

2.5 Portfolio Theory and the Global VaR Metric

One risk management method that can be used for multiple agents is a global VaR shared between agents when setting the fixed adaptive threshold after the initial learning period. This assumes that all agents have global observability, as per Eq. 3. Portfolio theory considers the risk across multiple assets or investments in a single portfolio. The agents can be modeled as singular investments, and the VaR measurement can be calculated using the Portfolio VaR, which considers the correlation between assets, thereby considering every agent’s initial experience of the environment. Correlated assets tend to move in the same way at the same time and have more overall risk than uncorrelated ones. Sometimes, this may give misleading information, but this is a good initial approximation of risk calculated for the entire swarm. First, the Portfolio VaR requires the calculation of the global (portfolio) rate of return,⁷

$$r_{\text{glob},t} = \sum_{i=1}^P w_i r_{i,t}, \quad (5)$$

for P agents (i.e. assets) with returns $r_{i,t}$ ($i = \{1, \dots, P\}$) at time t given by Eq. 2. This particular method assumes that the payoff (i.e. profit/loss) of assets is a linear function of normally distributed risk.⁷ The weight vector $\mathbf{w} = \{w_1, \dots, w_P\}$ accounts for the relative sizes of investments and defines the linear exposure to a certain risk factor (e.g. market or operational risk⁷), but for simplicity it is assumed that the weights are equal (i.e. $w_i = 1/P \ \forall i$). The formula for the Portfolio/Global VaR⁷ at time t is

$$\text{VaR}_{\text{glob},t} = -z_{1-\alpha} \sqrt{\sum_{i=1}^P \sum_{j=1}^P \Omega_{ij} \mathbf{V}_{ij} r_{i,t} r_{j,t}}, \quad (6)$$

where $z_{1-\alpha}$ is the critical z -value at the α confidence level,

$$\mathbf{\Omega} = \begin{pmatrix} 1 & \rho_{12} & \dots & \rho_{1P} \\ \rho_{21} & 1 & \dots & \rho_{2P} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{P1} & \rho_{P2} & \dots & 1 \end{pmatrix} \quad (7)$$

is the correlation matrix between assets, where $\rho_{ij} = \sigma_{ij}/\sigma_i\sigma_j$, and

$$\mathbf{V} = \begin{pmatrix} \sigma_1 & \sigma_{12} & \dots & \sigma_{1P} \\ \sigma_{21} & \sigma_2 & \dots & \sigma_{2P} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{P1} & \sigma_{P2} & \dots & \sigma_P \end{pmatrix}$$

is the volatility matrix, containing the standard deviations of agent returns over the initial time window. Like the Variance-Covariance VaR model,⁷ this is a parametric model so makes the assumption that the multivariate return distribution is normal. In some cases, the covariance of assets is used instead of the correlation, but correlation is a more convenient and clearer measure of linear dependence, and is able to calculate the strength of the information gain relationship. The method is also analytical because the individual VaR metrics are derived from closed-form solutions.⁷

(a) Agents have no communication arrays but instead only increase their step size if their individual return state falls below the global VaR limit, VaR_{glob} (Eq. 6). Only the agent highlighted in red increases speed here as a result.

(b) Time series plot of the returns of each agent in a population of 4 agents. The global VaR, VaR_{glob} , is shown as the red line, which is low owing to diversification in returns among the agents.

Figure 6. Swarm Level strategy in which agents are assumed to have global observability and a global threshold.

The Portfolio VaR is lowered by what is known as diversification of risk by spreading the risk over multiple assets. This relies on there being a correlation less than 1 between assets. Hence, the overall risk decreases as more assets are added to the portfolio, but this effect becomes less prominent the more assets there are. In the context of the swarm, the risk considered may be lower if the returns added over all agents outweighs the total exposure to hazards. Lower correlation coefficients in Eq. 7 also lowers the risk. If the assets are perfectly correlated (i.e. $\rho_{ij} = 1$ for every agent $i, j \in \mathcal{A}$), then VaR_{glob} is equal to the sum of the individual VaR metrics of agents. However, this is unlikely because correlations of assets are typically imperfect.⁷ An example of the strategy with global observability and a global VaR is shown in Fig. 6.

2.6 Performance Measurements

The performance metric is now adjusted to create a more conservative method of measuring the performance; one that takes into consideration the level of risk taken by the agents in order to get a reward, which is important for the comparison of different swarm population sizes. An important element of Portfolio Theory is the Sharpe ratio,²³ which is useful for measuring the trade-off between risk and return on investments. The Sharpe ratio, S_X , for an investment portfolio X is defined as

$$S_X = \frac{r_X - R_f}{\sigma_X}, \quad (8)$$

where r_X is the average rate of return of X (average of Eq. 5), R_f is the best available rate of return of a risk-free asset (termed as the theoretical risk-free/riskless rate) and σ_X is the standard deviation of X , used as a proxy for risk.²³ A high Sharpe ratio corresponds to a large risk-weighted return on investment (1 or higher is desirable) whereas a low Sharpe ratio close to 0 corresponds to a low risk-weighted return. The risk-weighted return is an indicator of how much risk is involved in gaining the return. For example, one might get a high return with a high level of risk, but this is an undesirable strategy to a rational agent and should be compared fairly to alternative strategies that have lower risks. The numerator of Eq. 8 measures the deviation from the risk-free rate, or the rate of return of the investment with no risk of loss. In the context of swarms, this could be the hypothetical return gained if robots only explored areas that are known to be free of radiation.

Every possible combination of assets can be plotted on a risk-return space in a particular time step to identify the portfolio with the maximum Sharpe ratio. An example of this risk-return space is shown in Fig. 7 for the portfolio composed in the last time step in Fig. 6(b), with adaptation from²⁴ and using functions from.²⁵ The mean portfolio returns are in the range $\{-0.5, 0.5\}$ and the risk proxy is in the range $\{0, 1\}$.

(a) Risk-return space annotated with key components. Portfolios with less standard deviation (risk) result in negative returns, hence some risk is needed to make a profit (i.e. information gain). (b) The Efficient Frontier showing the tangential intersection point which maximizes the Sharpe ratio. This particular combination of assets is known as the Efficient Portfolio.

Figure 7. Plot of the possible asset combinations from the simulation in Fig. 6 in a risk-return space.

The Efficient Frontier is the curve that defines the feasible region of optimal portfolios that have the highest possible expected return for a given level of risk (i.e. standard deviation). In other words, it is the intersection

between the set of portfolios with minimum risk and the set of portfolios with maximum return, known as the Markowitz efficient set.²⁴ This set assumes a positive correlation between risk and return. Points below the curve are suboptimal, meaning investors would prefer not to invest in anything below the Efficient Frontier. Portfolios cannot be constructed above the curve also. This means that the multi-agent system can return points on this curve, for which the constructed portfolios have the highest possible return, so the Efficient Frontier effectively creates a feasible region for an optimization task. To create the Efficient Frontier in Fig. 7, the procedure known as the ϵ -constraint method is used,²⁶ which involves the following steps:²⁷

- Compute a sample of the Efficient Frontier for a sufficiently large number of returns (this is 50 in each time step from the chosen window) ranging from the minimum to maximum possible return of an efficient portfolio.
- Use linear interpolation on the risk-return pairs to obtain a continuous curve representing the Efficient Frontier. It is assumed that 20 efficient portfolios is a sufficient amount to make the curve as smooth as possible.

The Capital Allocation line represents the risk-return trade-off and is tangent to the Efficient Frontier. The gradient is also a measure of the reward-to-variability measure. The intercept is the risk-free rate (R_f in Eq. 8), which is assumed to be 0, and the line's endpoint has a y -value of 1.

The algorithm that maximizes the Sharpe ratio is built into MATLAB's Financial Toolbox.²⁵ The optimization problem²⁷ is defined as

$$\text{Maximize } \frac{\boldsymbol{\mu}^T \mathbf{w} - R_f}{\sqrt{\mathbf{w}^T \boldsymbol{\Omega} \mathbf{w}}}, \quad \text{s.t. } \sum_{i=1}^P w_i = 1, \quad 0 \leq w_i \leq 1,$$

where $\boldsymbol{\mu}$ contains the means of each agent's returns, $\boldsymbol{\mu} = \{\mu_1, \dots, \mu_P\}$. The Markowitz model²⁴ formulates this as a continuous variable problem which instead has the simplified objective

$$\text{Minimize } \mathbf{w}^T \boldsymbol{\Omega} \mathbf{w}, \quad \text{s.t. } \sum_{i=1}^P w_i = 1, \quad 0 \leq w_i \leq 1 \quad \text{and} \quad \boldsymbol{\mu}^T \mathbf{w} \geq r_{\min},$$

where $r_{\min}$ is the minimum expected return. Given the value of the maximized Sharpe ratio at each time step, $S_{X_t}^{\max}$, as a multiplicative coefficient, the discounted reward $\tilde{\kappa}_t^a$ is then

$$\tilde{\kappa}_t^a = \kappa_t^a \cdot S_{X_t}^{\max} \quad (9)$$

for each agent $a \in \mathcal{A}$. Graphically, this is the point in Fig. 7(b) where the Efficient Frontier is tangent to the Capital Allocation line, hence is equivalent to the best possible risk-return trade-off in the feasible region of portfolio space. Due to the risk being calculated after the $N = 50$ time window, it is assumed that $S_{X_t}^{\max} = 1$ for $t \in \{0, 50\}$ to assign no discount to the reward. The Sharpe ratio that maximizes the risk-return trade-off is used because it gives the ratio for the theoretical optimal portfolio that could be achieved at time t . This is more useful than just using the Sharpe ratio of the current asset combination as it provides the best-case scenario given the current risk possibilities. The reward is discounted in light of the varying portfolio volatility, so only the efficient portfolio is considered. Hence, if the agents cannot achieve a very high trade-off, this will result in a lower $S_{X_t}^{\max}$ and incur a penalty on the reward gained, weighted by the magnitude of the risk-return trade-off. The average Sharpe ratio over the 30 simulations is also a good measure of the group efficiency, as it estimates the extent to which the agents are penalized.

Put simply, the swarm looks at the returns over a 50 time step window and uses expected values of the returns and volatility (r_X and σ_X in Eq. 8) to estimate the best combination of the P agent returns. These make statements about the future, meaning by assumption the use of the past returns has an implicit effect on future performance; the discounted reward is correlated with the Sharpe ratio so has a small influence on subsequent behavior. From the point of view of the swarm, the objective is to achieve the highest combination of rewards

whilst minimizing the potential to receive damage in the near future. An optimal portfolio would look similar to Fig. 8, with very low damage values but high rewards. Given the returns and risk generated, this would produce a very high Sharpe ratio, as the total loss is almost negligible. However, asset combinations such as this are very unlikely to be realized. After some time, the returns that agents can achieve are no longer positive if most of the environment has been explored, so this would result in a negative Sharpe ratio. It is assumed that after the agents are dead, the Sharpe ratio is no longer calculated, so the average is only taken over the time period up until this point.

(a) Environment Map with individual agent rewards labeled.

(b) Environment Map with the individual damage received labeled.

Figure 8. Theoretical example depicting the positions of agents (blue dots) that would produce a near-optimal portfolio. The risk proxy is $\sigma_X = 0.68$ and the expected rate of return is $r_X = 2.08$.

3. RESULTS

3.1 Homogeneous Strategy Tests

A homogeneous strategy test (i.e. a population of agents using the same VaR metric and broadcasting behavior) is used to measure the total discounted rewards and group efficiency that the multi-agent system acquires in the various environments. For a starting point before testing larger swarms, 4 agents are utilized, as in Fig. 6 and Fig. 5. This is the best start to comparing performance in the different environments, as the results may differ significantly from the one agent case due to the inclusion of the discounted reward system (Eq. 9). Also, the Local Observability (Obs.) and Swarm Level cases are extensions of the Damage Avoidance strategy, so act as a reference point to the baseline.

Firstly, a comparison of performance (Broadcasting vs. No Broadcasting) for Local Observability is assessed in Fig. 9(b). This shows that for the random noise environment, broadcasting threshold breaches gives better performance, as a pairwise test gives a p -value $= 1.81 \cdot 10^{-5}$. This is because agents are able to speed up within sufficiently small areas and thus cause neighbors in proximity to also avoid these areas. However, there is no significant improvement using this method in the other environment types, so using either type of network would be suitable. Moving forward, broadcasting is used to examine connections formed between agents.

The results from the homogeneous group tests reveal similar patterns of performance in the ‘correlated environment’ and ‘variable environment’ (Fig. 3. The Local Observability and Swarm Level VaR strategies both perform better than the Random Walk baseline, as shown by Fig. 10(a) and Fig. 10(b). The p -value for the test between Global (Swarm Level) and Local Observability is 0.026, so Swarm Level is significantly better. More tests must be carried out on heterogeneous groups to better evaluate the performance in these more extreme environments, as a strategy might perform better or worse based on the number of agents using it. Specifically,

Strategy	Mean & SE Rewards in Random Noise Environment	Mean & SE Rewards in Correlated Environment	Mean & SE Rewards in Variable Environment
Random Walk Baseline	1518±298.9 (0.901)	1372±294.7 (0.891)	1113±305.8 (0.875)
Local Obs. (No Broadcasting)	1756±350.1 (0.898)	1751±275.3 (0.898)	1456±313.4 (0.895)
Local Obs. (Broadcasting)	2334±525.0 (0.919)	1838±387.0 (0.919)	1488±365.4 (0.893)
Swarm Level	2338 ± 504.6 (0.927)	2200 ± 337.3 (0.907)	1712 ± 391.3 (0.903)

Table 1. Homogeneous Group Tests: measures the average group performance and efficiency (average Sharpe ratio, displayed in parentheses) of a homogeneous group ($P = 4$) in which all agents run the same strategy. The results of the best performing strategy in each environment are emboldened.

(a) Swarm networks composed of 4 agents, showing a broadcasting agent which has breached its threshold. Either this is acknowledged by the agent within proximity (left) or ignored (right). Arrows indicate the direction of agents, with magnitude proportional to step size.

(b) Box plot comparing performance when using individual thresholds (i.e. Local Observability) with or without broadcasting of threshold breaches. The damage map used is the random noise environment.

Figure 9. Comparison of the structure and performance of the different networks used in Local Observability.

these are carried out on the variable environment which needs further investigation, as the results are similar to that of the correlated environment but contain more outliers.

The reason that using a global VaR thresholds results in better performance than local thresholds is that the agents have access to the risk calculated over the first 50 time steps by all of the agents, as opposed to individual risk metrics. This means that the adaptive behavior takes the correlation of assets (from Eq. 7) into account and therefore agents have a good view of the ‘benchmark’ for the entire partially explored environment. This is unlike local thresholds in which agents are only adaptive to their initial exploration, so new areas that may have already been explored by others are not factored into their decision making. However, this comes at the expense of allowing global communication, which is generally not assumed for swarm systems.¹ As a compromise, a global VaR limit could be updated based on sporadic self-triggered communications back to the ‘base’ (user). This is something we will investigate in future work (section 4).

(a) Box plots comparing the performance of selected strategies used in the correlated damage environment.

(b) Box plots comparing the performance of selected strategies used in the variable damage environment. More outliers are present in the form of red '+'s.

Figure 10. Box plots of samples taken from the Homogeneous Group Tests in Table 1. Blue lines indicate the 95% confidence intervals, red lines are median rewards and red '+'s are outliers. '***' indicates p -value < 0.001, '**' indicates p -value < 0.01. These symbols are displayed in the midpoint between the two groups being compared.

3.2 Effect of Larger Swarm Sizes

The swarm size can be scaled up to discern whether the expected change in performance and efficiency is monotonically increasing with more agents, or if there is a limitation on this strategy for larger sizes, meaning the strategy would not be scalable. The swarm sizes to be investigated and compared to $P = 4$ are $P = 8, 16$ and 32 for the Swarm Level strategy. Any swarm larger than this would need to be tested using a larger grid size, and potentially swarms could be composed of hundreds of robots.¹⁷

These results show that the accumulated rewards reach a limit of about 3250 in Fig. 11(a), although lowering the damage capacity makes the trend more linear. This is because a system with fewer limiting constraints is able to explore the environment at a similar rate when the swarm size is doubled, so increasing the swarm size has a more significant effect. The change in average group efficiency with time (Fig. 11(b)) shows a different result. Increasing the swarm size has a negative effect on the efficiency (estimate of how close the portfolio combination is to optimal), suggesting that adding more agents results in a lower trade-off between risk and return. The agents do not form close to efficient portfolios, and this is likely due to the overcrowding of agents and the exploration of areas that have already been covered. This increases the discount added to the rewards collected due to a lower maximum Sharpe ratio, as more risk is taken for zero reward when there are more agents. Also, the efficiency does not differ much when the damage capacity is 250 or 500 in Fig. 11(b). The efficiency has a direct effect on the accumulated rewards, but for damage capacities lower than 1000, there seems to be nearly the same trend as the swarm size increases. The effect of risk diversification may be an issue in this Swarm Level strategy when setting the Portfolio VaR limit. This is because diversification may have the disadvantage of 'rationalizing' expected returns that may be lower than desired.

3.3 Boldness Distribution in the Swarm

We now consider the potential to deliberately skew individual agents' VaR thresholds to obtain diversity – diversification – in behavior. There has been some recognition of the opportunities available in introducing behavioral heterogeneity, for example in Ayanian's 'DART' (Diversity-enhanced Autonomy in Robot Teams).²⁸ It is also increasingly recognized in biology that heterogeneity can be functional, especially in close-knit animal

(a) Average performance (with error bars) as the number of agents in the environment changes.

(b) Average efficiency (estimate of how close the portfolio combination is to optimal) as the number of agents in the environment changes.

Figure 11. Graphs depicting the effect of increasing the swarm size over three different individual damage capacities. Each point on the graph is an average taken over 30 simulations in the ‘variable’ environment (c.f. Fig. 3). The agents follow a homogeneous Swarm Level strategy. As the group size increases, there is a higher discount added to the rewards collected due to a lower maximum Sharpe ratio, as more overall risk is taken for little reward.

teams such as ant colonies.²⁹ In our exploration task, such diversity may be especially relevant in certain environments. Hunt et al.⁵ explores a possible mechanic of dynamic risk limits – a changing boldness distribution – in a robot algorithm inspired by social spiders, where collective behavior results in efficient and rapid prey attack. In that case, a skewed distribution of many more shy spiders (robots) than bold spiders (robots) seems to be effective.

Structured Environment Scenario

To exemplify the potential usefulness of ‘boldness’ (risk appetite) variability, we consider briefly a new set of environments that have more structure in terms of reward and damage locations than the previous scenarios (those in Fig. 3). The environment includes areas of high reward (elements towards the right tail of the gamma PDF) that will not have any radiation damage (termed as ‘safe areas’) that require exploration. However, areas of particularly high risk/radiation (termed as ‘barriers’) must be traversed in order to complete the task.

Possible applications may be for discovering survivors in disaster areas³⁰ or retrieving valuable property from a radioactive area. It differs from the previous environments as the agents will not receive much of a reward unless the necessary risks are taken. The new reward and damage maps are shown in Fig. 12.

(a) Reward Map with 4 safe areas located at (20,20), (20,80), (80,20) and (80,80), each with radius 5.

(b) Damage Map with hazardous 'barriers' surrounding the safe areas, which have 0 damage associated with them.

(c) Example Simulation with 2 bold agents and 2 shy agents. View of the grid world from the damage map. The risk model used by the shy agents to set their thresholds is ES.

Figure 12. Simulated environments in which 'bold' and 'shy' agents navigate in the structured environment scenario.

With this new environment, it may be more beneficial to have a subsection of agents explore past the damage barriers whilst others take longer to do so, effectively making a trade-off between exploration and exploitation. A key factor of decision making in finance is risk preference, which is the attitude that investors hold towards risk.³¹ It is assumed that agents are rational, but some agents may have different levels of risk preference. In this case, using one type of VaR model for a specific type of agent can generally produce a higher magnitude of risk. This can be done either by adjusting the confidence level or using a 'better' risk model; a model that may be more useful in a market risk context (i.e. a better predictor of risk that considers extreme market moves),

but does not necessarily make the other model obsolete. Here we briefly consider Expected Shortfall as the ‘shy’ (more risk averse) agents’ VaR model. An alternative method that would make agents more risk averse is to increase the confidence level of the Historical VaR being used for the bold agents. This has the same effect of increasing the risk metric to allow for a lower risk preference. For example, if using the Historical VaR at the 99% confidence level, this lowers the percentile because now the worst 1% of returns are considered instead of the worst 5%.

Using an equally split population of bold and shy agents, Fig. 12(c) shows that the shy agents are less exploratory than their bold counterparts. Within the first 50 time steps, one of the bold agents starts to navigate through a barrier, and its return series starts to fall significantly as a result. However, the shy agents have not yet moved past any of these areas. This is because the Expected Shortfall chooses lower thresholds for the shy agents over time (due to more risk aversion), as opposed to the Historical VaR thresholds of the bold agents. This means that bold agents are usually moving faster than shy agents, which in turn results in a trade-off between more exploration subject to the cost of receiving more damage. Shy agents may not adapt to relatively low returns in this case, but this lowers the chance of entering risky areas and leaves this task to the bold agents. A shy agent has the benefit of increasing speed when it is in the bold agent’s communication map (due to the defining properties of Local Observability with broadcasting), so should eventually be able to reach the safe area without staying in the barrier as long. We will investigate this further in future work.

4. CONCLUSIONS

Here we have developed some simple control algorithms for use in exploring unknown hazardous environments, in which elements from Quantitative Finance have been adapted to the swarm robotics setting. In particular, we make an initial conceptual exploration of self-triggered avoidance behaviors, based either on an individual-level or swarm-level (global) Value at Risk (VaR) metric, and which can be broadcast to nearby neighbors. A swarm-level VaR violates the decentralization assumption commonly made in swarm robotics, but it could be implemented on the basis of sporadic communications with a base. By calculating VaR – and setting a VaR limit – at the swarm level, it allows the user to take advantage of the ‘diversification’ in agent returns. Agents that suffer short-term poor performance can be compensated for by better performance in others, without breaching the user’s global (swarm-level) VaR limit. This should help to operationalize the notion of attritability (expendability) within a managed risk framework.

An intrinsic feature, and possible limitation, of the Historical VaR is that the risk is estimated based on a series of past observations. This has the advantage of being simple to calculate for minimal swarm robots, and does not require any forward looking planning (e.g. Hakobyan et al.¹⁴). In certain circumstances, a reflexive behavior based on the instantaneous damage received at each time step could perhaps be a simpler and more reliable method. However, VaR is much more useful for exploration purposes as it also takes into account the potential reward to be gained, and is generally good for predicting areas of high damage under certain requirements. If we were to develop our very simple agent-based model further, we could consider introducing dynamic VaR thresholds. For example, Castello et al.³⁴ have investigated an Adaptive Response Threshold Method. Another extension would be to use a probabilistic model of failure,²⁰ which may be a more realistic method of modeling radiation instead of the existing additive mechanism. Also, our very simple random-walk movement model could easily be enhanced by employing a pheromone-based coordination strategy as used by ants, for example,³⁵ such that the agents leave ‘trails’ behind that reduce repeated exploration of the same areas.^{36,37}

Our next steps will be to develop realistic, physics-based simulations of robot swarms in the ‘Gazebo’ simulator. This will allow us to test the relevance of backward-looking, VaR based risk metrics for controlling swarms of real robots. We will implement realistic simulations of radiation hazards, using tools developed by the UK RAIN (Robotics and AI in Nuclear) Hub³² (Fig. 13(a)). It will also allow us to rigorously investigate the idea of different risk appetites (‘boldness’ levels) for the individual robots (Fig. 13(b)), and identify the conditions under which this is adaptive and the associated shape of the boldness distribution (e.g. positively skewed, with more ‘shy’ individuals). We intend to complement such simulation work with some real-world field trials, and in pursuit of the idea of attritable swarms we are testing simple, Raspberry Pi-based robots, such as the ‘Monsterborg’ from PiBorg.org³³ which retails for around \$300.

(a) Simulated radiation environments, developed with tools made by the UK RAIN (Robotics and AI in Nuclear) Hub.³² The simulated facility is shown on the left, and the radiation map on the right (c.f. Fig. 12 for example).

(b) Risk appetite (boldness) distributions in more realistic simulated robots. These robots are modeled on the ‘Monsterborg’.³³

Figure 13. Future work will examine realistic, physics-based simulations in the ‘Gazebo’ robot simulator.

Making swarms work effectively as a collective in dangerous environments promises to unlock a very wide range of applications, from responding to natural disasters to monitoring hazardous radiation. Our initial investigations indicate that there is considerable scope to consider financial risk management techniques in robot control. This seems particularly true for swarms where there is a natural correspondence between multiple agents of variable behavior and portfolio management of multiple assets. Approaches already widely employed in finance may therefore provide ‘off the shelf’ mathematical risk modeling opportunities for swarm robotics.

ACKNOWLEDGMENTS

EH acknowledges support from the Royal Academy of Engineering and the Office of the Chief Science Adviser for National Security under the UK Intelligence Community Postdoctoral Fellowship Programme.

REFERENCES

- [1] Hamann, H., [*Swarm Robotics: A Formal Approach (1st ed.)*], Springer International Publishing (2018).
- [2] Camazine, S., Deneubourg, J.-L., Franks, N., Bonabeau, B., and Theraulaz, G., [*Self-Organization in Biological Systems*], Princeton University Press (2001).

- [3] Colombi, J., Bentz, B., Recker, R., Lucas, B., and Freels, J., "Attritable design trades: Reliability and cost implications for unmanned aircraft," *11th Annual IEEE International Systems Conference, SysCon 2017 - Proceedings* (2017).
- [4] Gianni, M., "Towards Expendable Robot Teaming in Extreme Environments," *International Journal of Mechanical Engineering and Robotics Research* **8**(6), 830–838 (2019).
- [5] Hunt, E. R., Jenkinson, G., Wilsher, M., Dettmann, C. P., and Hauert, S., "SPIDER: a Bioinspired Swarm Algorithm for Adaptive Risk-Taking," *Artificial Life Conference Proceedings* (32), 44–51 (2020).
- [6] Hunt, E. R. and Hauert, S., "A checklist for safe robot swarms," *Nature Machine Intelligence* **2**(8), 420–422 (2020).
- [7] Jorion, P., [*Value at Risk: The New Benchmark for Managing Financial Risk (3rd ed.)*], McGraw-Hill Professional Publishing (2000).
- [8] Hull, J. C., [*Options, Futures, and Other Derivatives (6th ed.)*], Prentice Hall (2006).
- [9] Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D., "Coherent measures of risk," *Mathematical finance* **9**(3), 203–228 (1999).
- [10] Sharecast, "Daily closing values of the FTSE100." https://www.sharecast.com/index/FTSE_100/prices ((Accessed: 30/10/2019)).
- [11] Chow, Y., Tamar, A., Mannor, S., and Pavone, M., "Risk-sensitive and robust decision-making: a CVaR optimization approach," in [*Proceedings of the 28th International Conference on Neural Information Processing Systems-Volume 1*], 1522–1530 (2015).
- [12] Majumdar, A. and Pavone, M., "How Should a Robot Assess Risk? Towards an Axiomatic Theory of Risk in Robotics," in [*Robotics Research. Springer Proceedings in Advanced Robotics*], Amato, N. M., Hager, G., Thomas, S., and Torres-Torriti, M., eds., 75–84, Springer International Publishing, Cham (2020).
- [13] Singh, S., Chow, Y., Majumdar, A., and Pavone, M., "A framework for time-consistent, risk-sensitive model predictive control: Theory and algorithms," *IEEE Transactions on Automatic Control* **64**(7), 2905–2912 (2019).
- [14] Hakobyan, A., Kim, G. C., and Yang, I., "Risk-aware motion planning and control using CVaR-constrained optimization," *IEEE Robotics and Automation Letters* **4**(4), 3924–3931 (2019).
- [15] Hakobyan, A. and Yang, I., "Wasserstein Distributionally Robust Motion Planning and Control with Safety Constraints Using Conditional Value-at-Risk," in [*2020 IEEE International Conference on Robotics and Automation (ICRA)*], 490–496 (2020).
- [16] Rockafellar, R. T. and Uryasev, S., "Conditional Value-at-Risk for General Loss Distributions," *Journal of Banking & Finance* **26**(7), 1443–1471 (2002).
- [17] Rubenstein, M., Ahler, C., and Nagpal, R., "Kilobot: A low cost scalable robot system for collective behaviors," *Proceedings - IEEE International Conference on Robotics and Automation*, 3293–3298 (2012).
- [18] Hüttenrauch, M., Oic, A., and Neumann, G., "Deep reinforcement learning for swarm systems," *Journal of Machine Learning Research* **20**, 1–31 (2019).
- [19] Hunt, E. R., "Phenotypic Plasticity Provides a Bioinspiration Framework for Minimal Field Swarm Robotics," *Frontiers in Robotics and AI* **7**, 23 (2020).
- [20] Cheng, K. and Dasgupta, P., "Dynamic area coverage using faulty multi-agent swarms," *Proceedings of the IEEE/WIC/ACM International Conference on Intelligent Agent Technology, IAT 2007*, 17–23 (2008).
- [21] Kalman, R. E., "Mathematical Description of Linear Dynamical Systems," *Journal of the Society for Industrial and Applied Mathematics Series A Control* **1**(2), 152–192 (1962).
- [22] Zhou, L. and Tokekar, P., "Active Target Tracking With Self-Triggered Communications in Multi-Robot Teams," *IEEE Transactions on Automation Science and Engineering* **16**(3), 1085–1096 (2019).
- [23] Sharpe, W. F., "Mutual Fund Performance," *Journal of Business* **39**(S1), 119–138 (1966).
- [24] Markowitz, H. M., "Portfolio Selection," *The Journal of Finance* **7**(1), 77–91 (1952).
- [25] Mathworks, "Matlab financial toolbox." https://uk.mathworks.com/help/finance/index.html?s_tid=CRUX_lftnav ((Accessed: 12/30/2019)).
- [26] Steuer, R. E., Qi, Y., and Hirschberger, M., "Portfolio optimization: New capabilities and future methods," *Zeitschrift für Betriebswirtschaft* **76**(2), 199–220 (2006).

- [27] Calvo, C., Ivorra, C., and Liern, V., “On the computation of the efficient frontier of the portfolio selection problem,” *Journal of Applied Mathematics* **2012** (2012).
- [28] Ayanian, N., “DART: Diversity-enhanced Autonomy in Robot Teams,” *The International Journal of Robotics Research* **38**(12-13), 1329–1337 (2019).
- [29] O’Shea-Wheller, T. A., Hunt, E. R., and Sasaki, T., “Functional Heterogeneity in Superorganisms: Emerging Trends and Concepts,” *Annals of the Entomological Society of America* (11 2020). saaa039.
- [30] Higgins, F., Tomlinson, A., and Martin, K. M., “Threats to the swarm : Security considerations for swarm robotics,” *International Journal on Advances in Security* **2(2&3)**, 288–297 (2009).
- [31] Gal, D. and Rucker, D. D., “The Loss of Loss Aversion: Will It Loom Larger Than Its Gain?,” *Journal of Consumer Psychology* **28(3)**, 497–516 (2018).
- [32] UK RAIN Hub, “ROS-Compatible Radiation Sources and Sensors.” <https://rainhub.org.uk/our-technologies/ros-compatible-radiation-sources-and-sensors/> ((Accessed: 3/30/2021)).
- [33] PiBorg, “MonsterBorg — The Ultimate Raspberry Pi Robot.” <https://www.piborg.org/robots-1/monsterborg> ((Accessed: 3/30/2021)).
- [34] Castello, E., Yamamoto, T., Libera, F. D., Liu, W., Winfield, A. F. T., Nakamura, Y., and Ishiguro, H., “Adaptive foraging for simulated and real robotic swarms: the dynamical response threshold approach,” *Swarm Intelligence* **10(1)**, 1–31 (2016).
- [35] Hunt, E. R., Franks, N. R., and Baddeley, R. J., “The Bayesian superorganism: externalized memories facilitate distributed sampling,” *Journal of The Royal Society Interface* **17**(167), 20190848 (2020).
- [36] Fossum, F., Montanier, J., and Haddow, P. C., “Repellent pheromones for effective swarm robot search in unknown environments,” in *[2014 IEEE Symposium on Swarm Intelligence]*, 1–8 (2014).
- [37] Hunt, E. R., Jones, S., and Hauert, S., “Testing the limits of pheromone stigmergy in high-density robot swarms,” *Royal Society Open Science* **6**, 190225 (2019).