ARL-TR-9961 • AUG 2024

State-Space Perturbation to Assess Multi-Agent Coordination

by Erin Zaroukian, Rolando Fernandez, and Derrik Asher

NOTICES

Disclaimers

ARL-TR-9961 ● AUG 2024

State-Space Perturbation to Assess Multi-Agent Coordination

Erin Zaroukian, Rolando Fernandez, and Derrik Asher
DEVCOM Army Research Laboratory

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED START DATE	END DATE
August 2024	Technical Report	January 2019	February 2021

4. TITLE AND SUBTITLE
State-Space Perturbation to Assess Multi-Agent Coordination

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**

6. AUTHOR(S)
Erin Zaroukian, Rolando Fernandez, and Derrik Asher

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
DEVCOM Army Research Laboratory ATTN: FCDD-RLA-IC Aberdeen Proving Ground, MD 21005	ARL-TR-9961

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES
ORCIDs: Erin Zaroukian, 0000-0002-1381-085X; Derrik Asher, 0000-0002-1217-8581; Rolando Fernandez, 0000-0001-6950-6251

14. ABSTRACT
This report outlines state-space perturbation, a method developed to perturb the state space of a reinforcement learning agent to determine the degree to which its behavior is influenced by the behavior of other agents in its state space. Two applications are described: a predator–prey pursuit task and a turret-reconnaissance task, followed by a formal description of the method and a guide to performing state-space perturbation using internally developed code.

15. SUBJECT TERMS
simulation experiments, marl, coordinated behavior, methods, game theory, Military Information Sciences

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	C. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UU	20

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
Erin Zaroukian	(410) 278-3203

STANDARD FORM 298 (REV. 5/2020)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

1. Introduction

State-space perturbation was developed by Asher et al. (2019) and Fernandez et al. (2021) as a method to measure collaboration and the causal links that can develop between Multi-Agent Reinforcement Learning (MARL) agents. This method compares action outputs from a MARL agent's policy (baseline) to the same agent's action outputs when its state space has been systematically perturbed to alter perception of its teammates. By this method, an agent can be caused to "hallucinate" the positions of its teammates, and a researcher can determine whether this change in perception causes that agent to act differently, which would imply that the agent is coordinating its movements with its teammate. Perturbation in these two studies is briefly described in Fig. 1, with greater detail provided in Figs. 6 and 7 and in Sections 1.1 and 1.2.

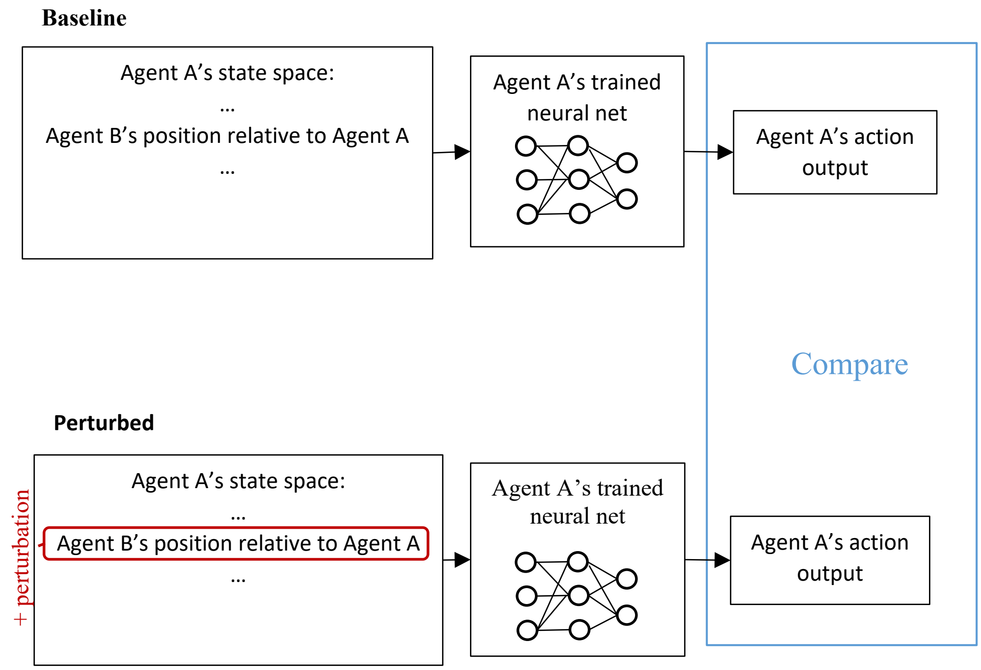

Fig. 1 General schematic of state-space perturbation to measure collaboration between Agent A and Agent B, where Agent B's relative position is perturbed within Agent A's state space

1.1 Predator–Prey Pursuit Task

In Asher et al. (2019), state-space perturbation is introduced in a predator–prey pursuit task, where three predators and one prey are trained to pursue/avoid each other in a continuous and bounded 2D environment using a Multi-Agent Deep Deterministic Policy Gradient (MADDPG) algorithm (Lowe et al. 2017). After gathering baseline actions, each predator's state space is independently perturbed by modifying that predator's perception of a partner predator's location, as illustrated in Fig. 2. These baseline and perturbed results were compared and aggregated to produce coordination profiles, or plots of percent change between baseline and perturbation for each predator with each perturbed teammate (Fig. 3).

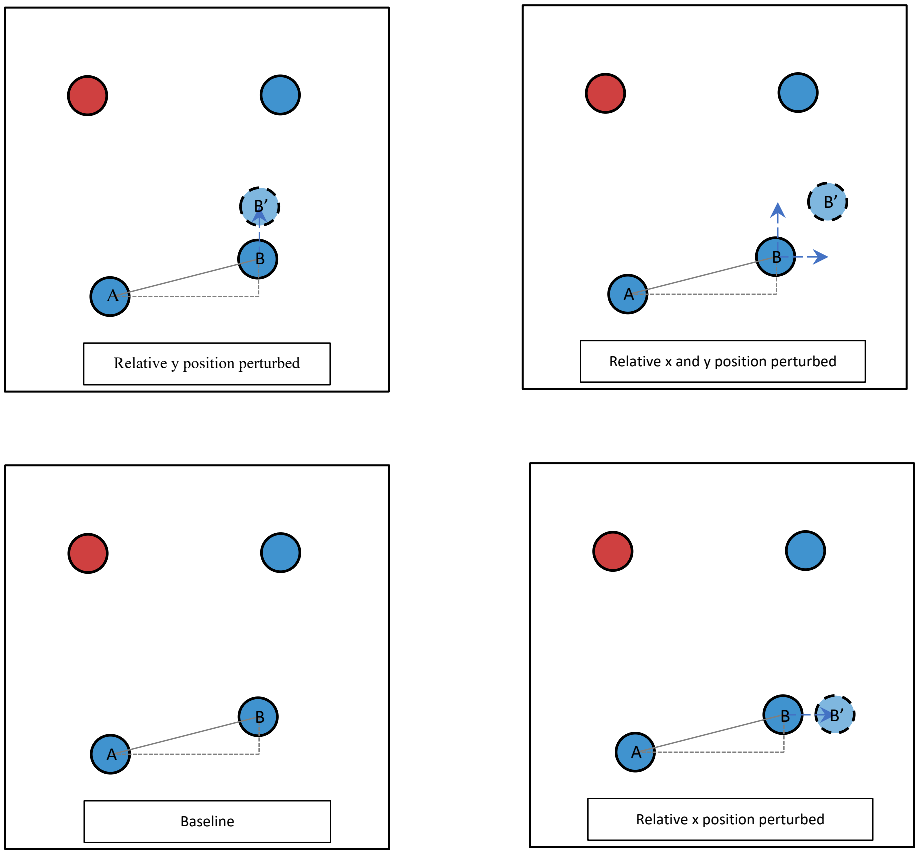

Fig. 2 Example schematic of Agent B perturbed in Agent A's state space. Bottom left: Baseline non-perturbed space showing three predators (including Agents A and B) and one prey. Upper left: The y component of Agent B's position relative to Agent A is perturbed in Agent A's state space, shown in location B'. Lower right: The x component of Agent B's position relative to Agent A is perturbed in Agent A's state space, shown in location B'. Upper right: Both the x and y component of Agent B's position relative to Agent A is perturbed in Agent A's state space, shown in location B'.

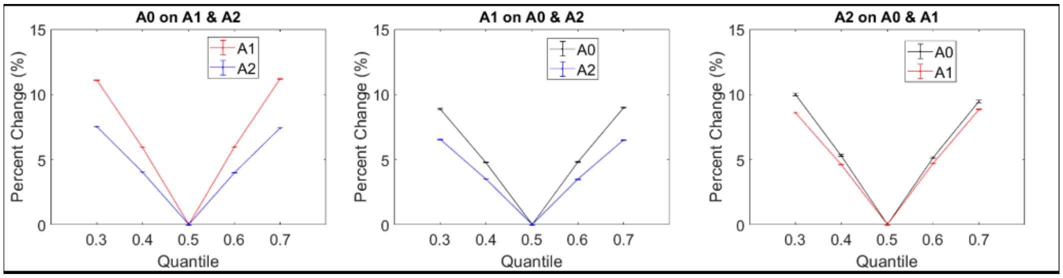

Fig. 3 **Coordination profiles (Reprinted from Asher et al. [2019]) showing the absolute value of the mean percent change in action, converted to velocity, from baseline (equivalent to 0.50 quantile) for predators A0, A1, and A2 when each of their perceptions of their teammates' position is independently perturbed by different amounts (0.3, 0.4, 0.5, 0.6, and 0.7 quantiles of the relative distances between all agents in baseline). This figure shows A0's relative distance perturbed independently in the state space of A1 and A2 (left), A1's relative distance perturbed independently in the state space of A0 and A2 (center), and A2's relative distance perturbed independently in the state space of A0 and A1 (right). The values plotted are medians with standard error of the median for data collapsed across perturbation dimension.**

Additionally, link node diagrams, as in Fig. 4, averaged ratios of mean percent change shown in Fig. 3.

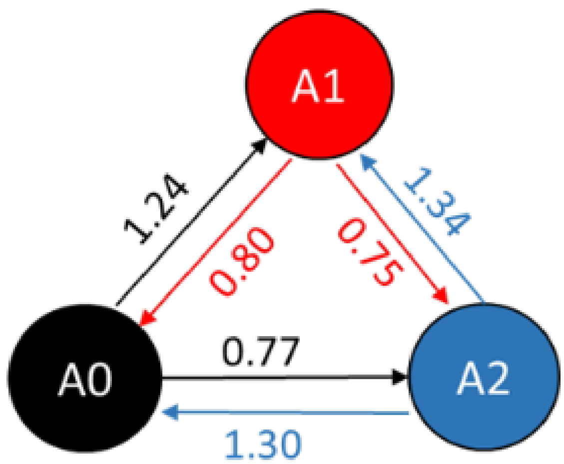

Fig. 4 **Link-node diagram (Reprinted from Asher et al. [2019]) showing the relative strength of impact of the agents on each other. Here, Agent A's relative impact on Agent B was calculated by collapsing data across perturbation dimensions and dividing the median of the mean percent change of B's velocity after A's position is perturbed in B's state space, summed across all five quantiles, by the median of the mean percent change of A's velocity after B's perceived position is perturbed in A's state space, again summed across all five quantiles.**

1.2 Turret-Reconnaissance Task

In Fernandez et al. (2021), state-space perturbation is extended to a turret-reconnaissance task, where two reconnaissance agents attempt to enter a turret's goal region before the turret can hit them. As in the previous study, this scenario utilized a continuous and bounded 2D environment where agents were trained with an MADDPG algorithm. The reconnaissance agents' state spaces were perturbed in the same manner as the predator agents' were above. A resulting coordination profile is shown in Fig. 5.

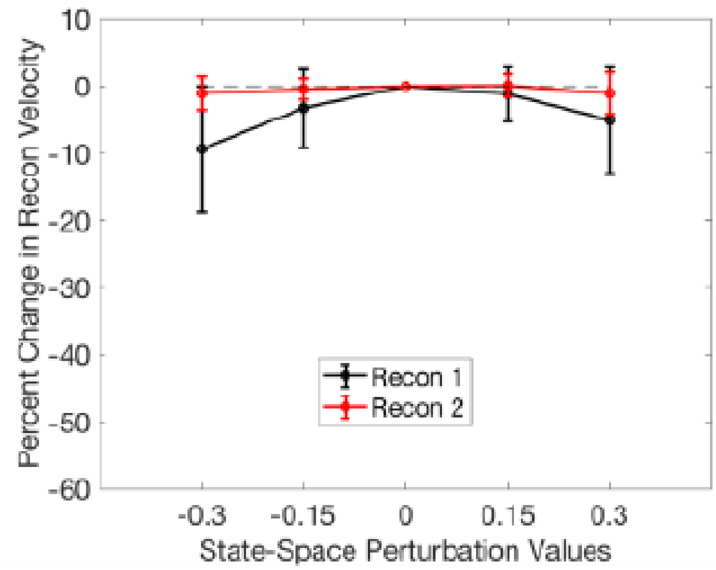

Fig. 5 Coordination profile (Reprinted from Fernandez et al. [2021]) showing the mean percent change in action (converted to velocity) from baseline for reconnaissance agents Recon 1 and Recon 2 when each of their perceptions of the other's position is independently perturbed by different amounts (–0.3, –0.15, 0, 0.15, and 0.3). The figure shows Recon 1's relative distance perturbed in Recon 2's state space (red), and Recon 2's relative distance perturbed in Recon 1's state space (black). Bars show standard error of the mean.

2. Method

At a high level, state-space perturbation manipulates the state of a single agent (here, its relative distance to other agents) to evaluate how those manipulations changed other agents' neural nets' action outputs. A formal description is provided in Section 2.1, followed by an outline of how to perform state-space perturbation using internally developed code with schematic representations.

2.1 Formal Description of State-Space Perturbation

A pair of agents, A and B, are situated in a continuous 2×2 space. A's behavior is determined through a function f_A that maps from A's state space to A's action output. A's state space ss_{A_i} contains a variety of information about A's environment at each time step $i \in \{0, 1, \dots n\}$, including B's position relative to A's position, stored as separate x and y components, $\mathrm{B}_{\mathrm{x,i}}$ and $\mathrm{B}_{\mathrm{y,i}}$. For simplicity, we will represent ss_{A_i} simply as $(\mathrm{B}_{\mathrm{x,i}}, \mathrm{B}_{\mathrm{y,i}})$, and the resulting action output from f_A will be represented as the x and y components of its acceleration, $(A_{\mathrm{a}_{x,i}}, \mathrm{A}_{\mathrm{a}_{y,i}})$.

$$f_A\big((B_{x,i}, B_{y,i})\big) = (A_{a_{x,i}}, A_{a_{y,i}})$$

A's velocity at time i is computed by adding this acceleration to A's velocity at time $i-1$*:

$$A_{v_{x,i-1}} + A_{a_{x,i}} = A_{v_{x,i}}$$

$$A_{v_{y,i}} + A_{a_{y,i}} = A_{v_{y,i}}$$

To assess the impact of B's behavior on A, we perturb B's position relative to A in dimension $\delta \in \{x, y, xy\}$, where x represents horizontal, y represents vertical, and xy represents both horizontal and vertical. That is, we perturb $B_{\delta,i}$ ($B_{x,i}$, $B_{y,i}$, or both) within ss_{A_i}. The perturbation is accomplished by adding a perturbation value Q_α for $\alpha \in \{3, 4, 5, 6, 7\}$ to $B_{\delta,i}$ at each time step i. Q_α is a quantile value determined as follows:

> For all time steps, rank the distances between all pairs of agents by pooling all x distances and all y distances (here, with only two agents A and B, this is equivalent to $(B_{x,i})_{i=0}^{n} \cup (B_{y,i})_{i=0}^{n}$ within ss_{A_i}) and partition into 10 subsets of equal size. The cut points between these subsets represent the 1, 2, 3, 4, 5, 6, 7, 8, and 9 quantile values, $Q_1, Q_2, \ldots Q_9$.

B's perturbed location is calculated by adding B's relative perturbed distance from A (i.e., $B_{\delta,i} + Q_\alpha$) to A's absolute location $(A_{x,i}, A_{y,i})$, and a validity filter is applied to check that the perturbed location remains within the 2 × 2 arena:

> $\delta = x$ (Perturbing $B_{x,i}$): $-1 < (A_{x,i} + B_{x,i} + Q_\alpha) < 1$
>
> $\delta = y$ (Perturbing $B_{y,i}$): $-1 < (A_{y,i} + B_{y,i} + Q_\alpha) < 1$
>
> $\delta = xy$ (Perturbing both $B_{x,i}$ and $B_{y,i}$): $-1 < (A_{x,i} + B_{x,i} + Q_\alpha) < 1$ and $-1 < (A_{y,i} + B_{y,i} + Q_\alpha) < 1$

If invalid, this time step i in this dimension δ for perturbation Q_α is excluded from analysis. Otherwise, $B_{\delta,i}$ is replaced with $B_{\delta,i} + Q_\alpha$ in ss_{A_i} and the resulting action output is computed as $(A_{a_{x,i}}^{Q_\alpha, B_{\delta,i}}, A_{a_{y,i}}^{Q_\alpha, B_{\delta,i}})$, the x and y component of A's acceleration after the δ component of B's relative distance from A has been perturbed by Q_α.

---

* Within the environment described in Section 2.2, velocity computations also include an environmental damping factor, and a maximum velocity is respected by scaling $\frac{(A_{v_{x,i}}, A_{v_{y,i}})}{\sqrt{A_{v_{x,i}}^2 + A_{v_{y,i}}^2}} * v_{max}$ when $\sqrt{A_{v_{x,i}}^2 + A_{v_{y,i}}^2} > v_{max}$.

$\delta = x$ (Perturbing $B_{x,i}$): $f_A\left((B_{x,i} + Q_\alpha, B_{y,i})\right) = (A_{a_{x,i}}^{Q_\alpha,B_{x,i}}, A_{a_{y,i}}^{Q_\alpha,B_{x,i}})$

$\delta = y$ (Perturbing $B_{y,i}$): $f_A\left((B_{x,i}, B_{y,i} + Q_\alpha)\right) = (A_{a_{x,i}}^{Q_\alpha,B_{y,i}}, A_{a_{y,i}}^{Q_\alpha,B_{y,i}})$

$\delta = x$y (Perturbing both $B_{x,i}$ and $B_{y,i}$): $f_A\left((B_{x,i} + Q_\alpha, B_{y,i} + Q_\alpha)\right) = (A_{a_{x,i}}^{Q_\alpha,B_{xy,i}}, A_{a_{y,i}}^{Q_\alpha,B_{xy,i}})$

As above, A's velocity post-perturbation at i is computed by adding this acceleration to A's velocity at time $i - 1$.

$$A_{v_{x,i-1}}^{\square} + A_{a_{x,i}}^{Q_\alpha,B_{\delta,i}} = A_{v_{x,i}}^{Q_\alpha,B_{\delta,i}}$$

$$A_{v_{y,i-1}}^{\square} + A_{a_{y,i}}^{p_Q,B_{\delta,i}} = A_{v_{y,i}}^{Q_\alpha,B_{\delta,i}}$$

The percent change in A's x and y velocity due to the perturbation can be compared to its velocity at the same time step without perturbation (baseline) by computing the ratio of the difference between the perturbed and unperturbed velocities and the unperturbed velocity.

Percent change:

$$A_{\omega_{x,i}}^{Q_\alpha,B_{\delta,i}} = \frac{A_{v_{x,i}}^{\square} - A_{v_{x,i}}^{Q_\alpha,B_{\delta,i}}}{A_{v_{x,i}}^{\square}} * 100$$

$$A_{\omega_{y,i}}^{p_\alpha,B_{\delta,i}} = \frac{A_{v_{y,i}}^{\square} - A_{v_{y,i}}^{p_\alpha,B_{\delta,i}}}{A_{v_{,i}}^{\square}} * 100$$

The percent change mean is simply the mean of the x and y percent changes $A_{\omega_{x,i}}^{Q_\alpha,B_{\delta,i}}$ and $A_{\omega_{y,i}}^{Q_\alpha,B_{\delta,i}}$.

Percent change mean: $\overline{A_{\omega_\iota}^{p_\alpha,B_{\delta,\iota}}} = \frac{A_{\omega_{x,i}}^{Q_\alpha,B_{\delta,i}} + A_{\omega_{y,i}}^{Q_\alpha,B_{\delta,i}}}{2}$

Coordination profile:

To represent results of perturbations in a coordination profile, for each pair of agents A and B across all i, α, δ values, compute a summary statistic (here, median, $med(\forall i, \alpha, \delta, \overline{A_{\omega_\iota}^{p_\alpha,B_{\delta,\iota}}})$), then plot and compare values for different pairs of agents.[†]

Link-node diagram:

[†] In the environment described in Section 2.2, outliers are removed before computing summary statistics.

Using the summary statistics from coordination profiles, calculate the relative impact of B on A by dividing the sum of A's summary statistic across all perturbation values by the sum of B's summary statistic across all perturbation values (e.g., using median as the summary statistic $\frac{\sum_{\delta=3}^{7} med(\forall i,\alpha,\delta,\overline{A_{\omega_\iota}^{p_{\alpha,B_{\delta,\iota}}}})}{\sum_{\delta=3}^{7} med(\forall i,\alpha,\delta,\overline{B_{\omega_\iota}^{p_{\alpha,A_{\delta,\iota}}}})}$).

2.2 State-Space Perturbation using MADDPG

The following describes in detail how state-space perturbation was performed for a predator–prey task in OpenAI Gym (Brockman et al. 2016) using the MADDPG algorithm with internally developed code.[‡] Much of this process is illustrated in Figs. 6 and 7.

- Train model using deep MARL.
  - Run `MADDPG/scripts/train_MADDPG.sh`
- Generate baseline data.
  - Run `MADDPG/scripts/perturbation/eval_predatorprey_MADDPG.sh`
- Choose perturbation values.
  - In Asher et al. (2019), the distance between each pair of agents at each time step was computed from the baseline data, and the values at the 0.3, 0.4, 0.5, 0.6, and 0.7 quantiles were selected as the perturbation values.
  - In Fernandez et al. (2021), the values at these quantiles that were chosen are approximated at –0.3, –0.15, 0, 0.15, and 0.3.
- Independently perturb each agent's state space and generate new action value data.
  - Run `MADDPG/scripts/perturbation/perturb_predatorprey_MADDPG.sh`
    - This script perturbs each agent's state space 2 (predator teammates) × 5 (perturbation values corresponding to 0.3, 0.4, 0.5, 0.6, and 0.7

[‡] Code is located at https://gitlab.sitcore.net/Multi-Agent-Teaming/MADDPG and https://gitlab.sitcore.net/Multi-Agent-Teaming/Multi-Agent-Particle-Environment. Contact authors for access.

quantiles) × 3 (perturbing teammate's relative x position, perturbing teammate's relative y position, and simultaneously perturbing teammate's relative x and y positions) = 30 times for each time step in each episode.

- Invalid data (where the perturbation caused an agent to be located outside of the 2 × 2 arena) is removed.
- Action outputs (accelerations) are converted to velocities, which are typically easier to interpret.
- For each agent, each perturbed partner, and each perturbation dimension (x, y, and x and y simultaneously), `percent_changed` is calculated at each time step as $\frac{\text{baseline velocity} - \text{changed velocity}}{\text{baseline velocity}} * 100$ (see Fig. 8; note that velocity is divided into x and y components).
  - "`percent_changed_means`" is mean of the x and y components in the "`percent_changed`" folder (see Fig. 9).

- Analysis: compare perturbed results to non-perturbed results.
  - For each agent and each perturbed partner, aggregate `percent_changed_means` across perturbation dimensions.
  - Remove outliers (e.g., values more than 3 median absolute deviations) (median(|X_i - median(X)|)) from median.
    - Extreme values tend to occur when baseline velocities cross over 0 (e.g., from positive x to negative x), such that the percent change is very large.[§]
  - Plot values in a "coordination profile."
    - For each perturbed agent, plot the (absolute value of the) median (as in Asher et al. [2019][**]), or mean (as in Fernandez et al. [2021]), as appropriate, of `percent_changed_means` for each partner across all time steps, episodes, and perturbation dimensions (x, y, and x and y simultaneously) for each perturbation value.
  - If desired, create link node diagram.

[§] No system was used in Asher et al. (2019) or Fernandez at al. (2021) to remove the very small values for very small changed velocity. This step was omitted for convenience, and these values were assumed to be even across percentiles.
[**] See also, https://influentialpoints.com/Training/standard_error_of_median.htm.

- Using the summary statistics from coordination profiles, calculate the relative impact of one agent on another by dividing the latter's summary statistic (e.g., mean of `percent_changed_means` across all time steps, episodes, and perturbation dimensions, see Fig. 10), summed across all perturbation values, by the former's.

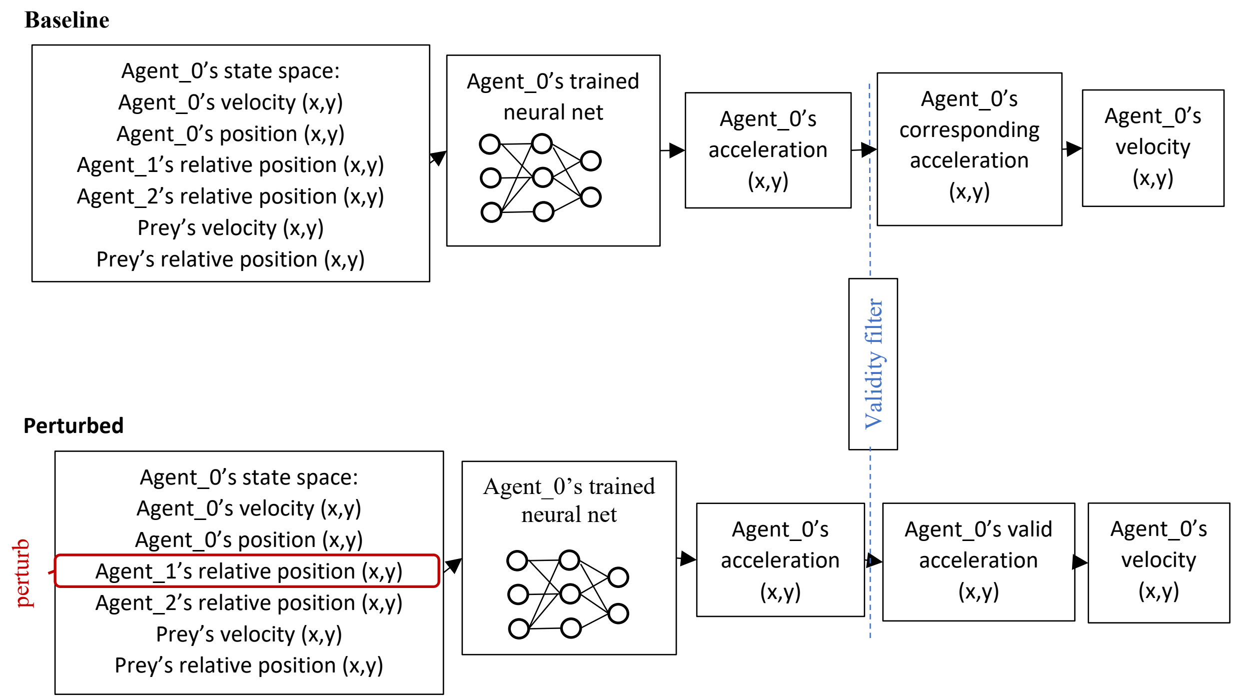

Fig. 6 A more detailed schematic of state-space perturbation in a predator–prey pursuit task, where Agent_1's relative position is perturbed in Agent_0's state space

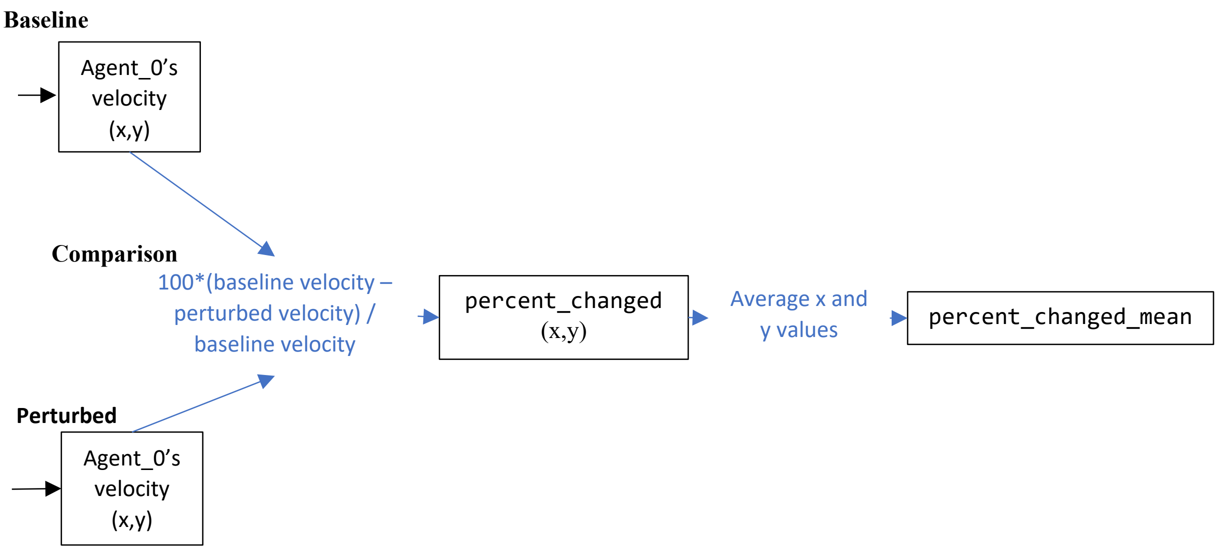

Fig. 7 Continued schematic of state-space perturbation illustrating how baseline and perturbed data are compared by computing the percent changed for the agent's baseline and perturbed x and y velocity components at each time step, and these x and y components are then averaged together at each time step

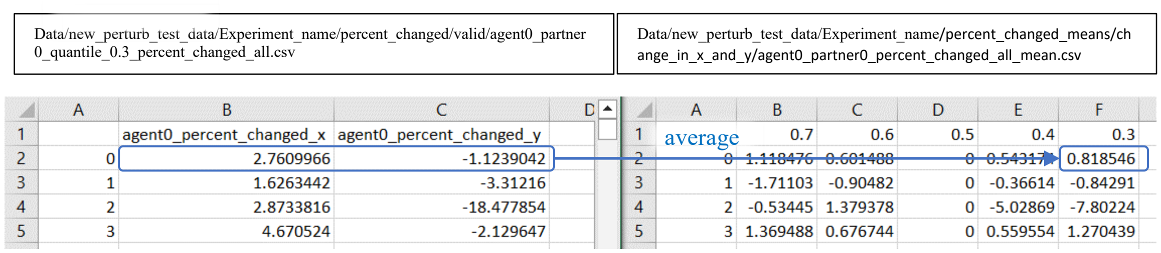

Fig. 8 Example of percent changed and the calculation of mean percent changed for Agent0, where its Partner0's relative position is perturbed by the 0.3-quantile value simultaneously in both the x and y dimensions

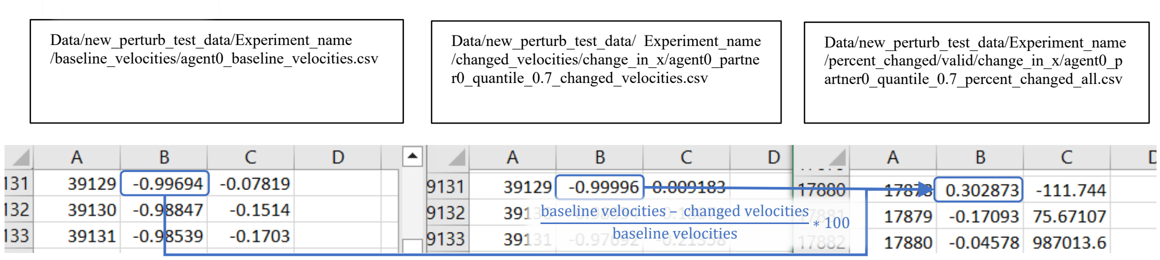

Fig. 9 Example of baseline velocities, changed velocities, and the calculation of percent changed for Agent0, where its Partner0's relative position is perturbed by the 0.7-quantile value in the x dimension

A’s relative impact on **B** =

median(

For each episode, timestep, quantile, and perturbation dimension,

mean(

percent change of the **x** component of **B**’s velocity after **A**’s relative [**x, y, x and y**] position is perturbed by the [**0.3, 0.4, 0.5, 0.6, 0.7**] quantile value in **B**’s state space

percent change of the **y** component of **B**’s velocity after **A**’s relative [**x, y, x and y**] position is perturbed by the [**0.3, 0.4, 0.5, 0.6, 0.7**] quantile value in **B**’s state space

)

)

divided by

median(

For each episode, timestep, quantile, and perturbation dimension,

mean(

percent change of the **x** component of **A**’s velocity after **B**’s relative [**x, y, x and y**] position is perturbed by the [**0.3, 0.4, 0.5, 0.6, 0.7**] quantile value in **A**’s state space

percent change of the **y** component of **A**’s velocity after **B**’s relative [**x, y, x and y**] position is perturbed by the [**0.3, 0.4, 0.5, 0.6, 0.7**] quantile value in **A**’s state space

)

)

Fig. 10 Description of how links in the link-node diagram are calculated

3. Conclusion

This state-space perturbation method can quantify the coordination between two deep-MARL agents’ movements, as shown in both a predator–prey pursuit task and a turret-reconnaissance task, by showing the relative strength by which one agent’s actions are influenced by another’s. Future work will determine how well this method performs with different tasks, as well as how it correlates with other measures of coordination.

4. References

Asher D, Garber-Barron M, Rodriguez S, Zaroukian E, Waytowich N. Multi-agent coordination profiles through state space perturbations. Proceedings of the 2019 International Conference on Computational Science and Computational Intelligence (CSCI); 2019 Dec. IEEE; c2019. p. 249–252.

Brockman G, Cheung V, Pettersson L, Schneider J, Schulman J, Tang J, Zaremba W. OpenAI Gym; 2016. arXiv:1606.01540.

Fernandez R, Basak A, Howell B, Hsu C, Zaroukian E, Perret J, Humann J, Dorothy M, Sharma P, Nivison S, Bell Z, Asher D. Multi-agent collaboration in an adversarial turret reconnaissance task. In: Intelligent Human Systems Integration 2021: Proceedings of the 4th International Conference on Intelligent Human Systems Integration (IHSI 2021): Integrating People and Intelligent Systems; 2021 Feb 22–24; Palermo, Italy. Springer International Publishing; c2021. p. 38–43.

Lowe R, Wu YI, Tamar A, Harb J, Pieter Abbeel O, Mordatch I. Multi-agent actor-critic for mixed cooperative-competitive environments. Advances in Neural Information Processing Systems. 2017;30.

Standard error of the median. n.d. [accessed 22 Aug 2024]. https://influentialpoints.com/Training/standard_error_of_median.htm

1 (PDF)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA
1 (PDF)	DEVCOM ARL FCDD RLB CI TECH LIB