Georgia Tech

MISDIRECTION IN ROBOT TEAMS: EXPLOITING ORGANIZATIONAL PRINCIPLES FOR OPERATIONAL ADVANTAGE

RONALD C. ARKIN, MICHAEL J PETTINATI, AKSHAY KRISHNAN AND SHENKANG CHEN

GEORGIA TECH, MOBILE ROBOT LAB

CREATING THE NEXT®

Supported by the National Science Foundation CNS EAGER Grant #1848653

Misdirection in Robot Teams:
Exploiting Organizational Principles for Operational Advantage

EAGER: CNS: Misdirection in Robot Teams: Exploiting Organizational Principles for Operational Advantage
Award # 1848653 10/2018 - 3/2021 Ronald Arkin Georgia Tech

Challenge

This project focuses on higher-level strategies for multi-robots: to misdirect and to counter-misdirect. As multi-robot systems become more autonomous, distributed, networked, numerous, and with more capability to make critical decisions, the prospect for intentional and unintentional misdirection must be anticipated and exploited.

Solution - Developed Multiple Strategies

- Push misdirection – Fear-based herding
- Pull misdirection – Judas goat/ Pied piper
- Counter Misdirection
- Exploit biological parallels

Key Innovations

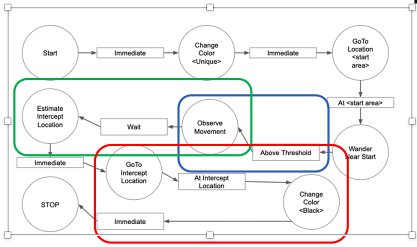
- Use of shills for misdirection
- Counter-misdirection for multi-robots

Fear-based push approach
https://commons.wikimedia.org/wiki/File:Karjus.jpg

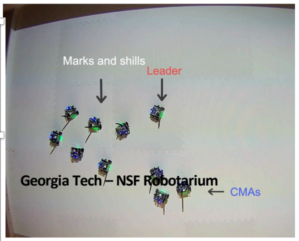
Pull Approach
https://en.wikipedia.org/wiki/File:Pied_Piper2.jpg

Added Shills and Counter-Misdirection Agents

Counter Misdirection FSA



Validation on Physical Robots



Scientific Impact

- Multiagent deception is a phenomenon that will continue to become more commonplace. It is important to develop these methods to better understand the phenomenon as well as to create methods to counter its effects if and when they are deployed.

Broader Impact

- Deception, unfortunately, is becoming commonplace in the cyberworld. It is important to study both the basis for such deception as well as methods to counter it.
- There are deep ethical questions associated with the use of deception and this project considers and discusses those effects

PROJECT MOTIVATION

Context: Multiagent robot teams in competitive environments

Objective: Move a team of agents from one location to another (a location that is beneficial for their adversaries)

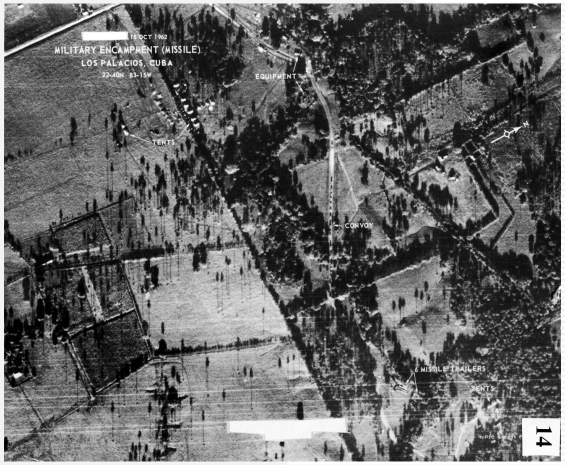
Examples:

Military

Entertainment

Approach:

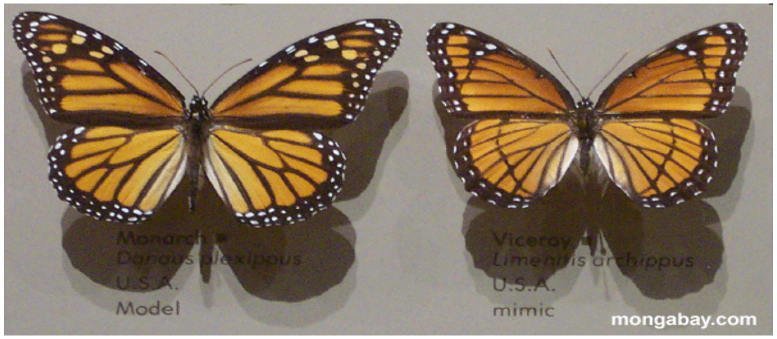
Develop and test different models of misdirection



https://commons.wikimedia.org/wiki/File:RoboCup_2019_Team_NimbRo.jpg

Why Deception?

- Deception is commonly used by animals
- The use of deception by primates may indicate Theory of Mind (Byrne & Whiten, 1990)
- "the development of deception follows the development of other skills used in social understanding" (Vasek, 1984)
- "another price you pay for higher-order intentionality is the opportunity [for] … deception" (Dennett, 1983)
- The Turing test is fundamentally based on deception as an indicator of human-level intelligence
- May be a hallmark of social intelligence (Byrne & Whiten, 1990)



Mimicry

Camouflage

Robot Deception

Robot deception: robots convey false information or conceal true information.

Research on robotic deception:

- Deceptive robot behaviors in search and rescue (Shim 2015).
- Deceptive robot behaviors in a gameplay scenario (Dragan 2014).
- Behavioral strategies for deception inspired by animals: mobbing (Davis 2012).
- Ethical considerations (Arkin 2010).

☑ Deceptive communication signals evolved in a simulated evolutionary environment (Floreano 07)

☑ Robot Deception in HRI

- ☑ Cheating robot in rock-paper-scissors game (Scasselati 10)
- ☑ A deceptive robot referee (Vazquez 11)
- ☑ Deception in robotic physical therapy system (Brewer 06)

Gameplay (Dragan 2014)

Mobbing (Davis 2012)

Misdirection in Robot Teams: Exploiting organization principles for operational advantage

Motivating Examples:

- Confidence tricks in humans (con artists and shills)
- Force deception in military operations
- Feints and ruses in sports
- Tactical (Intentional) Deception in Capuchin Monkeys and the Great Apes.
- Cultural deception in humans.
- Political misinformation and coordinated conspiracies.
- Strategic lying by human groups and reputation games.

WHAT GETS AGENTS MOVING?

- Other Agents
  - Repelling Agents
    - E.g. Predators
  - Pulling Agents
    - E.g. “Follow Me”, Conspecifics moving with intent
- Thresholds
  - Agents act when they see a certain number of others act in the same way.
    - Granovetter (1978)
    - “Threshold models of collective behavior”

ne.0112884.g001_a.png

https://commons.wikimedia.org/wiki/File:Followership.png

CREATING THE NEXT®

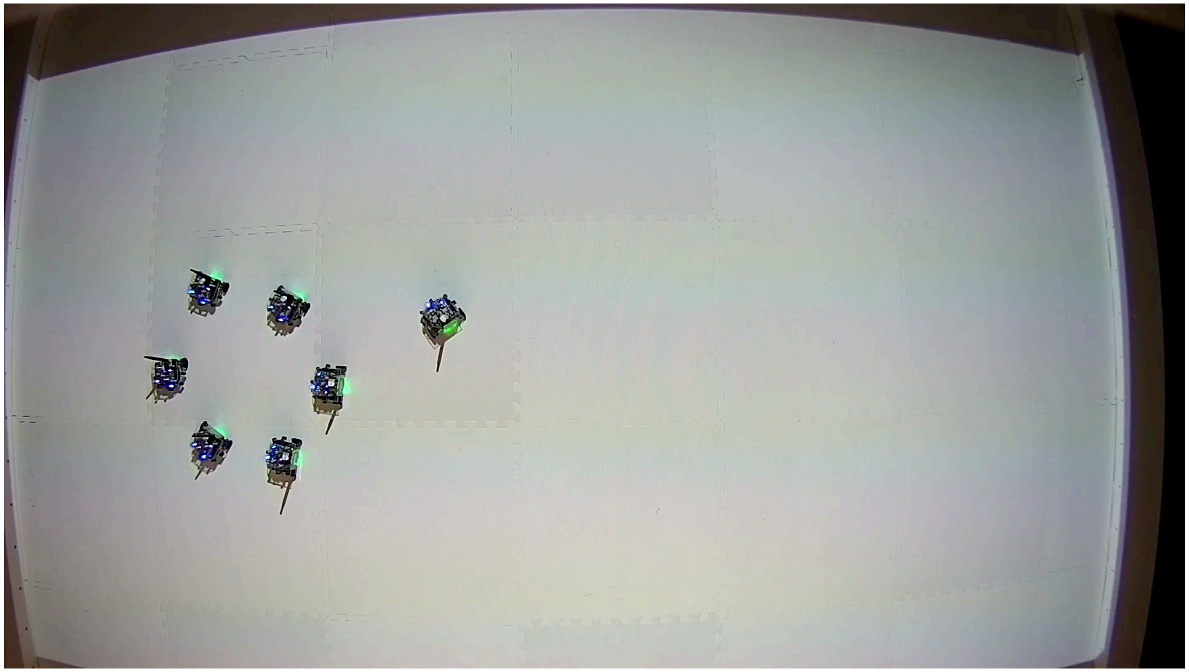
MOVING ROBOT TEAMS: USING SHILLS TO INDUCE MOVEMENT

- Three-shell game
- Crowd members pushes individual marks to play
- Shills draw marks into game by winning when it isn't populated with other marks

(Maurer 1947)

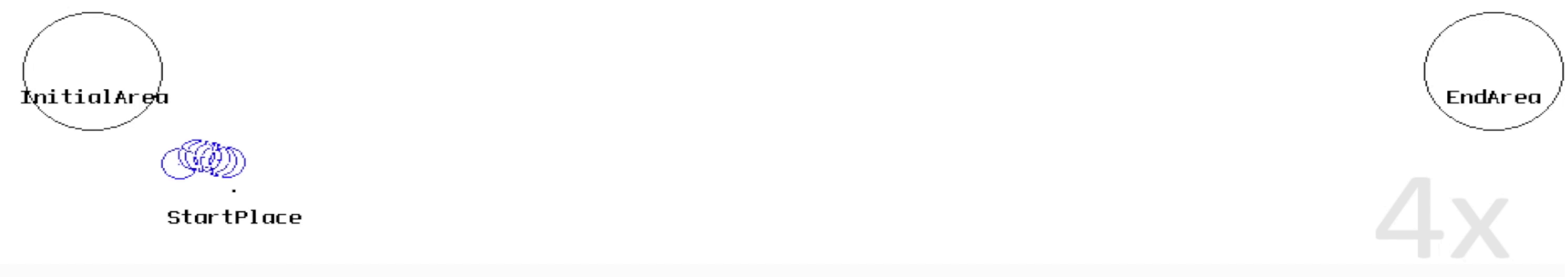
Misdirection in Robot Teams using Shills

“Wolves in Sheep’s Clothing: Using Shill Agents to Misdirect Multi-Robot Teams” (Pettinati 2020).



PUSH AND PULL APPROACH

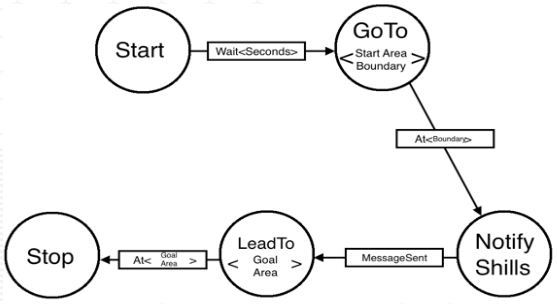
A shill agent (a leader proxy, brown) helps to efficiently move the group, which is being pushed from behind by a herding agent (orange).



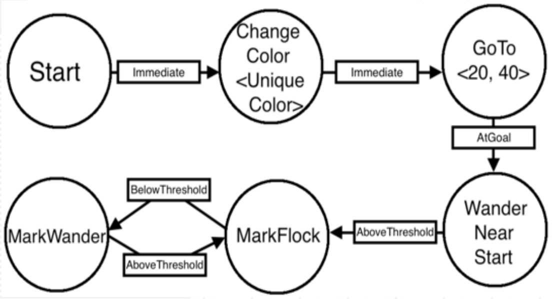
A shill agent and a herding agent move the marks from start to goal.

THE BEHAVIOR ASSEMBLAGES FOR EACH AGENT TYPE ALONG WITH THE COMPOSING BEHAVIORS.

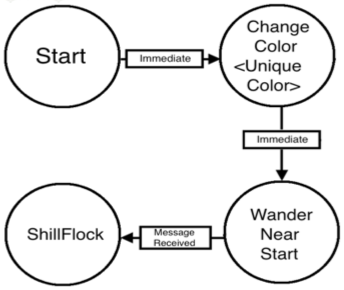
Robotic Agent	Behavior Assemblage	Composing Behaviors
Leader	Lead To Goal	• *Go-To-Goal* • *Avoid-Obstacles* • *Wander*
Mark	Wander Near Start (Simulation Outset)	• *Wander* • *Stay Near Start* • *Avoid-Obstacle* • *Off Robots*
	Mark Mill Around (Below Flock Threshold)	• *Wander* • *Avoid-Obstacle* • *Off Robots*
	Mark Flock (Above Flock Threshold)	• *Lek Behavior* • *Wander* • *Avoid-Obstacle* • *Off Robots*
Shill	Wander Near Start (Simulation Outset)	• *Wander* • *Stay Near Start* • *Avoid-Obstacle* • *Off Robots*
	Shill Flock (Leader Signaled)	• *Follow Leader* • *Lek Behavior* • *Wander* • *Avoid-Obstacle* • *Off Robots*



(Leader)



(Mark)



(Shill)

MOVING ROBOT TEAMS: "PUSH" APPROACHES

- Herding (Fear-based, predator-like agent, depend on selfish herd)
  - Goats/Sheep/Ducks/Cows - Farming
    - Vaughan, Sumpter et al. (1998, 2000)
      - "Experiments in Automatic Flock Control"
      - "Robot sheepdog project achieves automatic flock control"
  - Birds – Avoiding Strikes at Airport
    - Paranjape et al. (2018)
    - "Robotic herding of a flock of birds using an unmanned aerial vehicle"
- Capuchin Monkeys
  - Pushing conspecifics away from food source (specific point)
    - Wheeler (2009)
    - "Monkeys crying wolf?"

https://commons.wikimedia.org/wiki/File:B-6543_(7788321152).jpg

https://commons.wikimedia.org/wiki/File:Capuchin_Costa_Rica.jpg

https://commons.wikimedia.org/wiki/File:Karjus.jpg

Push Approach

- A repulsing agent (orange) tries to move a group from one location to another. Green agents have reached a threshold, which causes them to flee from this agent. Yellow agents haven’t reached this threshold.

- A shill agent (a leader proxy, brown) helps to efficiently move the group.

Push Approach: Robotarium results

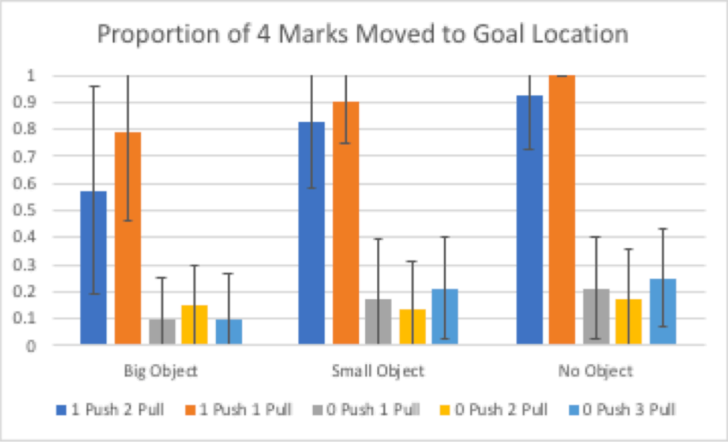
Push : 1 leader, 1 shill, no obstacles

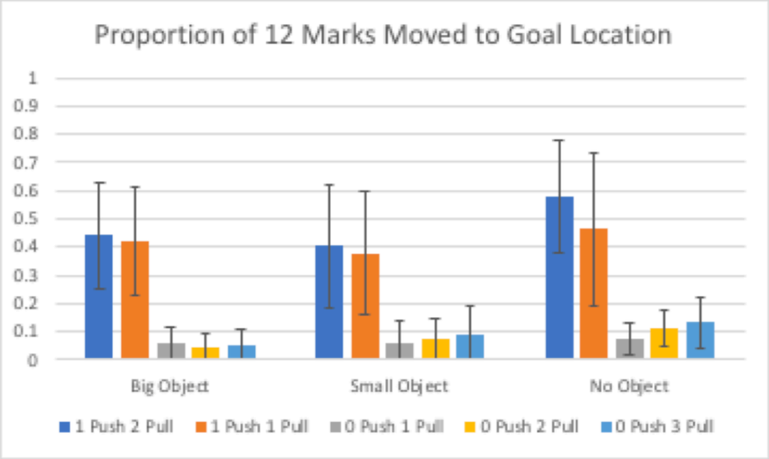
Push:1 leader, 2 shills, no obstacles

HERDING EFFICIENCY

Environment	Proportion of 4 Marks Moved to Goal Location				
	1 Pushing / 2 Pulling Agents Mean/Median (Standard Deviation) (n = 20)	*1 Pushing / 1 Pulling Agents Mean/Median (Standard Deviation) (n = 20)*	*1 Pulling Agent Mean/Median (Standard Deviation) (n = 20)*	*2 Pulling Agents Mean/Median (Standard Deviation) (n = 20)*	*3 Pulling Agents Mean/Median (Standard Deviation) (n = 20)*
Big Object	.575/.75 (.381)	.7875/1.0 (.327)	.1/0 (.150)	.15/.25 (.150)	.1/0 (.170)
Small Object	.825/1.0 (.245)	.9/1.0 (.150)	.175/0 (.216)	.138/0 (.172)	.213/.25 (.186)
No Object	.925/1.0 (.200)	1.0/1.0 (0.0)	.213/.25 (.186)	.175/.25 (.183)	.25/.25 (.181)

Environment	Proportion of 12 Marks Moved to Goal Location				
	1 Pushing / 2 Pulling Agents Mean/Median (Standard Deviation) (n = 20)	*1 Pushing / 1 Pulling Agents Mean/Median (Standard Deviation) (n = 20)*	*1 Pulling Agent Mean/Median (Standard Deviation) (n = 20)*	*2 Pulling Agents Mean/Median (Standard Deviation) (n = 20)*	*3 Pulling Agents Mean/Median (Standard Deviation) (n = 20)*
Big Object	.442/.458 (.188)	.421/.50 (.192)	.058/.083 (.055)	.046/.042 (.050)	.054/.083 (.056)
Small Object	.404/.417 (.217)	.379/.333 (.220)	.063/.042 (.076)	.075/.083 (.071)	.092/.083 (.10)
No Object	.579/.625 (.196)	.463/.583 (.268)	.075/.083 (.060)	.113/.083 (.062)	.133/.125 (.091)





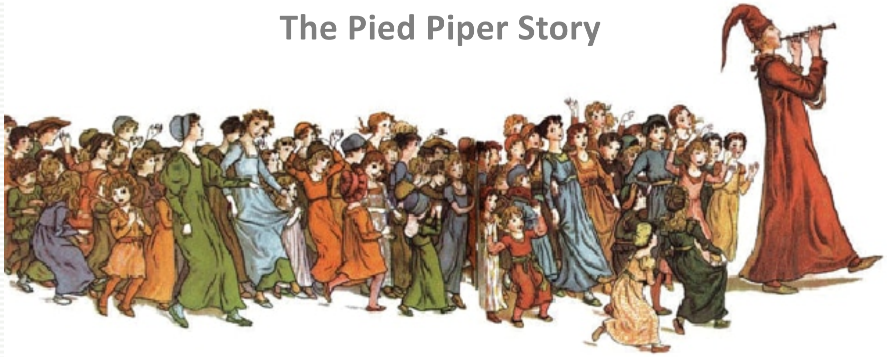
MISDIRECTING ROBOT TEAMS: A “PULL” APPROACH

Objective: Move a team of robots from one location to another.

Solution: Exploit the fact that agents in a group follow the movements of their neighbors.

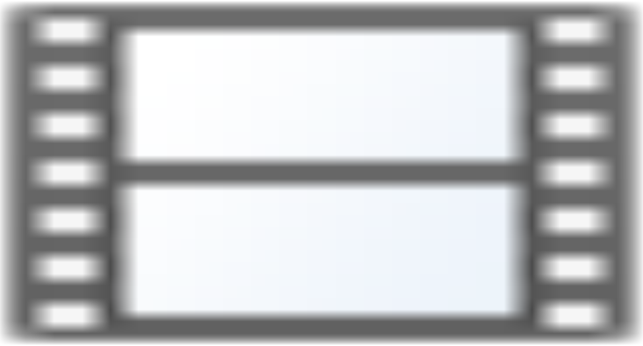
The Judas Goat

The Pied Piper Story

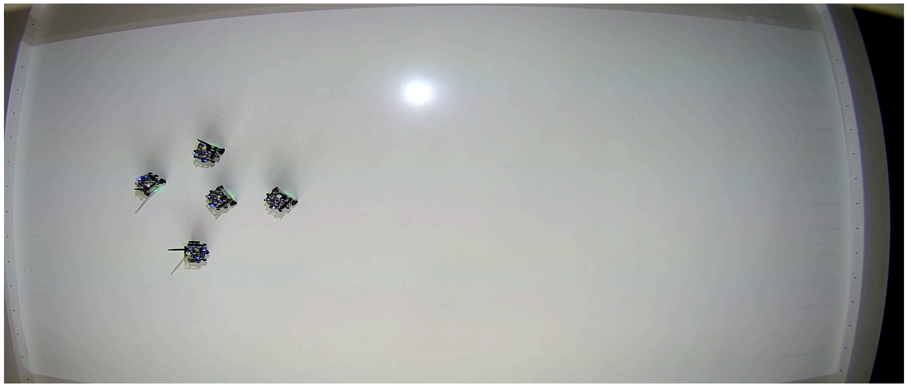


CREATING THE NEXT®

Pull Approach: Robotarium results
1 shill, a leader and 3 other agents

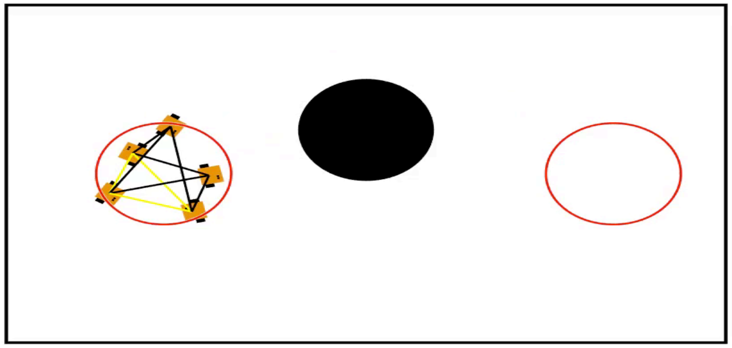


Pull : No obstacle, group does not split



Pull: Obstacle, group does not split

Pull : Obstacle, group splits



Simulation: Leaders and shills are connected only by black lines, agents have yellow (light) lines among themselves

DISCUSSION

- ***When teams of marks are "naive" (their thresholds for flocking are universally low), shills are not necessary to successfully misdirect them.***
- ***When teams of marks contain agents with higher flocking thresholds, a leader alone is often not able to successfully misdirect them.***
- ***The weight of a shill's lekking behavior must be low enough to prevent it from dominating the follow the leader behavior. It cannot follow too closely.***
- ***If the deceptive team is going to function effectively, shill agents must be able to view the leader agent throughout the deception or the shill agent must have knowledge of the goal location.***

Misdirection and Counter-Misdirection in Robot Teams

- Misdirection:
  - Mislead agents to the wrong locations that may be traps or other remote locations to gain advantage over them.
- Counter-misdirection:
  - Goal: stop the misdirection process or negate its effects.
  - Two main components: misdirection **detection**, misdirection **stoppage**.

Counter-misdirection in Robot Teams: Overview

Following the previous work, we developed a novel counter-misdirection approach for behavior-based multi-robot teams by deploying a new type of agent: counter-misdirection agents (**CMAs**).

Three different groups:

- **Mark group**: threshold-based marks.
- **Misdirection team**: a leader with multiple shills.
- **Counter-misdirection team:** a team of counter-misdirection agents (**CMAs**).

Counter-Misdirection Agent (CMA)

Goal: stop misdirected marks from reaching the goal location.

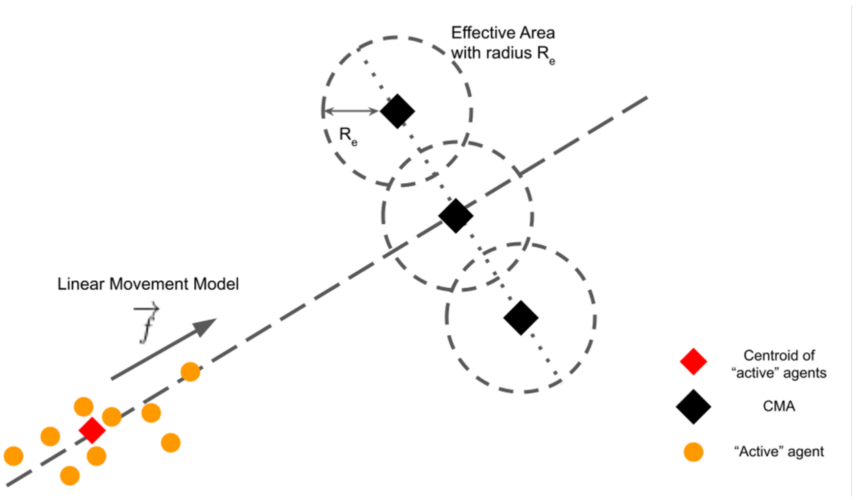
Challenges:

- Cannot identify shills and their leader from marks.
- No knowledge on where the goal location is.

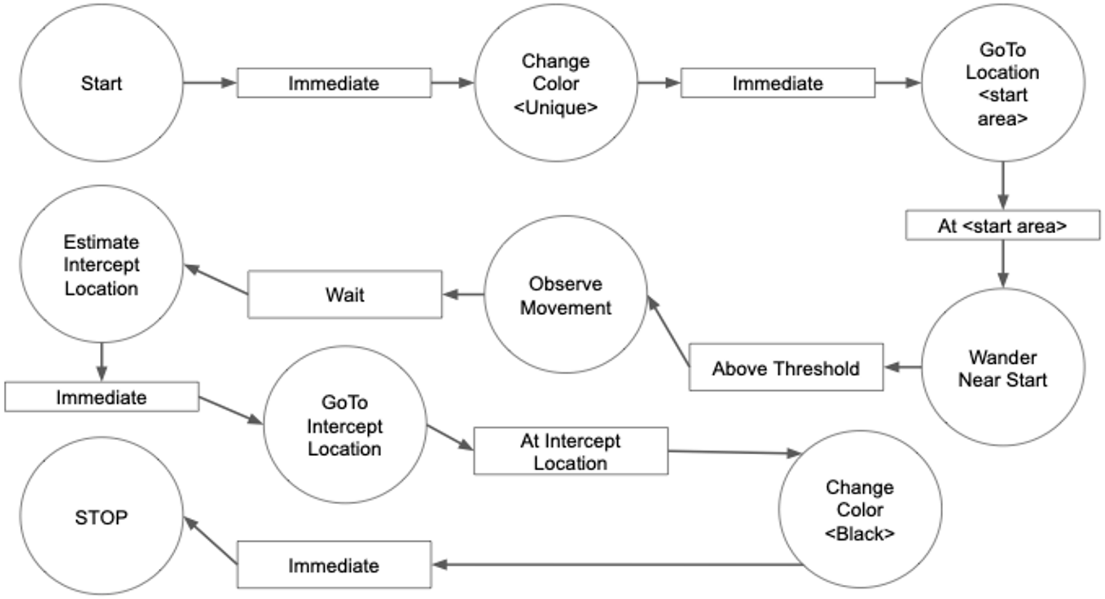
How:

1. Detect misdirection process.
2. Estimate intercept location.
3. Create a repulsive field.

Multiple CMAs can form a “barrier” collectively without explicit communication.

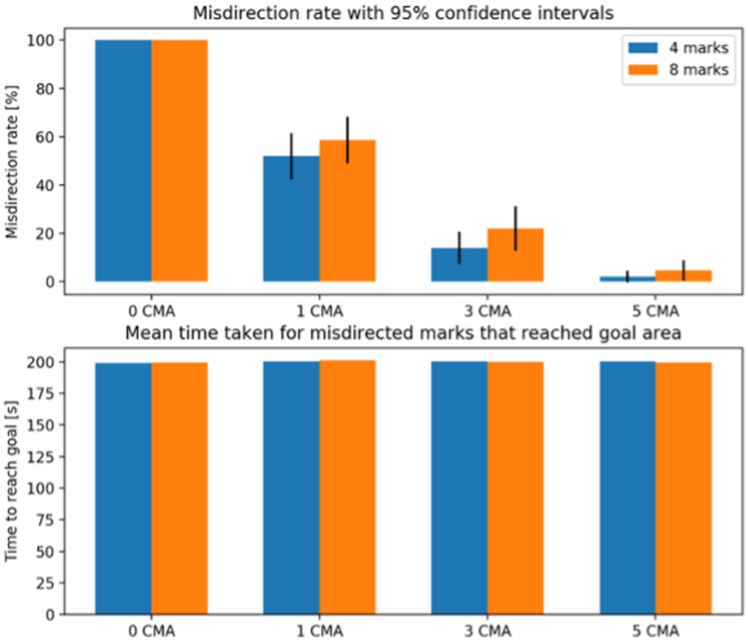


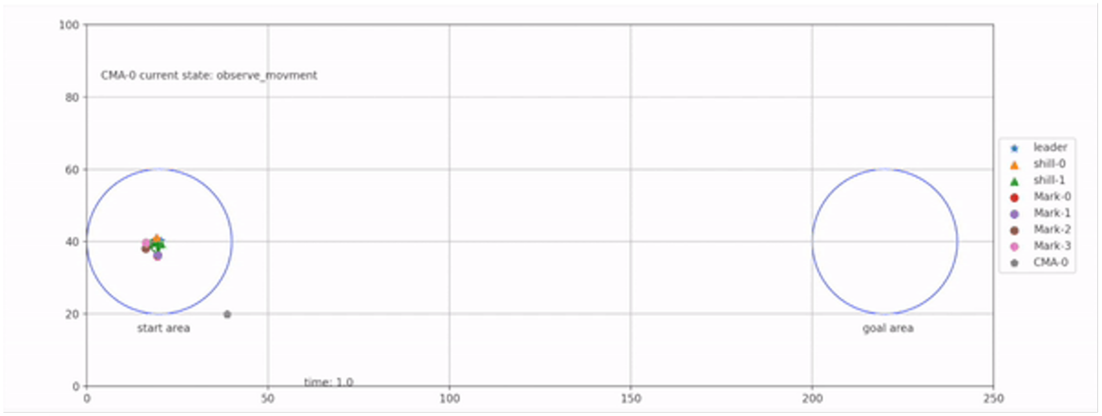
CMA behaviors

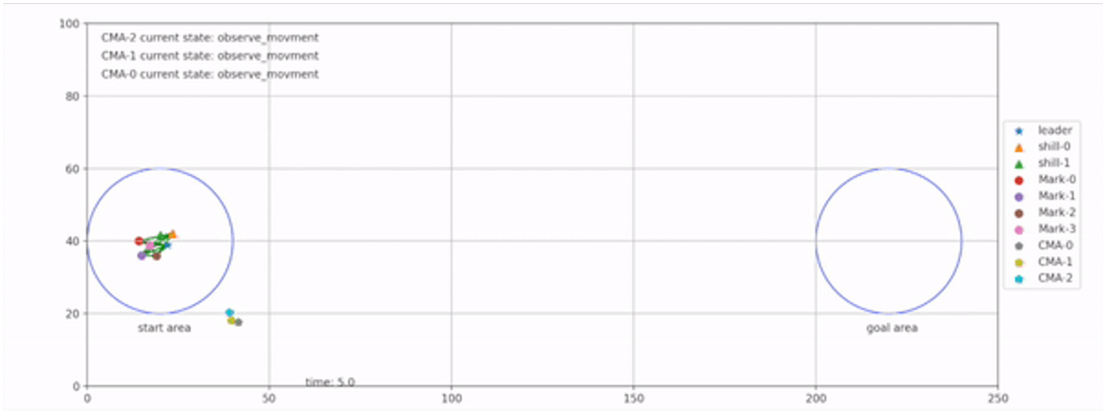


Behaviors Assemblage	**Composing Behaviors**
Wander Near Start (Simulation Outset)	*Wander, Stay-Near-Start, Avoid-Obstacle, Off-Robots.*
Observe Movement	*Stop, Observe*
Estimate Intercept Location	*Stop, Estimate-Location*
Goto Intercept Location	*Goto-Intercept, Color-Pushed-Back, Avoid-Obstacle, Wander.*

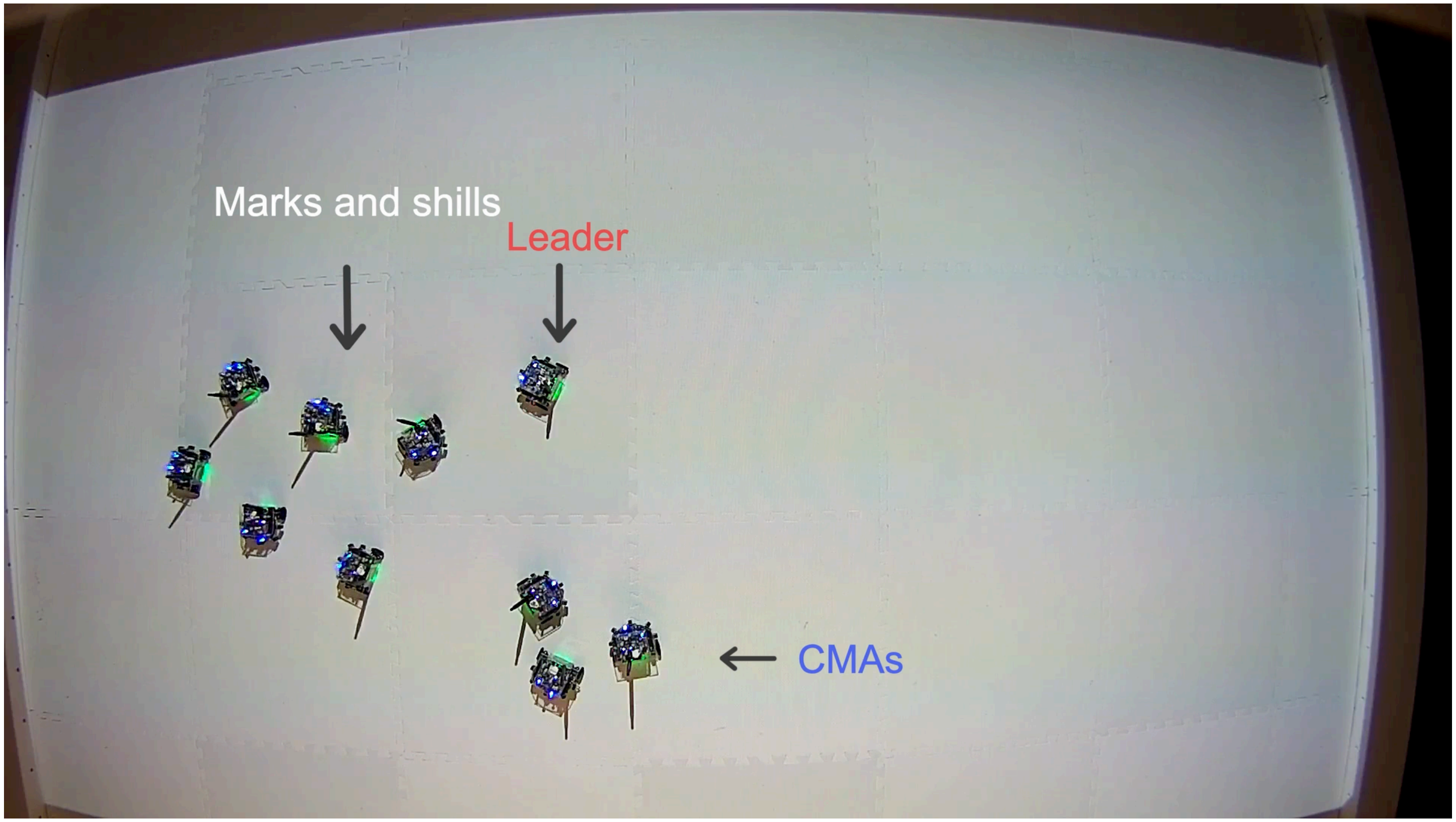
Simulation Results







Validation on Physical Robots



Discussion

- As more CMAs (counter misdirection agents) are deployed, this counter misdirection approach can lower the misdirection rate substantially.
- The counter-misdirection approach is suitable for variable numbers of marks and different inter-robot distances.
- The leader's random movements significantly affect the efficiency of the counter-misdirection strategy.

Conclusion

- Misdirection and counter-misdirection have significant potential in the field of mobile robotics, especially for multi-robot systems. However, there has been little study on robotic counter-misdirection to date.

- We developed a simple and effective behavior-based counter-misdirection approach for multi-robot teams using a team of counter-misdirection agents (CMAs).

Ethical Considerations for Deceptive Robots

- One might question the intent behind creating deceptive as it is entirely possible that the tools could conceivably be used for nefarious purposes.

- How does one ensure that it is only used in an appropriate context?

- Is there an inherent right, whereby humans should not be lied to?

- Kantian theory clearly indicates that lying is fundamentally wrong.

- But from a utilitarian perspective there may be times where deception has societal value, even apart from the military, with the goal of enhancing that individual's survival:
  - calming down a panicking individual in a search and rescue operation
  - in the management of patients with dementia

- In this case, even from a deontological perspective, the intention is good, let alone from a utilitarian consequentialist measure.

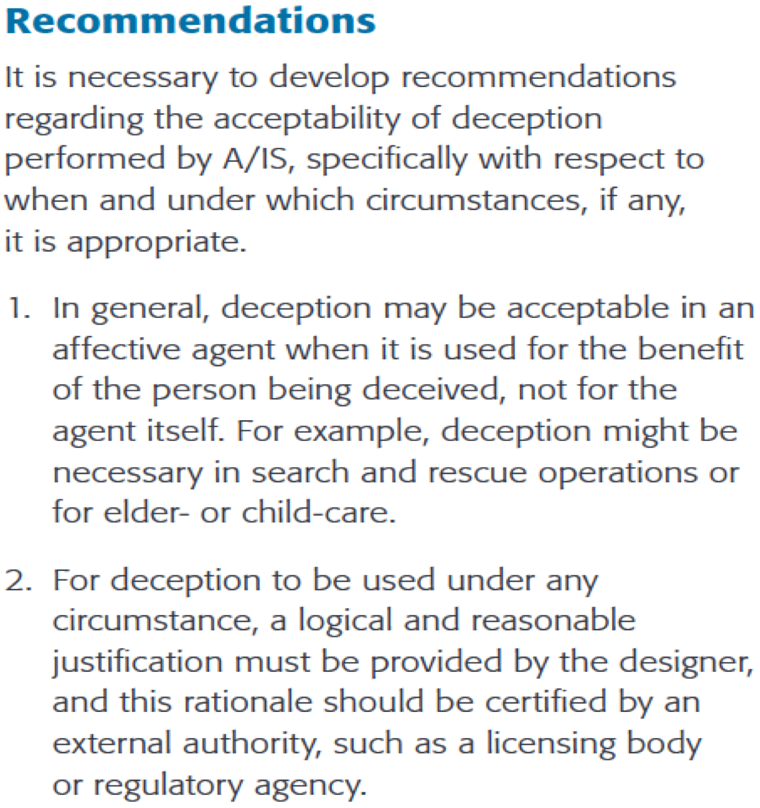
- But does that warrant allowing a robot to possess such a capacity?

IEEE Deception Guidelines

ETHICALLY ALIGNED DESIGN

First Edition

A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems



Recommendations

It is necessary to develop recommendations regarding the acceptability of deception performed by A/IS, specifically with respect to when and under which circumstances, if any, it is appropriate.

1. In general, deception may be acceptable in an affective agent when it is used for the benefit of the person being deceived, not for the agent itself. For example, deception might be necessary in search and rescue operations or for elder- or child-care.
2. For deception to be used under any circumstance, a logical and reasonable justification must be provided by the designer, and this rationale should be certified by an external authority, such as a licensing body or regulatory agency.

For further information . . .

Mobile Robot Laboratory Web site

- http://www.cc.gatech.edu/ai/robot-lab/
- Multiple relevant papers available

IEEE Global Initiative for Ethical Considerations in AI and Autonomous Systems
https://standards.ieee.org/develop/indconn/ec/autonomous_systems.html

IEEE Social Implications of Technology Society
http://www.ieeessit.org/

CS 4002 – Robots and Society Course (Georgia Tech)
http://www.cc.gatech.edu/classes/AY2021/cs4002a_spring/