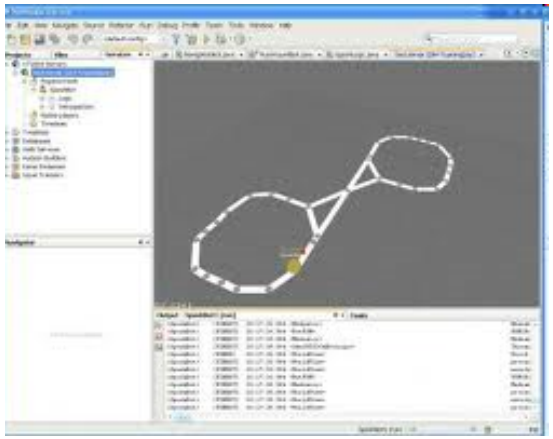
# Botprize 2010

Jacob Schrum, Igor Karpov, and Risto Miikkulainen
{schrum2,ikarpov,risto}@cs.utexas.edu

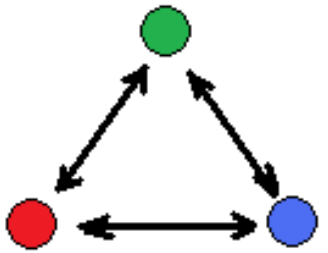

# Unreal Tournament 2004

- Commercial videogame
- First Person Shooter genre
- Play vs. humans and bots
- Programming API: Pogamut
  - Gamebots message protocol

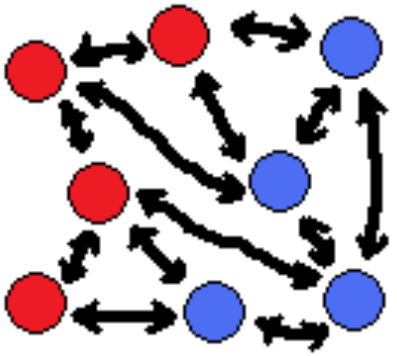


# Turing Test For Bots

- Can humans tell bots from other humans?
- Botprize 2008, 2009
  - In style of traditional Turing Test
    - Bot vs. Judge vs. Confederate
    - 3 individuals per match



- Botprize 2010
  - Judging game
    - Multiple humans vs. multiple bots
    - All humans are judges and players



# Judging Game

- Special judging gun
  - Replaces the Link Gun
- Primary and alternate fire look identical
  - Primary fire against bots
  - Alternate fire against humans
- Correctly judge opponent:
  - Kills opponent, +10 frags
- Incorrectly judge opponent:
  - Shooter dies, -10 frags
- Bots can use this gun!

# Competition

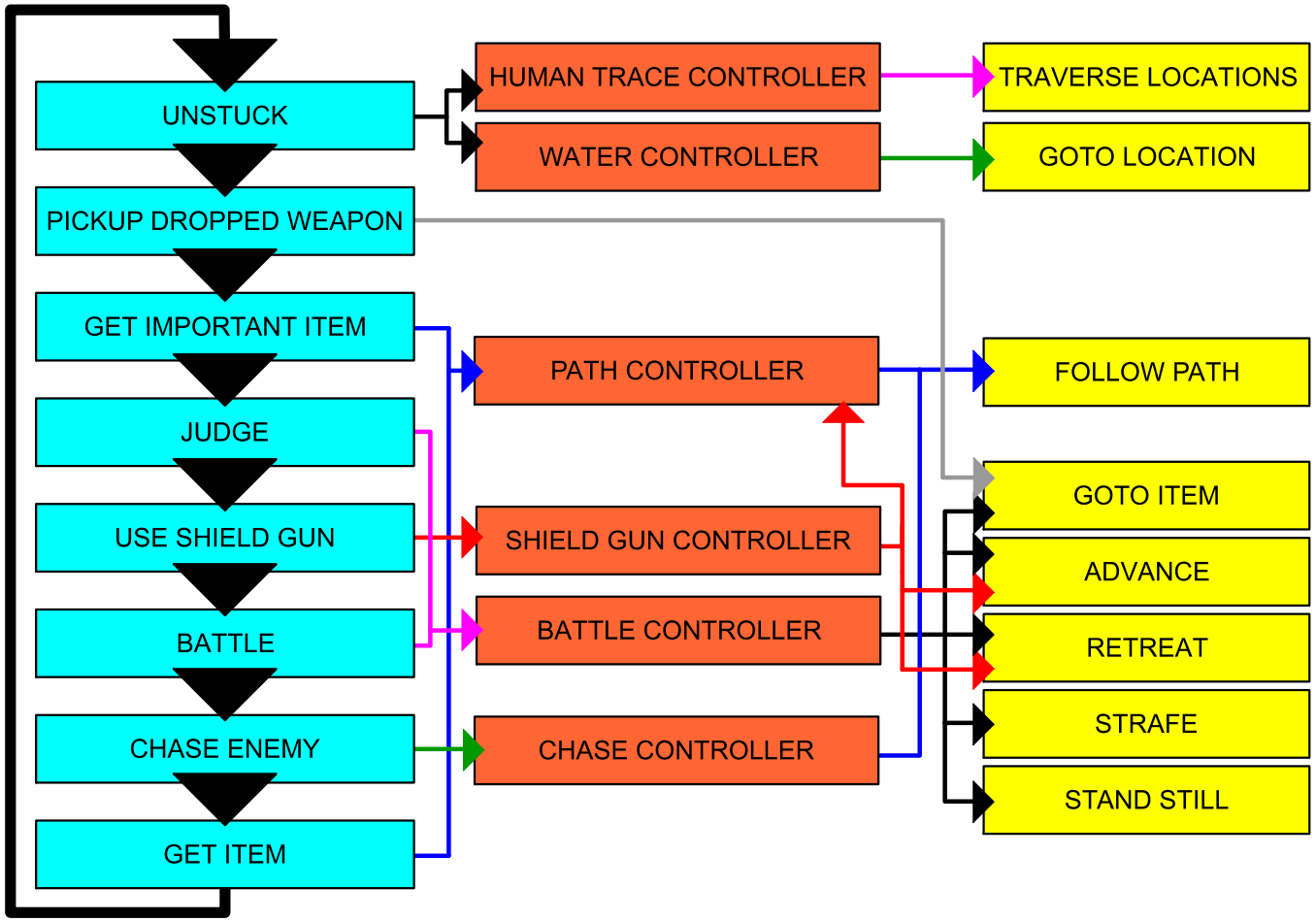
- 3 sessions, 1 hour each
- 4 matches per session, 15 minutes each
- 5 competing bots, 6-7 judges, and 1-2 native UT bots per session
- 3 large custom levels used:



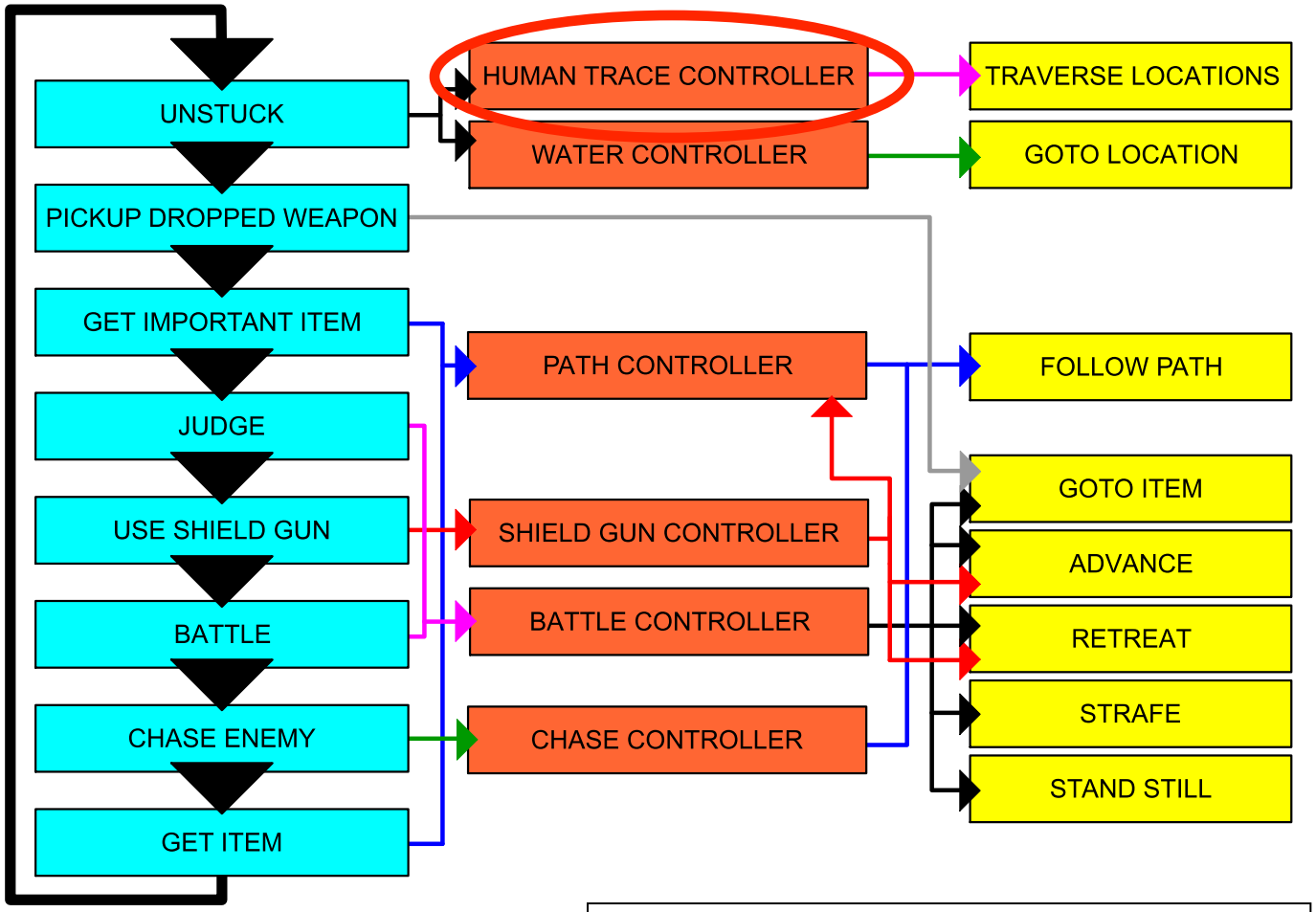
# Our Bot (Demo)


14:49
20
0
0
Coleman
NOW VIEWING
Loren
You got the Flak Cannon.
85
13

# Agent Architecture

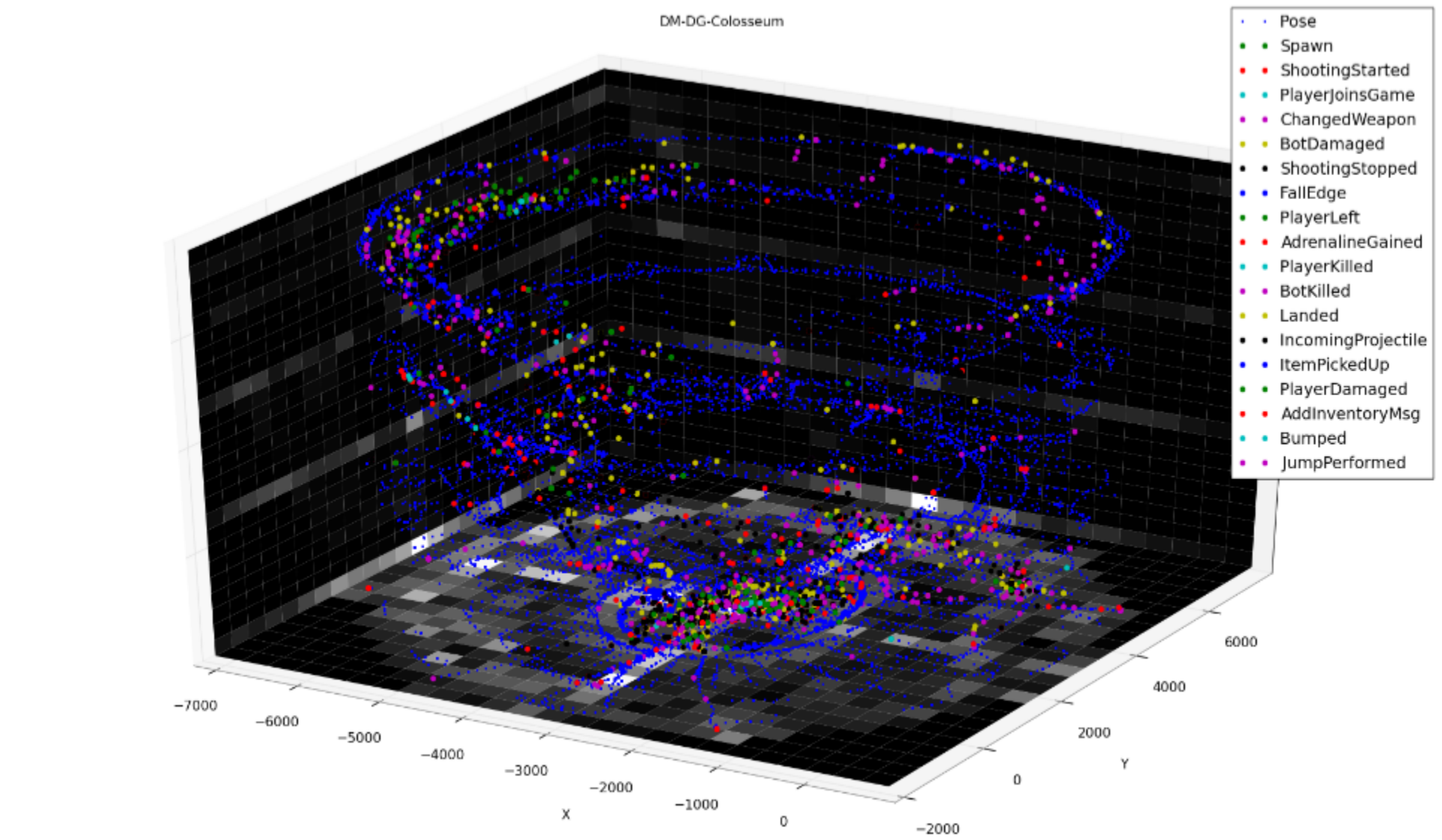


# Agent Architecture




Use human traces to get unstuck

# Human Trace Data



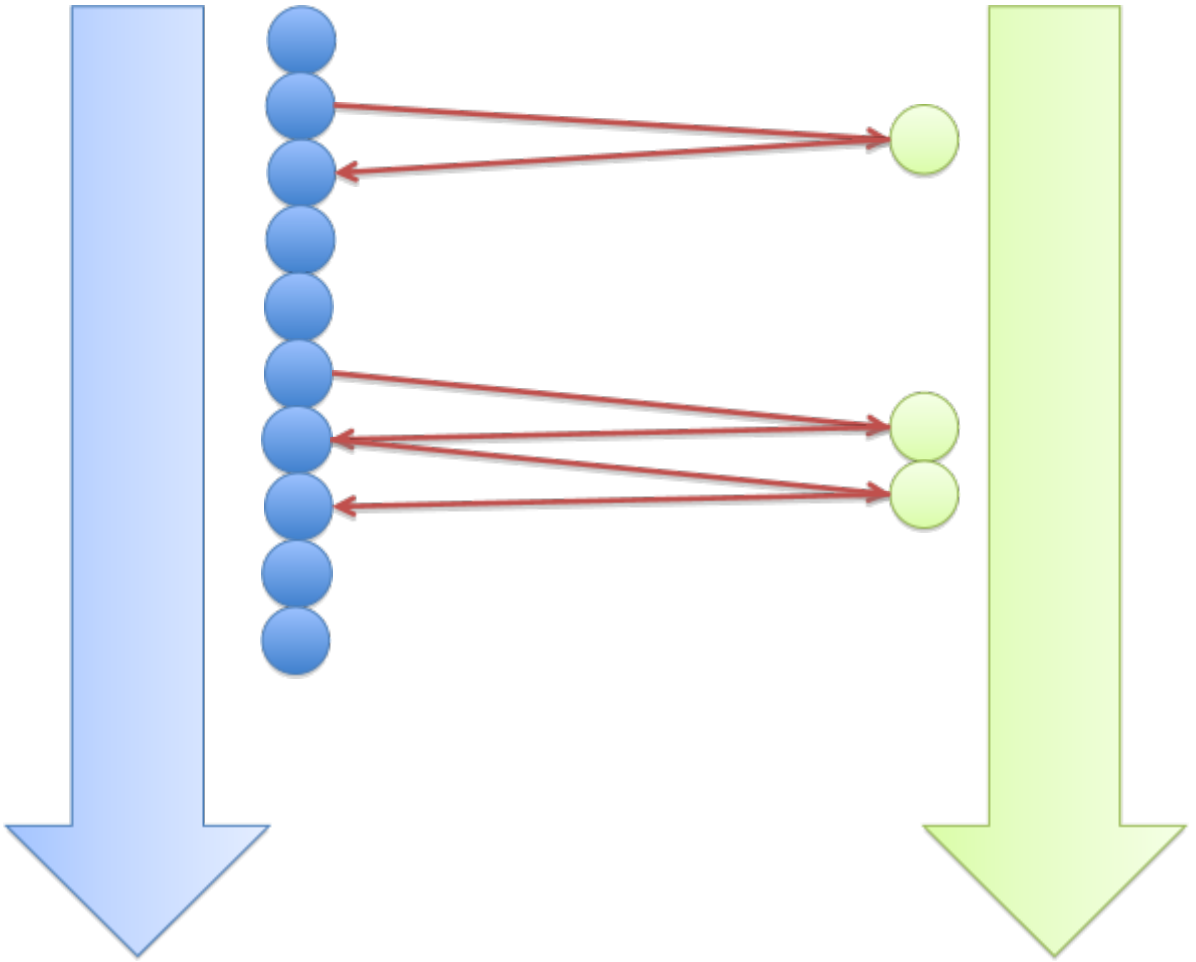
# Replaying Human Experience

- Record
  - Player pose
    - position, orientation, velocity and acceleration
  - Events
    - fall, damage, weapons, items, jumps, etc.
- Index for lookup by
  - Region of origin
  - Future events
- Replay (when stuck)
  - Short relative path from origin

# What is in the Database?

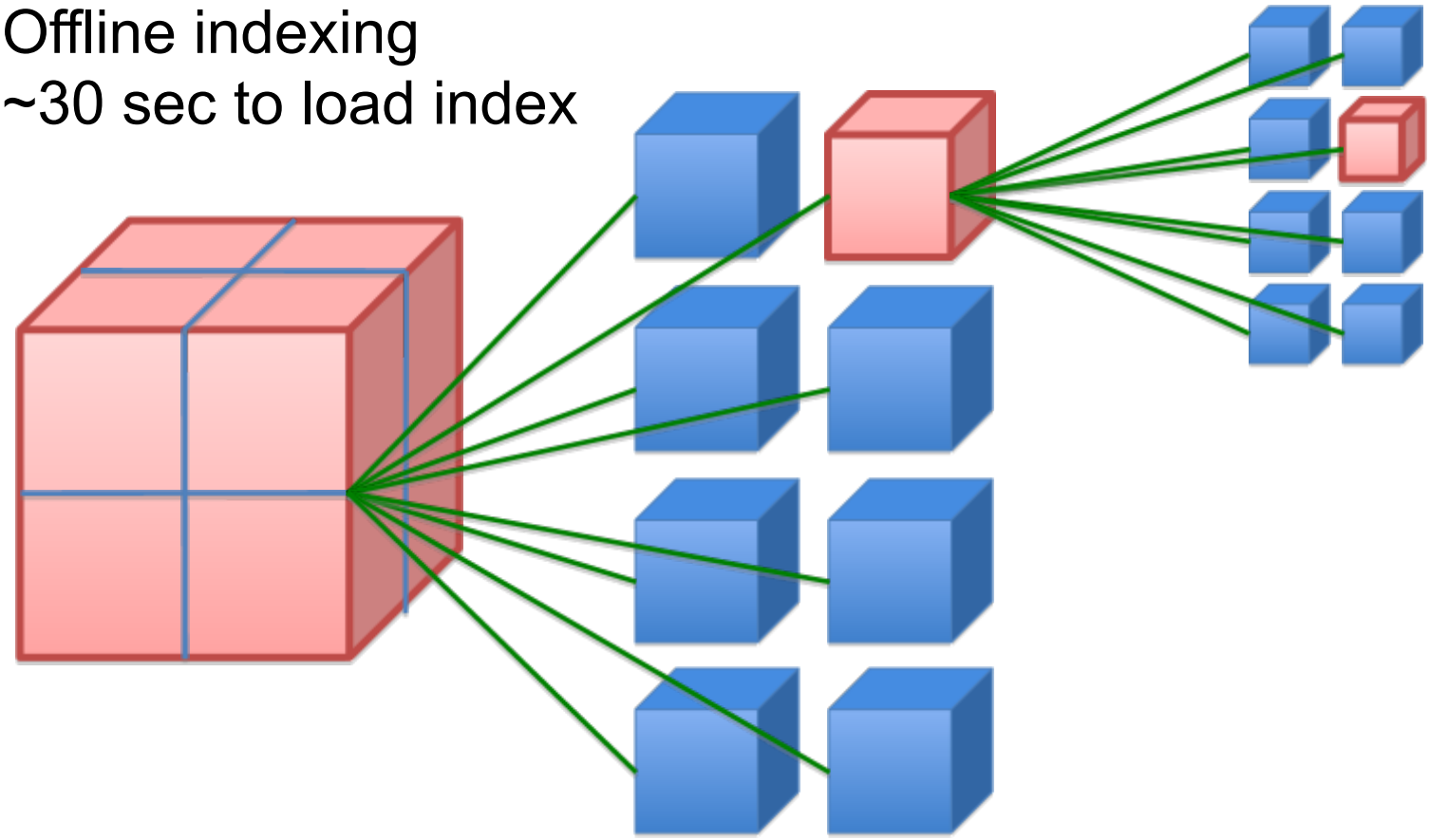
t, x, y, z, rx, ry, rz, vx, vy, vz, ax, ay, az

t, e

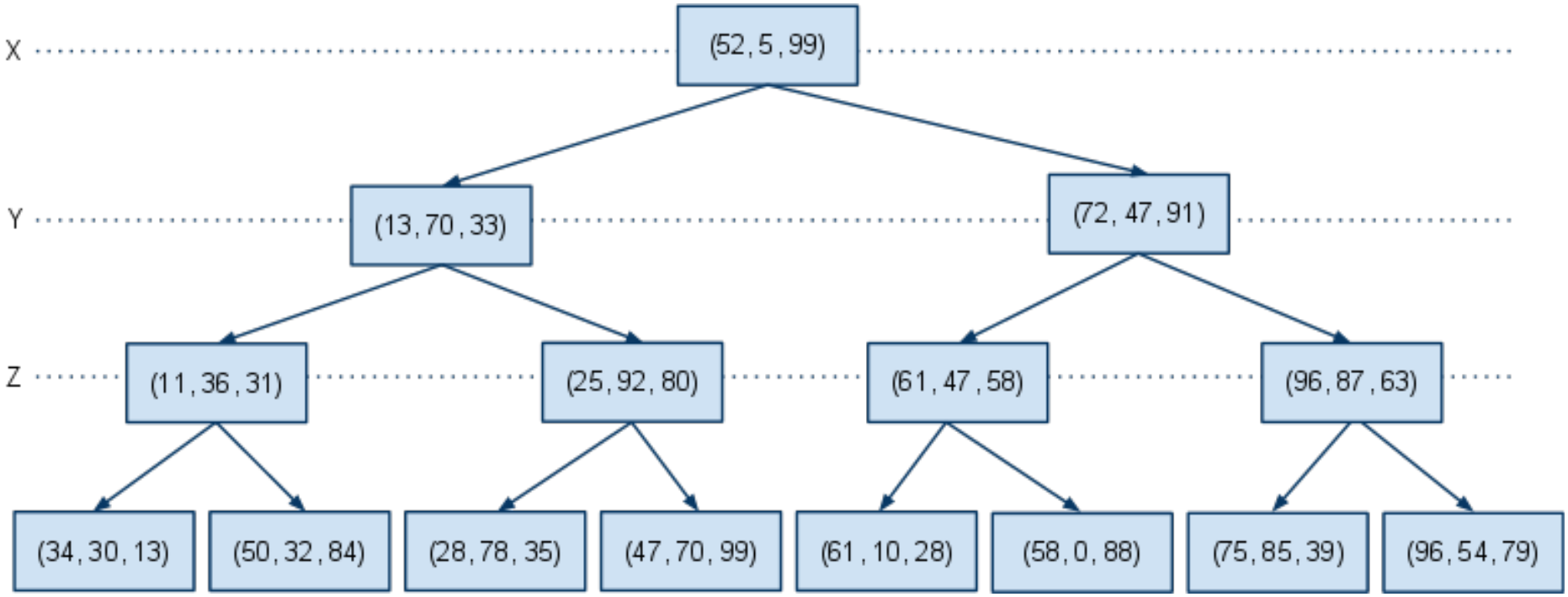


# Indexing the Data: Octrees

- O(log N) lookup
- Offline indexing
- ~30 sec to load index



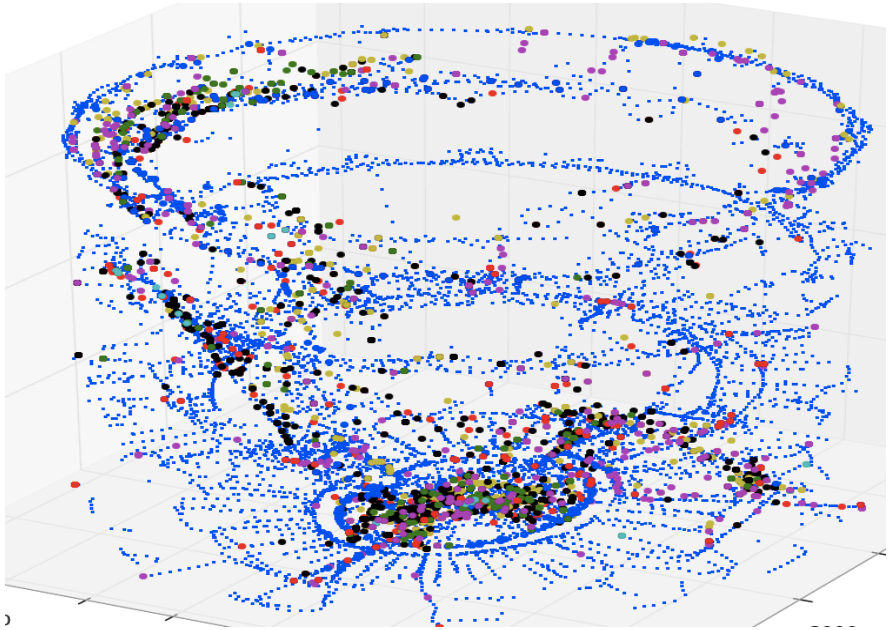
# Indexing the Data: KD-Trees

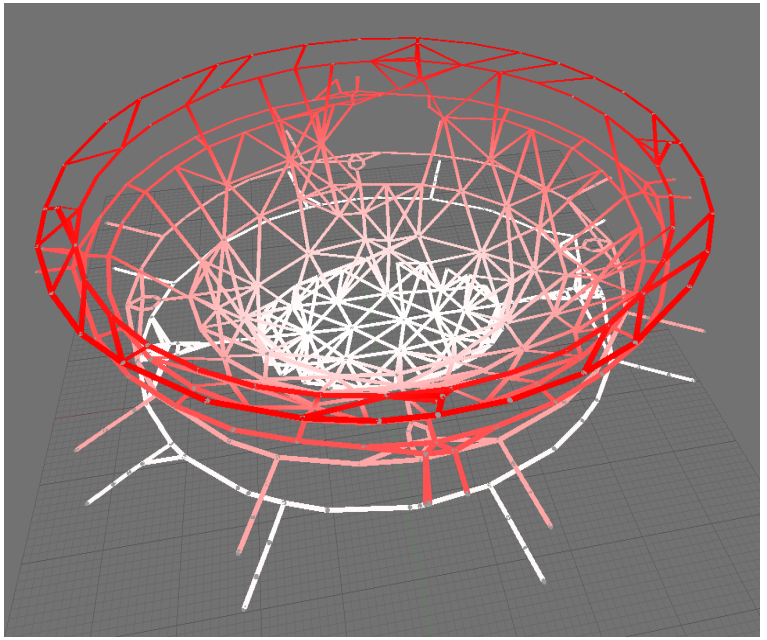



- O(log N) nearest neighbor search
- Offline indexing
- ~30 sec to load index

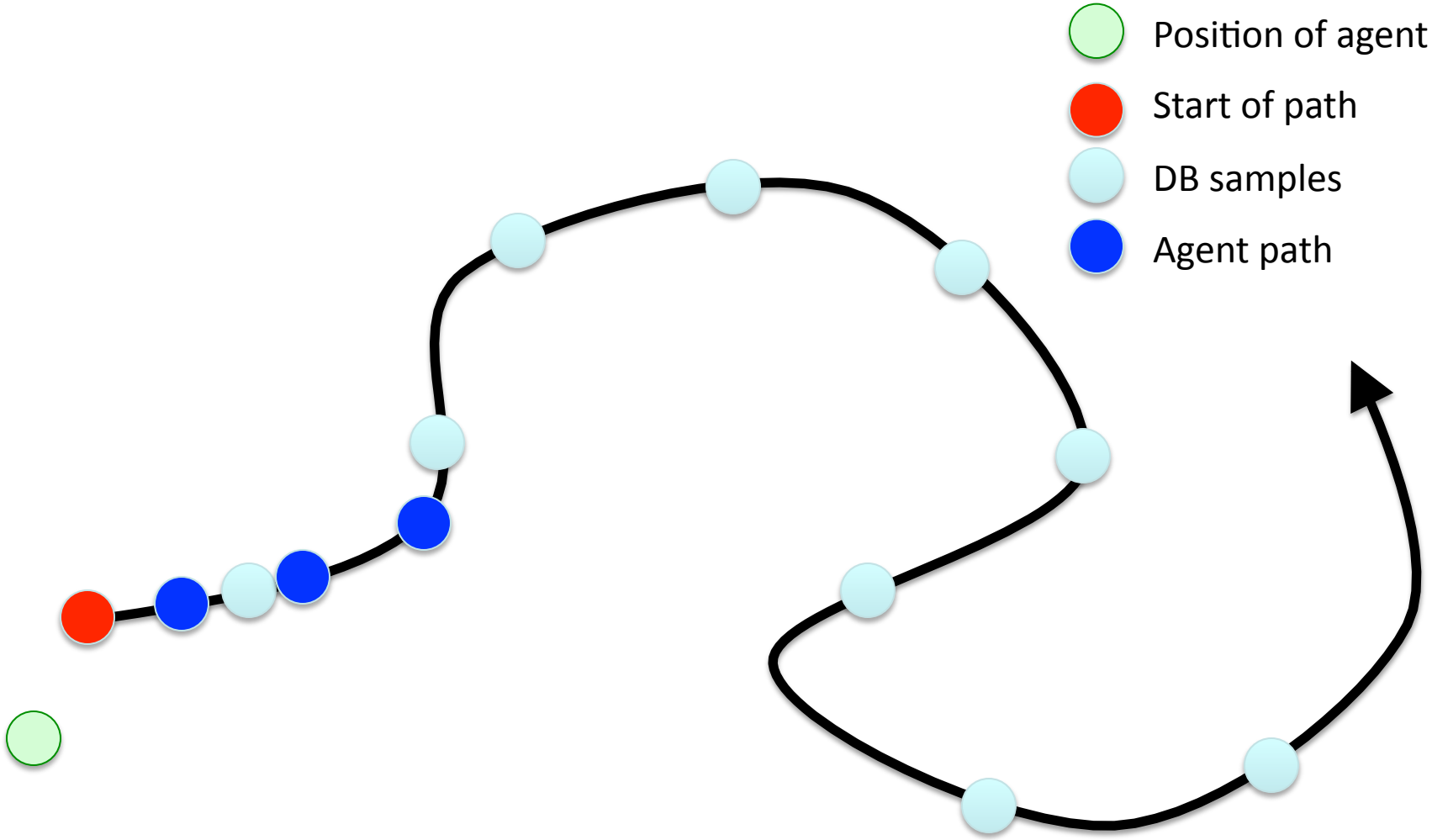
# Indexing the Data: Navpoint Graph

- Each level has graph of navpoints (under 300)
- Store navpoints in a KD-tree (quick)
- For each point in human DB, find closest navpoint (offline)
- Retrieve all points within navpoint's Voronoi region
- From here, use random or nearest selection (online)

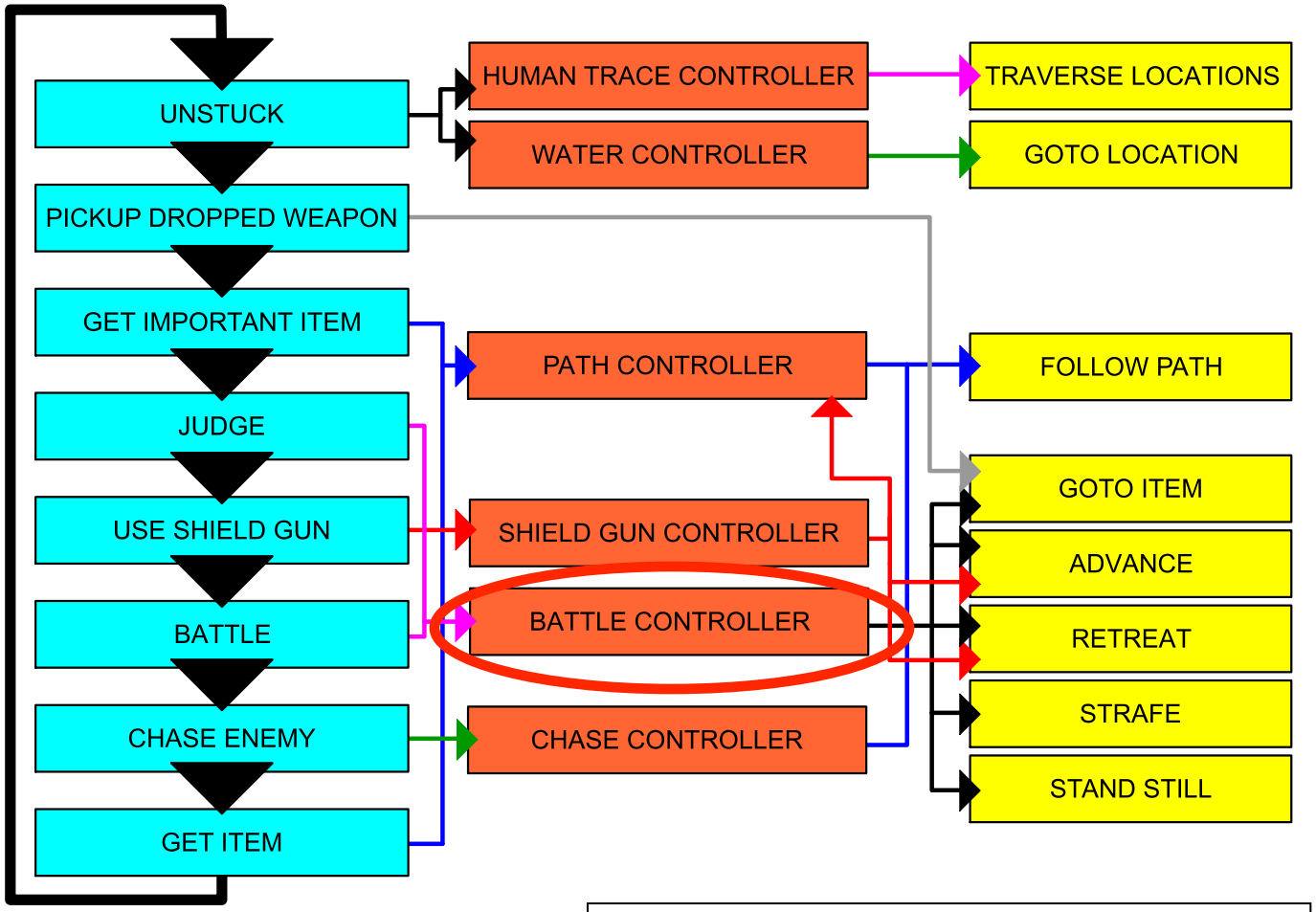




# Generating the path

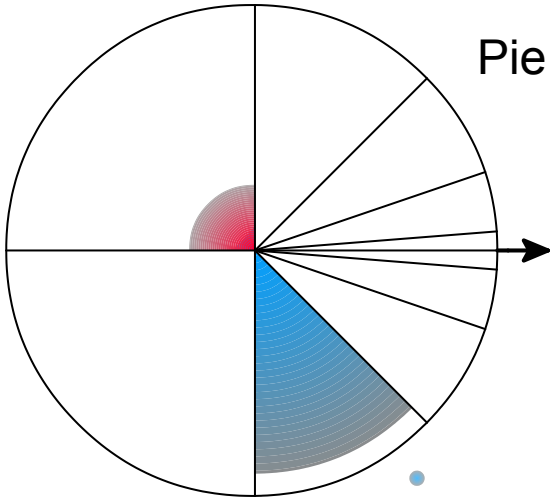


# Agent Architecture




Evolve controller that fights well

# Battle Controller Inputs



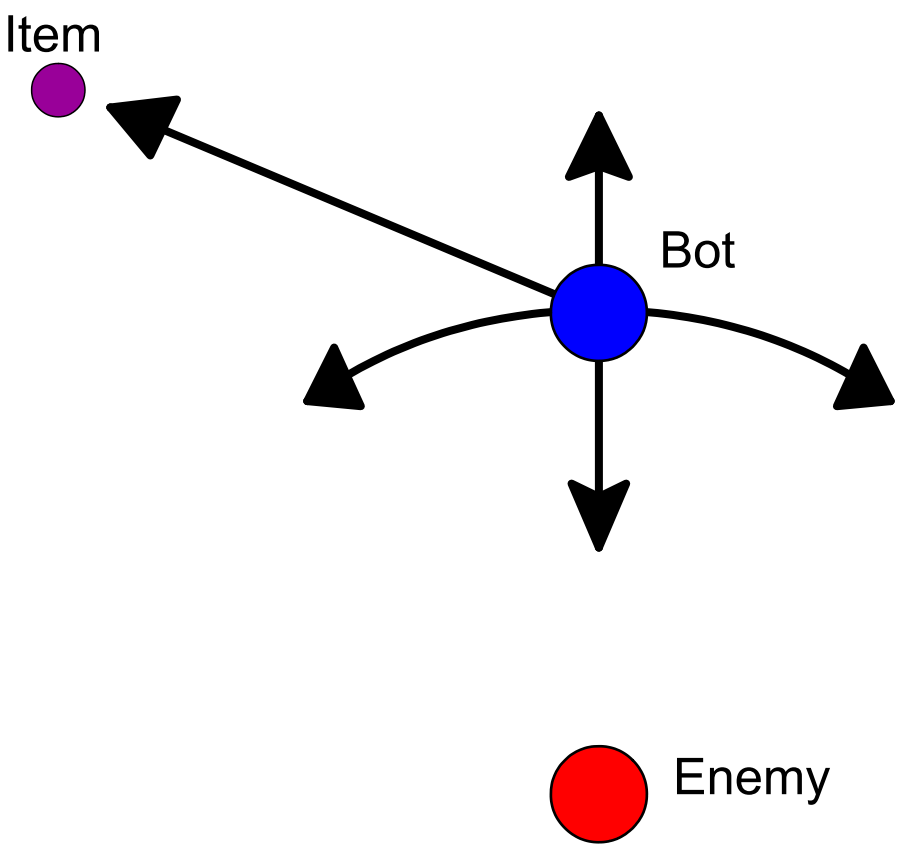
Pie slice sensors for enemies

Ray traces for walls/level geometry


Other misc. sensors for current weapon properties, nearby item properties, etc.

# Battle Controller Outputs

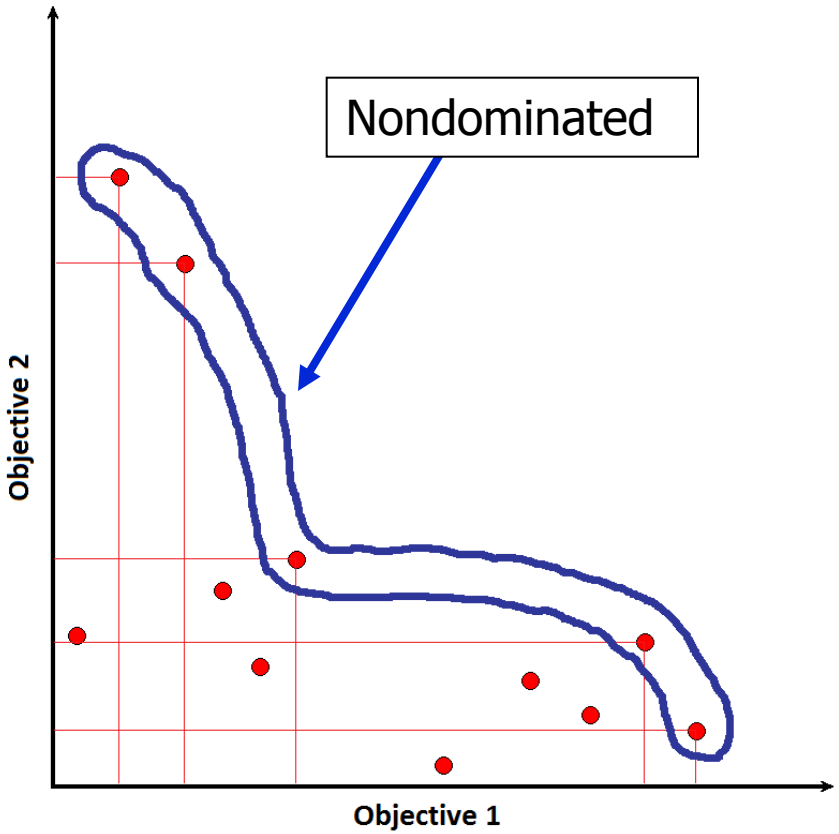
- 6 movement outputs
  - Advance
  - Retreat
  - Strafe left
  - Strafe right
  - Move to nearest item
  - Stand still
- 3 additional outputs
  - Shoot?
  - Alternate fire?
  - Jump?



# Mutiobjective Optimization

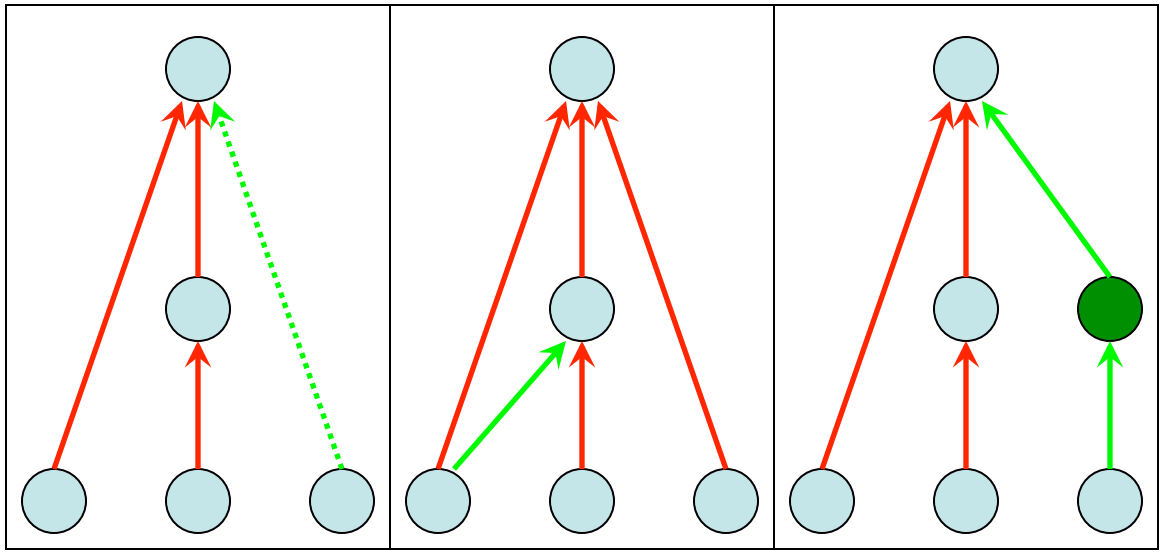
- Pareto dominance: $\vec{v} \succ \vec{u}$ iff
  - $\forall i \in \{1,\ldots,n\}: v_i \geq u_i$
  - $\exists i \in \{1,\ldots,n\}: v_i > u_i$
- Assumes maximization
- Want nondominated points
- NSGA-II used in this work

- What to evolve?
  - NNs as control policies



# Constructive Neuroevolution

- Genetic Algorithms + Neural Networks
- Build structure incrementally (complexification)
- Good at generating control policies
- Three basic mutations (no crossover used)

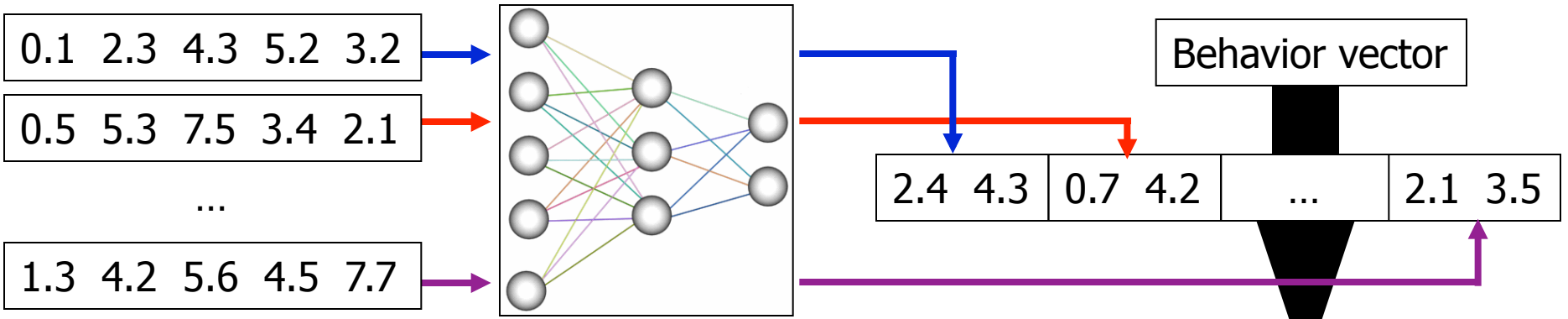


# Objectives

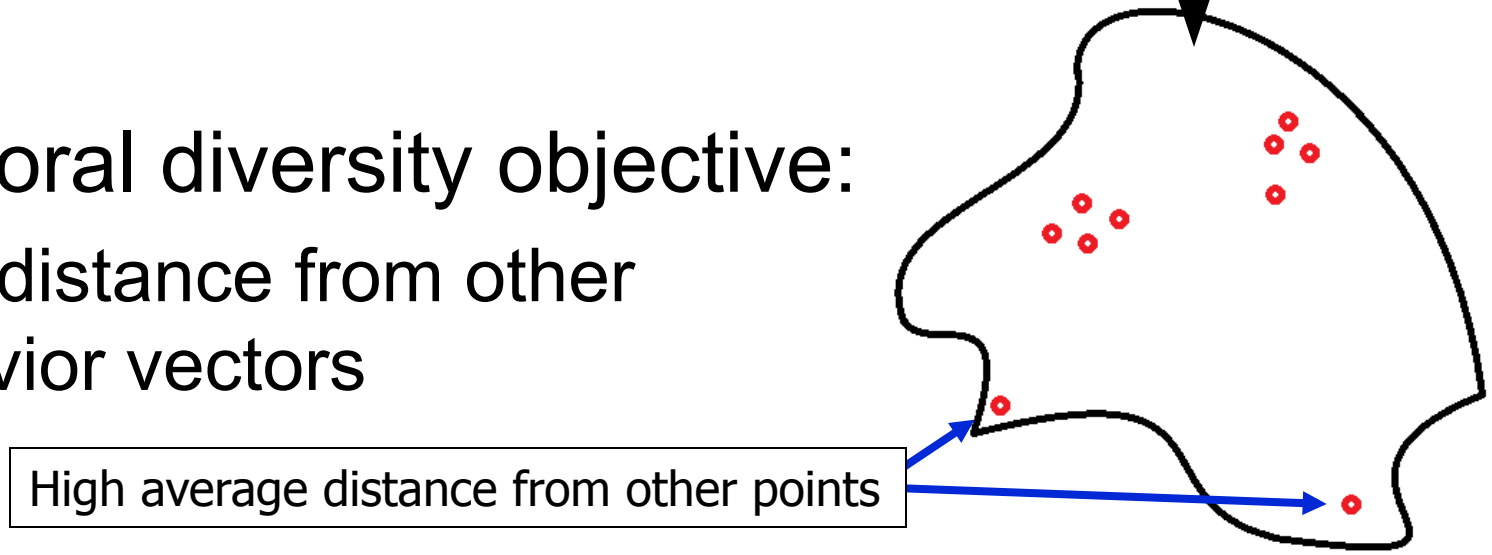
- Damage dealt
- Accuracy
- Damage received (negative)
- Geometry collisions (negative)
- Actor collisions (negative)
- Behavior diversity


# Behavioral Diversity

- Behavior vector:
  - Given input vectors, concatenate outputs




- Behavioral diversity objective:
  - AVG distance from other behavior vectors



# Botprize 2010 Results

| Bot Name | Humanness % | Judging Accuracy % |
|---|---|---|
| Conscious-Robots | 31.82% | N/A |
| **UT^2** | **27.27%** | **45.74 %** |
| ICE-2010 | 23.33% | N/A |
| Discordia | 17.78% | 54.83 % |
| w00t | 9.30% | 53.84 % |

Also, native UT bot had humanness of 35.3982%.

Native bot and winner did not judge at all.

| Human Player | Humanness % |
|---|---|
| Mads Frost | 80.00% |
| Simon and Will Lucas | 59.09% |
| Ben Weber | 48.28% |
| Nicola Beume | 47.06% |
| Minh Tran | 42.31% |
| Gordon Calleja | 38.10% |
| Mike Preuss | 35.48% |

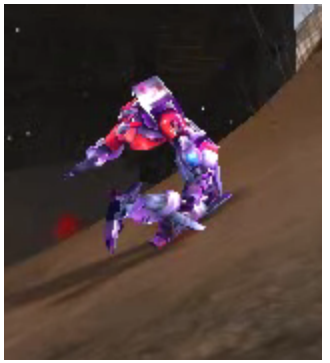
| Human Player | Judging Accuracy % |
|---|---|
| Gordon Calleja | 78.57% |
| Nicola Beume | 67.21% |
| Minh Tran | 64.29% |
| Ben Weber | 64.08% |
| Mike Preuss | 59.70% |
| Mads Frost | 57.69% |
| Simon and Will Lucas | 54.79% |

# Insights

- Judging for the bot is not important
  - Better to not judge then do it wrong
- Different judges, different expectations
  - Combat, dodging, jumping, etc.
  - Perhaps mimicry of opponents would help
- Human judges expect reaction/response
  - Shoot and miss, run away and wait
- Human judges like to observe
  - From roof tops, through sniper scope

# Why Did We Lose?

- Specific weapon issues (sniping)
- Some tricks in our judging behavior
- Problems with following
- Perhaps perceived as too skilled
- Still got stuck a few times
- Some weird firing glitches



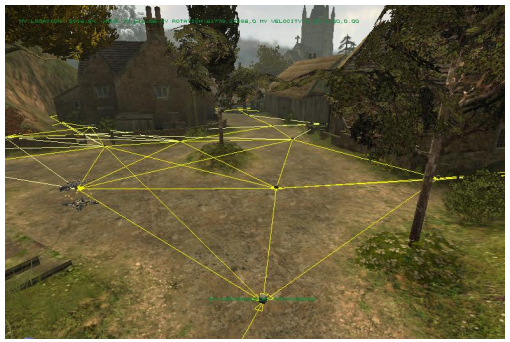
- Mostly minutiae!

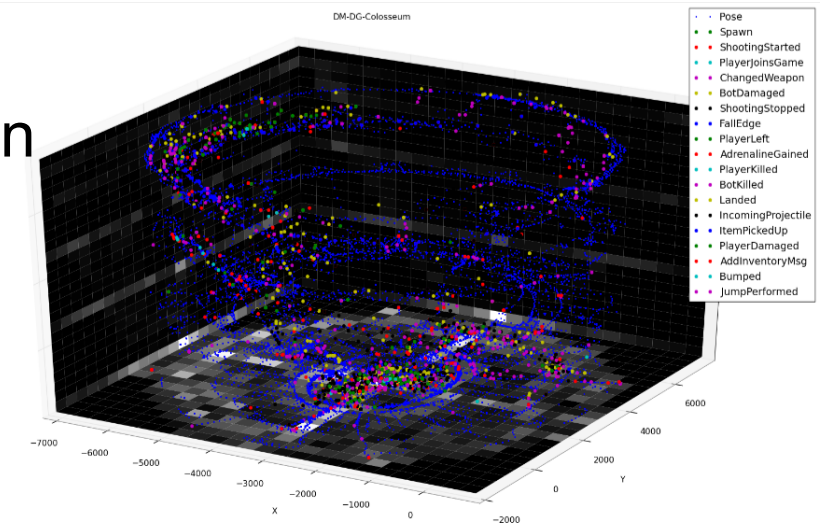
# Believable Bots

- Will be writing a book chapter on our bot
- Experiments evaluating bot performance
  - Human Trace Controller gets bot unstuck
  - Evolved Battle Controller good at combat

# Human Trace Experiments

- Do the human traces help the agent get unstuck?
  - Time stuck with full system, w/o filtering, w/random paths
- Does the performance improve with more data?
  - Time stuck with 1, 2, 3 players, etc.
- Does the indexing method make a difference?
  - Random vs. nearest starting point
  - Constrained by Octree region
  - Constrained by Navpoint region





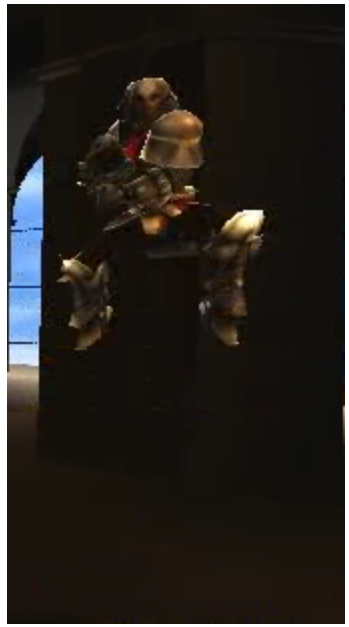
# Evolution Experiments

- Does evolution improve combat?
  - Bot vs. random combat action selector
- Are all the different actions useful?
  - Usage of each type of movement action
  - Ablation studies
- Importance of weapons
  - Above experiments with limited weapon access




# Future Work


- Human Traces
  - Generalize to unseen levels
  - Induce better navigation graphs
  - Make intelligent decisions about when to jump
  - Use to improve following
  - Supervised learning
- Evolution
  - Different features/input representation
  - Apply to other control modules
  - Apply to selection between modules
  - Reduce reliance on scripted behavior



# Future Work

- Theory of Mind
  - Planned behavior transitions
    - e.g. a chasing bot expects to enter combat mode
  - Mimicry: expectation of similarity
    - Match opponent's level of dodging, aggressiveness, ammo wasting, etc.
    - Establish communication
  - Deliberation
    - Sniping humans don't move as much
    - Better human judges don't make snap decisions



# Questions?



Jacob Schrum
Igor Karpov
Risto Miikkulainen
{schrum2,ikarpov,risto}@cs.utexas.edu

# Botprize 2010 Results

**Most human bots**

| bot name | team | affiliation | humanness % |
|---|---|---|---|
| Conscious-Robots | Raul Arrabales<br>Jorge Muñoz | Carlos III University, Madrid | 31.8182 % |
| UT^2 | Igor Karpov<br>Jacob Schrum<br>Risto Miikulainen | University of Texas, Austin | 27.2727 % |
| ICE-2010 | Akihiro Kojima<br>Daichi Hirono<br>Takumi Sato<br>Seiji Murakami<br>Ruck Thawonmas | Intelligent Computer Entertainment Lab.<br>Ritsumeikan University, Japan | 23.3333 % |
| Discordia | Casey Rosenthal<br>Clare Bates Congdon | University of Southern Maine | 17.7778 % |
| w00t | Daniel Büscher<br>Matthias Gorzellik<br>Jannis Seyfried<br>Björn Witt | Institut für Informatik<br>Albert-Ludwigs Universität<br>Freiburg, Deutschland | 9.3023 % |

**Most human humans**

| player | affiliation | humanness % |
|---|---|---|
| Mads Frost | IT University Copenhagen | 80.0000 % |
| Simon and Will Lucas | University of Essex | 59.0909 % |
| Ben Weber | UC Santa Cruz | 48.2759 % |
| Nicola Beume | TU Dortmund University | 47.0588 % |
| Minh Tran | Edith Cowan University | 42.3077 % |
| Gordon Calleja | IT University Copenhagen | 38.0952 % |
| Mike Preuss | TU Dortmund University | 35.4839 % |

**Best bot judges**

| bot name | accuracy % |
|---|---|
| Discordia | 54.8387 % |
| w00t | 53.8462 % |
| UT^2 | 45.7447 % |

**Best human judges**

| player | accuracy % |
|---|---|
| Gordon Calleja | 78.5714 % |
| Nicola Beume | 67.2131 % |
| Minh Tran | 64.2857 % |
| Ben Weber | 64.0845 % |
| Mike Preuss | 59.7015 % |
| Mads Frost | 57.6923 % |
| Simon and Will Lucas | 54.7945 % |

# Judgment Counts

| UT^2 | total | correct | incorrect | ratio |
|---|---|---|---|---|
| by humans | 33 | 24 | 9 | 0.27 |
| by bots | 4 | 4 | 0 | |
| total | 37 | 28 | 9 | 0.24 |
| **Conscious-R** | **total** | **correct** | **incorrect** | **ratio** |
| by humans | 44 | 30 | 14 | 0.32 |
| by bots | 6 | 3 | 3 | |
| total | 50 | 33 | 17 | 0.34 |
| **Frost** | **total** | **correct** | **incorrect** | **ratio** |
| by humans | 10 | 8 | 8 | 0.8 |
| by bots | 4 | 3 | 3 | |
| total | 14 | 11 | 11 | 0.79 |
| **Swill** | **total** | **correct** | **incorrect** | **ratio** |
| by humans | 22 | 9 | 13 | 0.59 |
| by bots | 9 | 3 | 6 | |
| total | 31 | 12 | 19 | 0.61 |