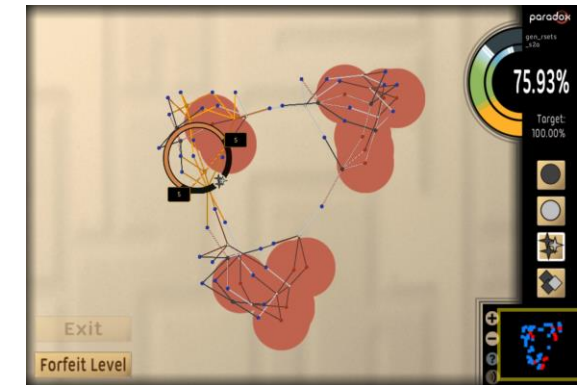
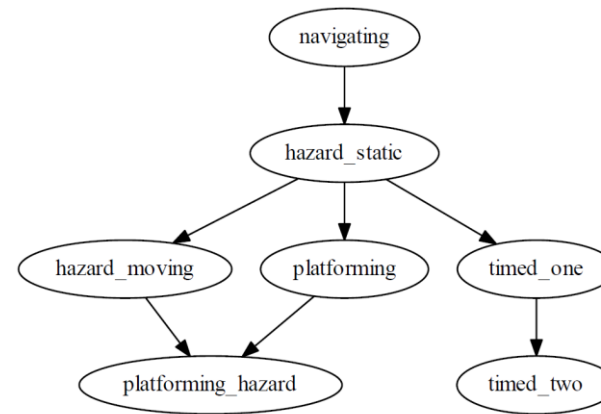
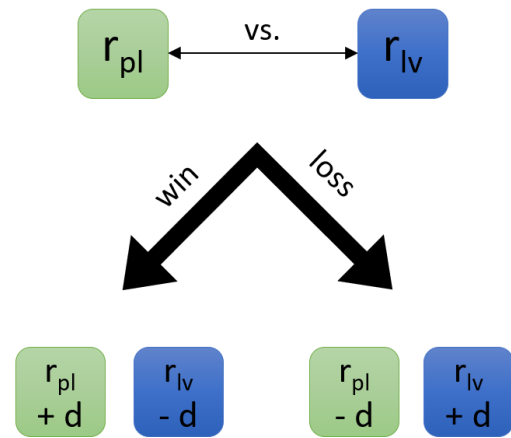


Evaluating and Comparing Skill Chains and Rating Systems for Dynamic Difficulty Adjustment

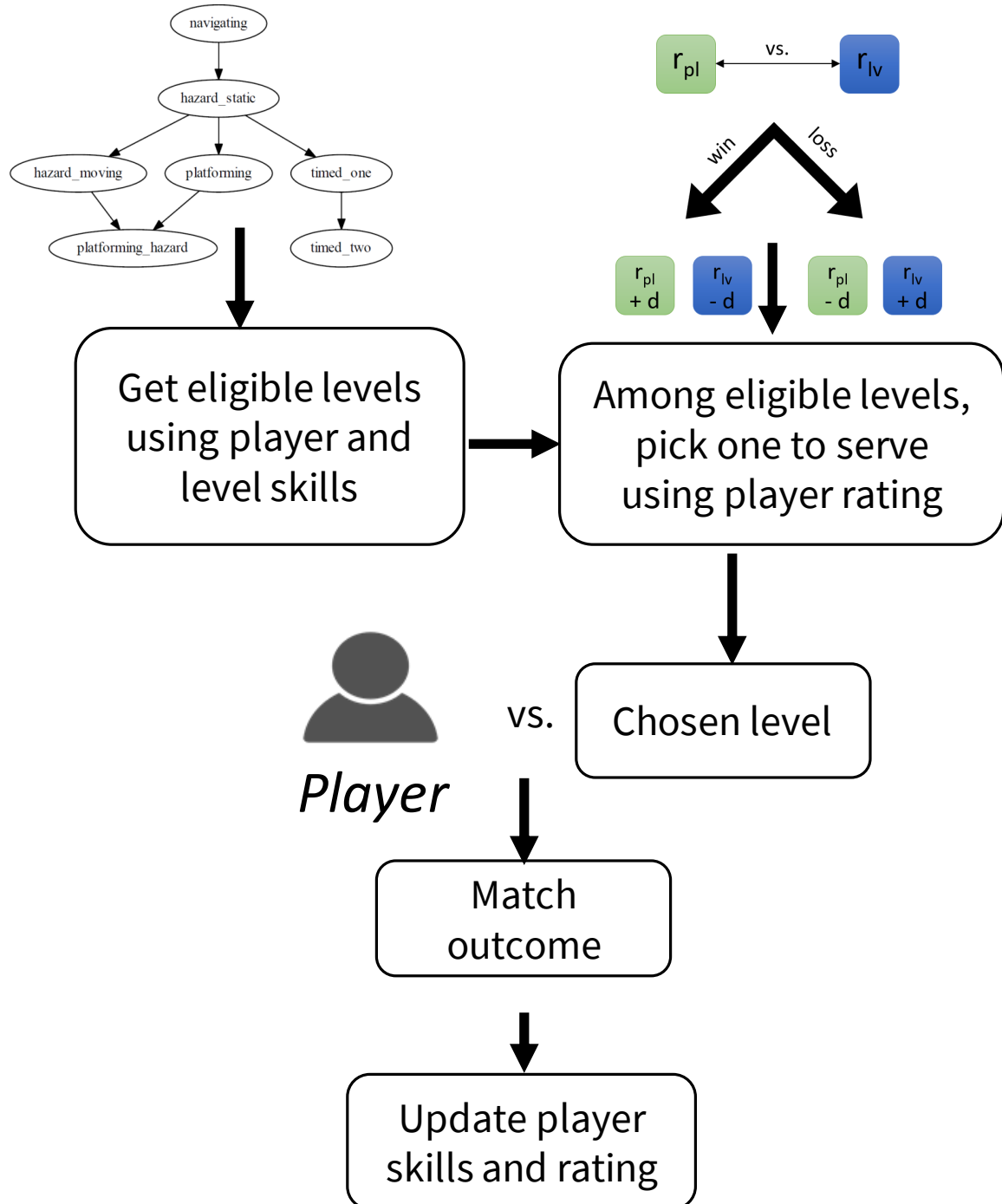
Anurag Sarkar and Seth Cooper
Northeastern University



- **Rating systems** assign ratings to players based on skill and levels based on difficulty
- Dynamic difficulty adjustment (DDA) via player-vs-level matchmaking
- **Skill chains** define the order of player skill acquisition during gameplay
- Used to define level progressions of varying difficulty
- Rating systems and skill chains have both been used for difficulty balancing in **human computation games**
- Existing skill models combining rating systems and skill chains only applied to specific type of HCGs
- Relative pros and cons of using rating systems and skill chains not clear

Model

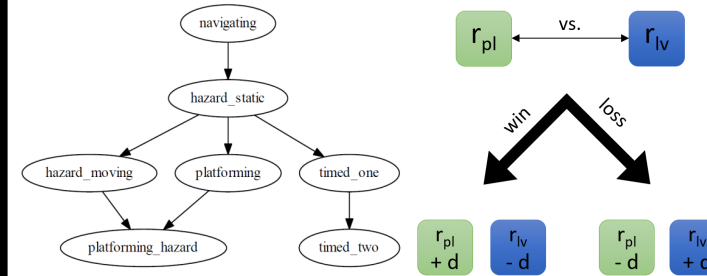
Generalized combined skill model to all HCGs



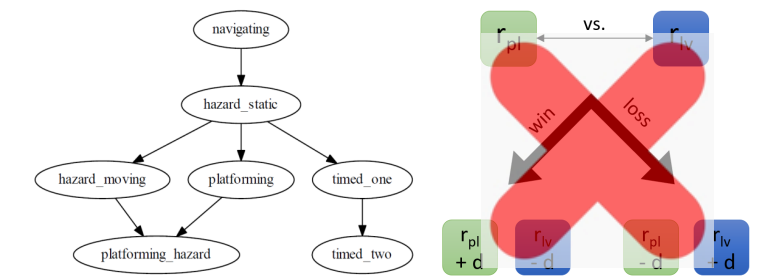
Progressions

Defined progressions to evaluate relative merits of skill chains and rating systems in model

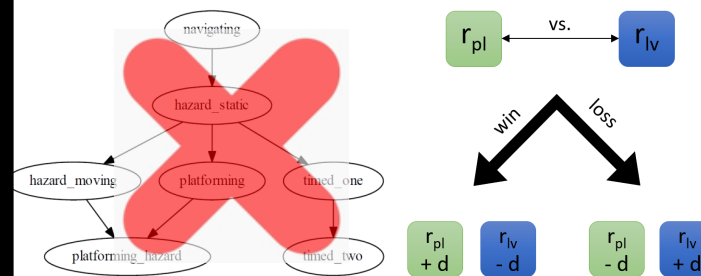
SKILL-RAT



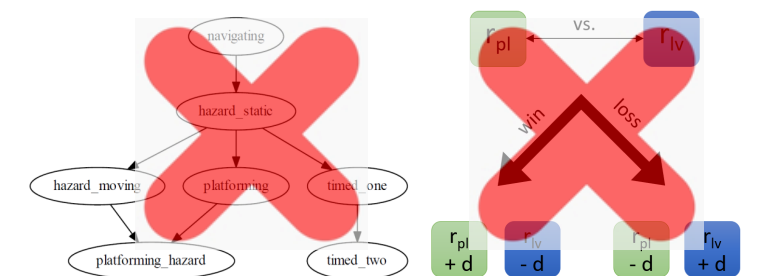
SKILL-ONLY



RAT-ONLY



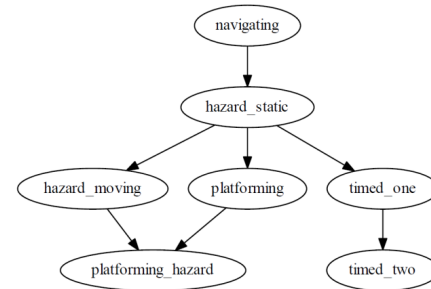
RANDOM



Games

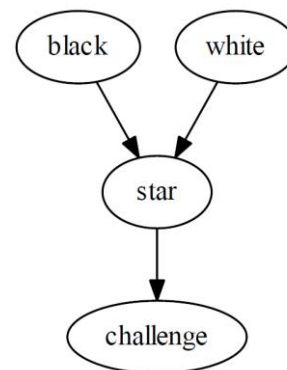
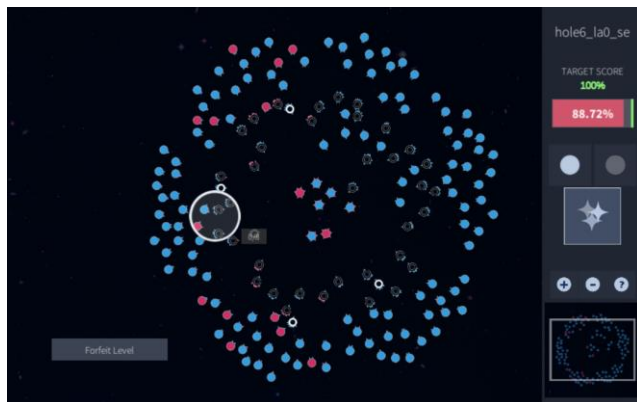
Evaluations using two human computation games

Iowa James



- 2D platformer HCG modeling item collection
- Skills based on running and jump mechanics

Paradox



- 2D puzzle HCG modeling constraint satisfaction
- Skills based on using value-assigning tools

Experiment

- For each game, recruited players using Amazon Mechanical Turk
- Players assigned randomly to one of 4 progressions
- Three broad measures
 - *Quantity and difficulty of completed levels*
 - *Skill acquisition rates*
 - *Failure and completion rates for different types of levels*

Results

Quantity and Difficulty of Completed Levels

| Variable | SKILL_RAT | SKILL_ONLY | RAT_ONLY | RANDOM |
|-----------------------------------|-------------------|-------------------|-------------------|-------------------|
| Play Time ($p = .29$) | 355 | 489 | 419 | 269 |
| Final Player Rating ($p = .19$) | 1406 | 1401 | 1353 | 1358 |
| Max Level Rating† ($p < .001$) | 1669 ^a | 1839 ^b | 1662 ^a | 1517 ^a |
| Levels Completed† ($p < .001$) | 3 ^a | 2 ^b | 3 ^{ab} | 1 ^c |
| Levels Failed ($p = .1$) | 2.5 | 4 | 3 | 4 |
| Max Skillset Size ($p = .14$) | 2 | 2 | 2 | 1 |

Iowa James

| Variable | SKILL_RAT | SKILL_ONLY | RAT_ONLY | RANDOM |
|-----------------------------------|------------------|------------------|------------------|-----------------|
| Play Time ($p = .81$) | 443 | 481 | 466 | 395 |
| Final Player Rating ($p = .09$) | 1069 | 1122 | 1075 | 1395 |
| Max Level Rating† ($p < .001$) | 758 ^a | 758 ^a | 602 ^b | 0 ^b |
| Levels Completed† ($p < .001$) | 3 ^{ab} | 3 ^a | 2 ^b | 0 ^c |
| Levels Failed† ($p = .03$) | 1 ^a | 2 ^{ab} | 4 ^b | 2 ^{ab} |
| Max Skillset Size† ($p < .001$) | 2 ^{ab} | 3 ^a | 2 ^{bce} | 0 ^e |

Paradox

- Variable analysis showing median values
- Using **skill chains** can lead to players completing harder levels
- Adding **ratings system** can lead to players completing more levels

Skill Acquisition Rates

| Skill | SKILL_RAT | SKILL_ONLY | RAT_ONLY | RANDOM |
|--------------------|-----------|------------|----------|--------|
| navigating | 96 | 91 | 97 | 58 |
| hazard_static | 57.9 | 55.7 | 52.9 | 46.2 |
| hazard_moving | 4 | 18.6 | 19.1 | 21.2 |
| timed_one | 34.2 | 13.4 | 20.6 | 19.2 |
| platforming | 38.2 | 19.6 | 17.7 | 21.2 |
| timed_two | 2.6 | 4.1 | 19.1 | 7.7 |
| platforming_hazard | 1.3 | 3.1 | 4.4 | 5.8 |

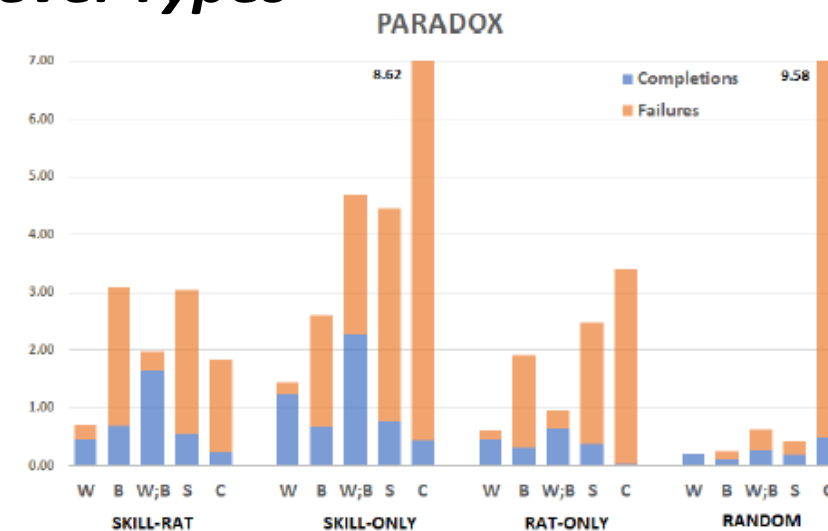
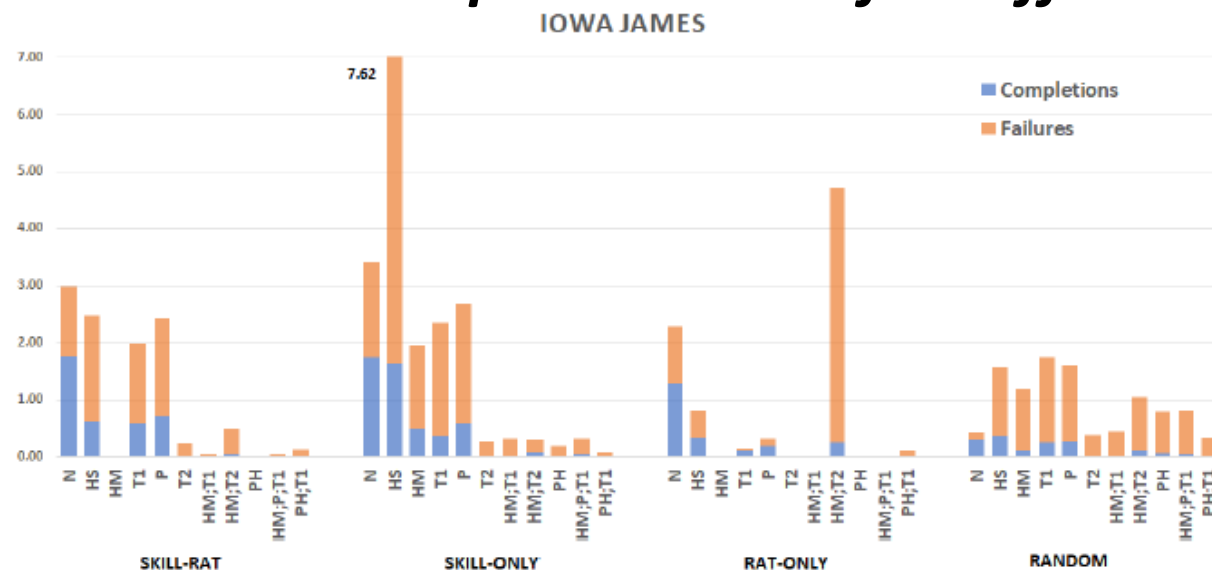
Iowa James ($\chi^2(12) = 34.5; p < .001$)

| Skill | SKILL_RAT | SKILL_ONLY | RAT_ONLY | RANDOM |
|-----------|-----------|------------|----------|--------|
| white | 86.5 | 93.8 | 79.6 | 46.5 |
| black | 88.5 | 87.7 | 61.1 | 46.5 |
| star | 48.1 | 50.6 | 42.6 | 30.2 |
| challenge | 9.6 | 17.3 | 3.7 | 20.9 |

Paradox ($\chi^2(5) = 25.9; p = .002$)

- Percentage of players acquiring individual skills in each progression
- Using **skill chains** can lead to players acquiring skills at a higher rate

Failure and Completion Rates for Different Level Types



- Using only **rating systems** causes players to complete fewer different types of levels

Number of completions and failures for each level type in both games