Meet Your Match Rating: Providing Skill Information and Choice in Player-vs-Level Matchmaking

Anurag Sarkar and **Seth Cooper**

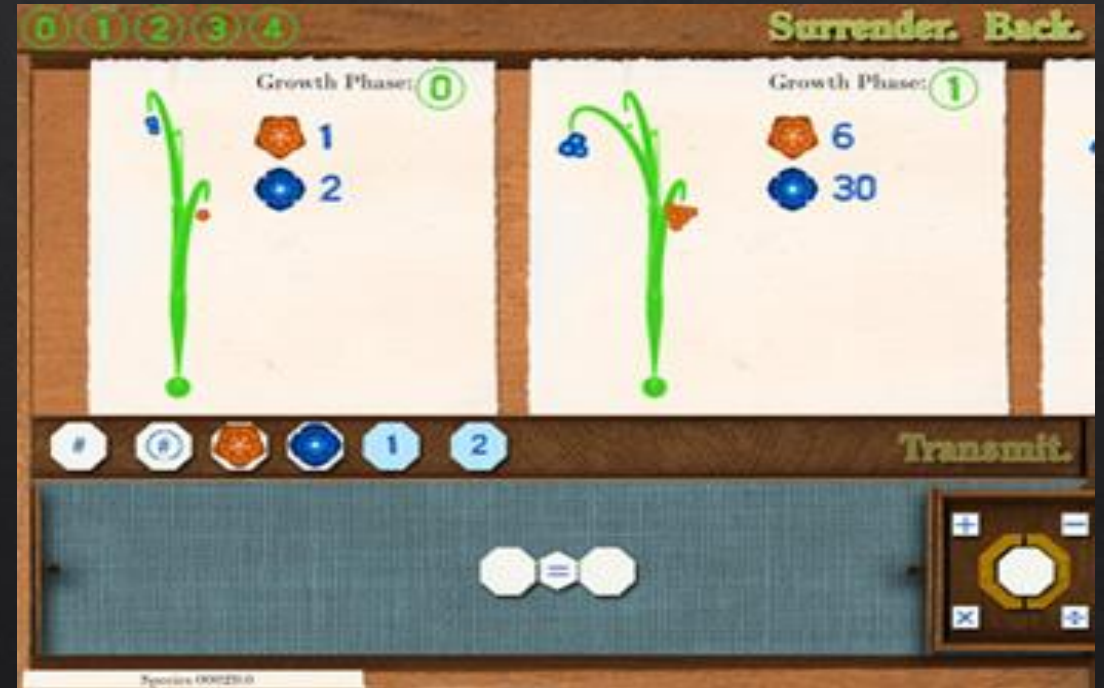*College of Computer and Information Science*

Northeastern University

Human Computation Games (HCGs)

Games that motivate large numbers of people to solve tasks that are hard to automate

Nanocrafter

Xylem

Engagement and Difficulty Balancing

- ◈ HCGs suffer from
  - ◈ Poor engagement
  - ◈ Poor player retention

# Engagement and Difficulty Balancing

◈ HCGs suffer from

  ◈ Poor engagement

  ◈ Poor player retention

◈ REASON - Lack of difficulty balancing in HCGs

Engagement and Difficulty Balancing

◈ HCGs suffer from

    ◈ Poor engagement

    ◈ Poor player retention

◈ REASON - Lack of difficulty balancing in HCGs

    ◈ No *a priori* knowledge of difficulty of tasks to be solved

    ◈ Not possible to modify tasks without compromising validity of solutions

Solution: Player Rating Systems

◈ In our previous work, we used rating systems to order levels for players by mapping player skill and level difficulty to ratings

◈ Rating system could then be used to match players of certain skill with levels of comparable difficulty

◈ Serving levels in an order determined by the Glicko-2 system was shown to improve player engagement
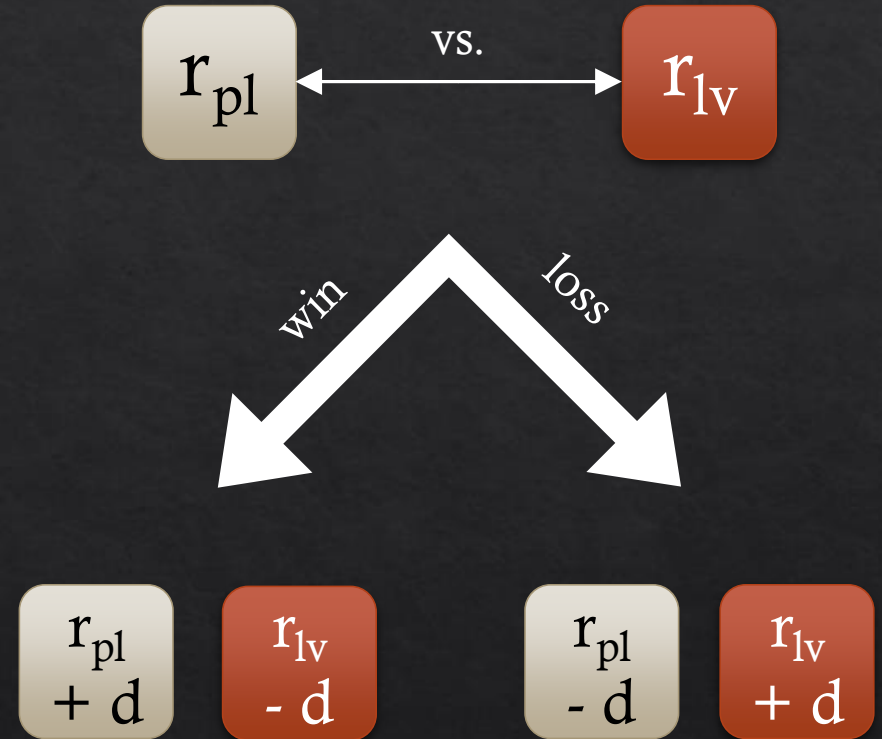
r_{pl} vs. r_{lv}

win   loss

$r_{pl} + d$ $r_{lv} - d$ $r_{pl} - d$ $r_{lv} + d$

Skill Feedback & Choice

◈ Previous work performed matchmaking unbeknownst to players with players being oblivious to the rating system

Skill Feedback & Choice

◈ Previous work performed matchmaking unbeknownst to players with players being oblivious to the rating system

◈ Further engagement benefits could be achieved by

◈ Informing players of the matchmaking system

Skill Feedback & Choice

◈ Previous work performed matchmaking unbeknownst to players with players being oblivious to the rating system

◈ Further engagement benefits could be achieved by

   ◈ Informing players of the matchmaking system

   ◈ Offering them choice of next level difficulty

Self Determination Theory (SDT)

- Theory of human motivation and psychology

Self Determination Theory (SDT)

- Theory of human motivation and psychology

- Three innate psychological needs
  - *Relatedness*
  - *Autonomy*
  - *Competence*

Self Determination Theory (SDT)

◆ Theory of human motivation and psychology

◆ Three innate psychological needs

　◆ *Relatedness (HCGs may already tap into this)*

　◆ *Autonomy*

　◆ *Competence*

Self Determination Theory (SDT)

⬩ Theory of human motivation and psychology

⬩ Three innate psychological needs

   ⬩ *Relatedness (HCGs may already tap into this)*

   ⬩ *Autonomy (by offering choice)*

   ⬩ *Competence (by providing skill feedback)*

Hypotheses

◈ *H1 – Informing players of ratings and rating system will lead to better engagement and experience than not informing them*

Hypotheses

◈ *H1 – Informing players of ratings and rating system will lead to better engagement and experience than not informing them*

◈ *H2 – Additionally offering choice of level difficulty will lead to even better engagement and experience than when only informing them of the rating system*
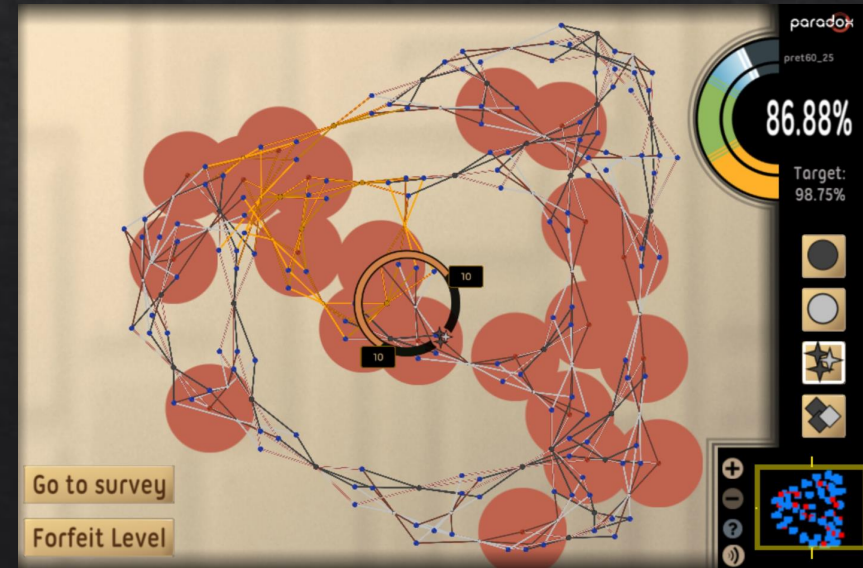
Paradox

- 2D puzzle game for crowdsourced formal verification of software

- Each level represents a MAX-SAT problem

- Players assign values to variables, schedule optimizations

- Player completes level by reaching target score
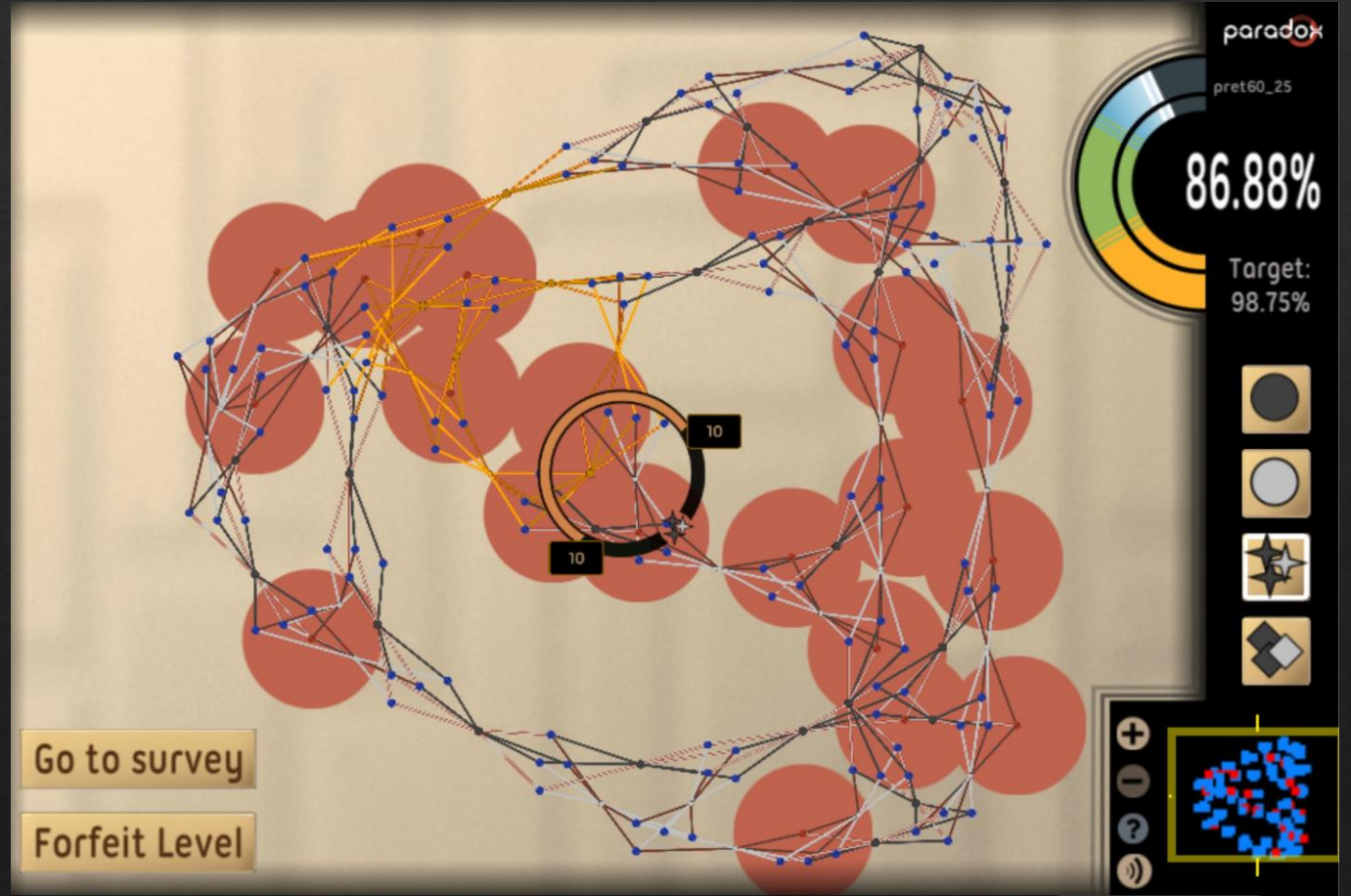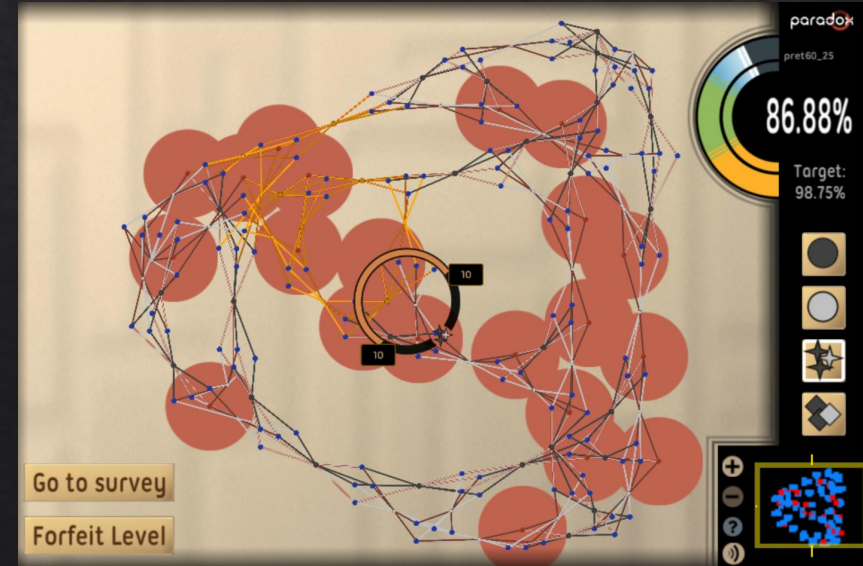
Participant Recruitment and Study

◈ Players recruited using Mechanical Turk

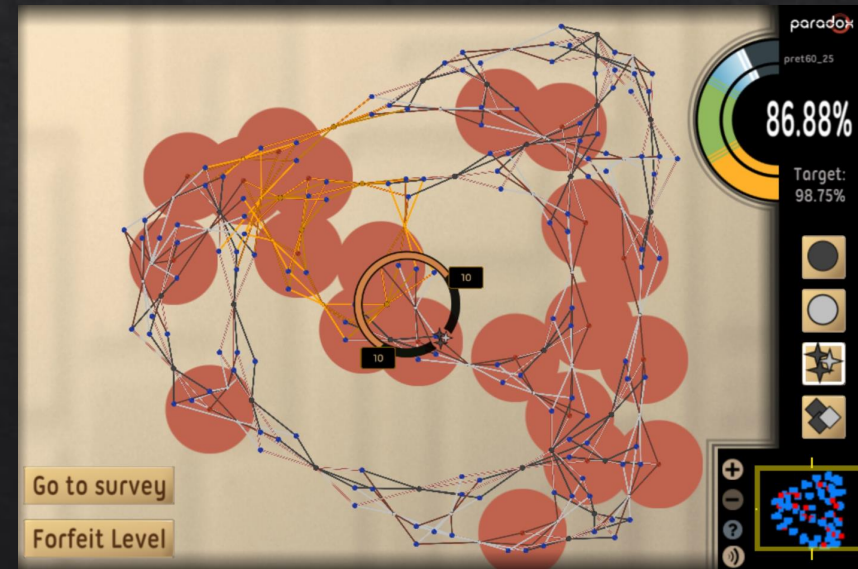# Participant Recruitment and Study

◇ Players recruited using Mechanical Turk

◇ Two part study

◇ Feedback & Choice Experiment using Glicko-2 rating system

◇ Post-Game Survey using Intrinsic Motivation Inventory (IMI)

# Participant Recruitment and Study

◈ Players recruited using Mechanical Turk

◈ Two part study

  ◈ Feedback & Choice Experiment using Glicko-2 rating system

    ◈ 9 mandatory tutorial levels

    ◈ 55 optional challenge levels

  ◈ Post-Game Survey using Intrinsic Motivation Inventory (IMI)

Participant Recruitment and Study

◈ Players recruited using Mechanical Turk

◈ Two part study

   ◈ Feedback & Choice Experiment using Glicko-2 rating system

      ◈ 9 mandatory tutorial levels

      ◈ 55 optional challenge levels

   ◈ Post-Game Survey using Intrinsic Motivation Inventory (IMI)

      ◈ 25 questions
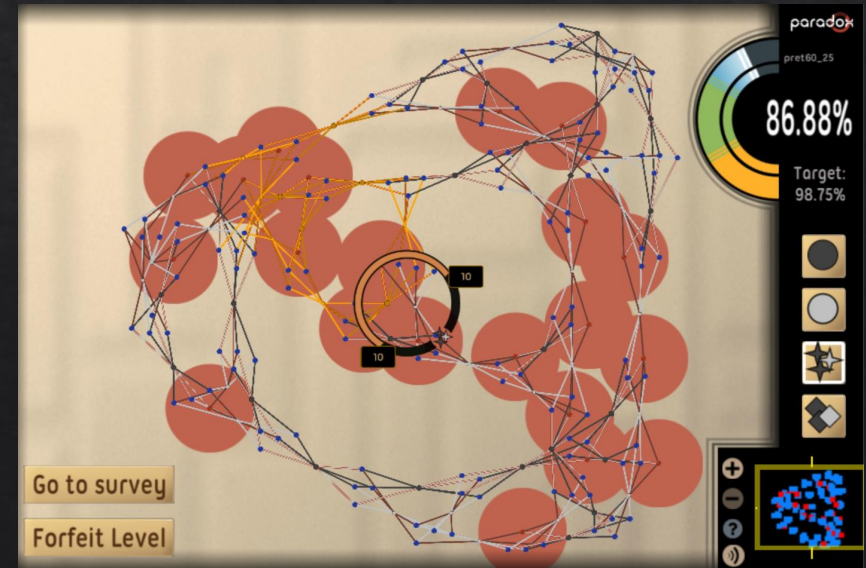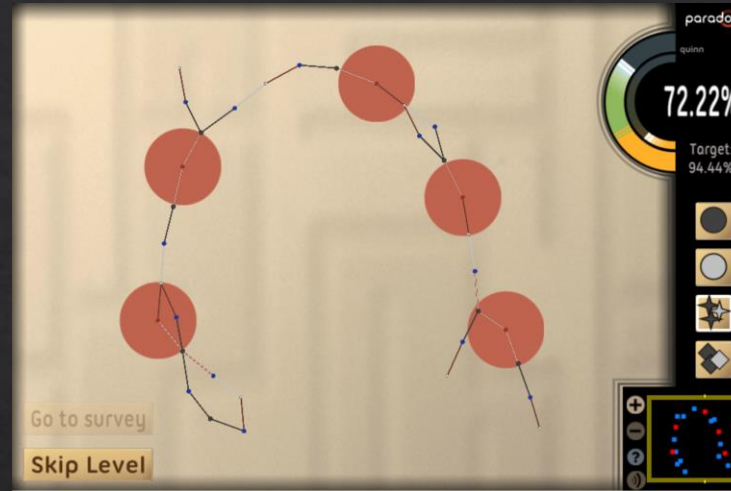
Ratings Feedback & Choice Experiment

- ◈ Player-level pairings considered as matches

- ◈ Match outcomes:
  - ◇ Level Completed => Player wins
  - ◇ Level Forfeited => Level wins
  - ◇ Level Skipped => Ignore

Ratings Feedback & Choice Experiment

- Player-level pairings considered as matches

- Match outcomes:
  - Level Completed => Player wins
  - Level Forfeited => Level wins
  - Level Skipped => Ignore

- Three experimental conditions
  - BLIND

BLIND

Ratings Feedback & Choice Experiment
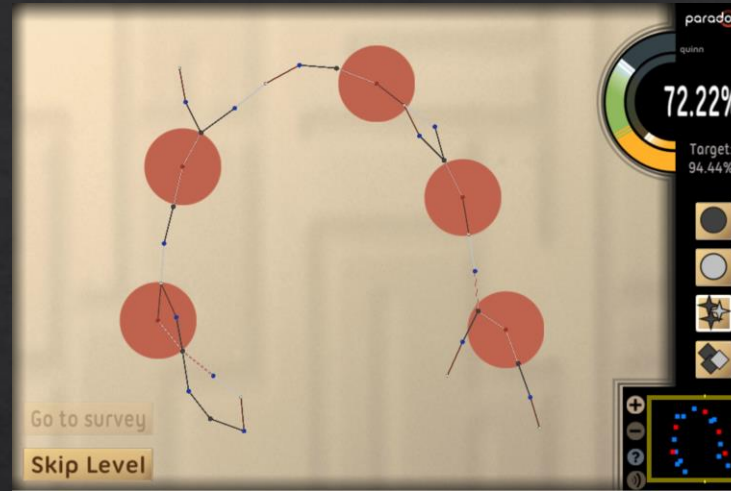
◈ Player-level pairings considered as matches

◈ Match outcomes:
  - ◇ Level Completed => Player wins
  - ◇ Level Forfeited => Level wins
  - ◇ Level Skipped => Ignore

◈ Three experimental conditions
  - ◇ BLIND
  - ◇ RATINGS

BLIND

RATINGS

Ratings Feedback & Choice Experiment

◈ Player-level pairings considered as matches

◈ Match outcomes:

   ◇ Level Completed => Player wins

   ◇ Level Forfeited => Level wins

   ◇ Level Skipped => Ignore
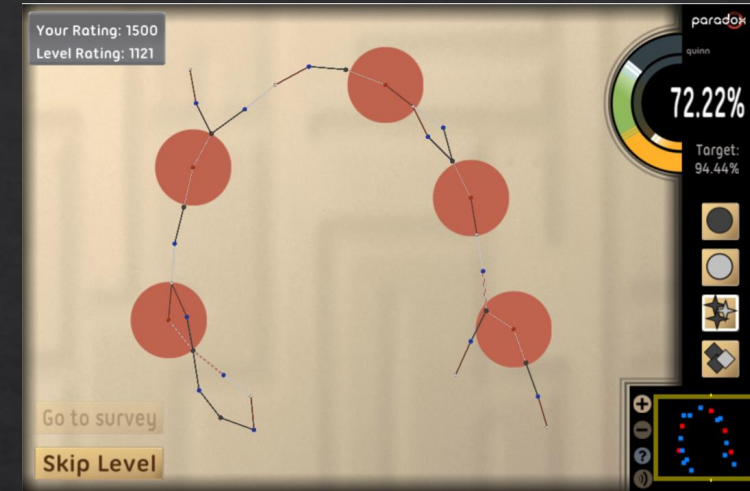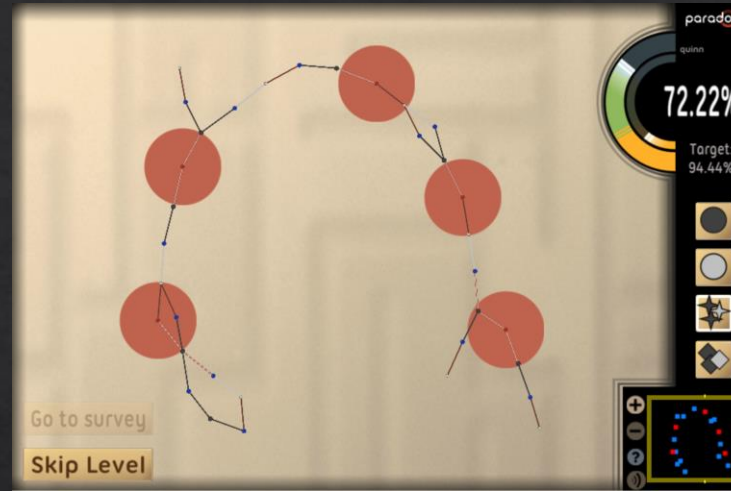
◈ Three experimental conditions

   ◇ BLIND

   ◇ RATINGS

   ◇ CHOICE
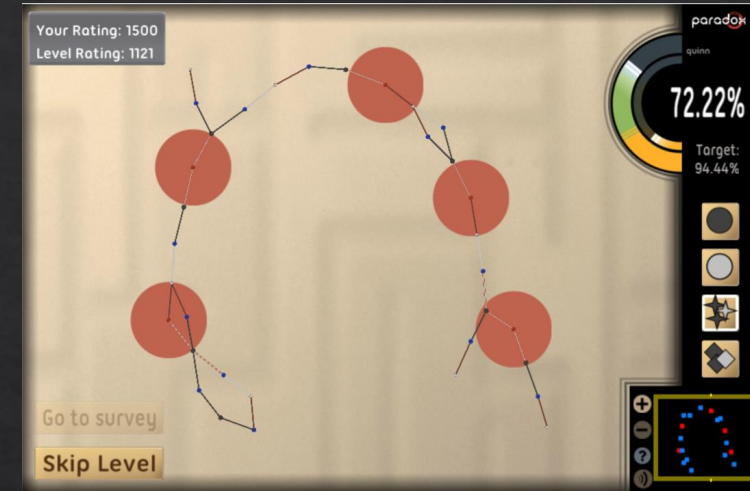
BLIND

RATINGS

CHOICE

Blind and Ratings Condition

BLIND

RATINGS

Compute desired win rate using player's rating

Blind and Ratings Condition

BLIND

RATINGS

Compute desired win rate using player's rating

Compute player's win expectancy versus each remaining level

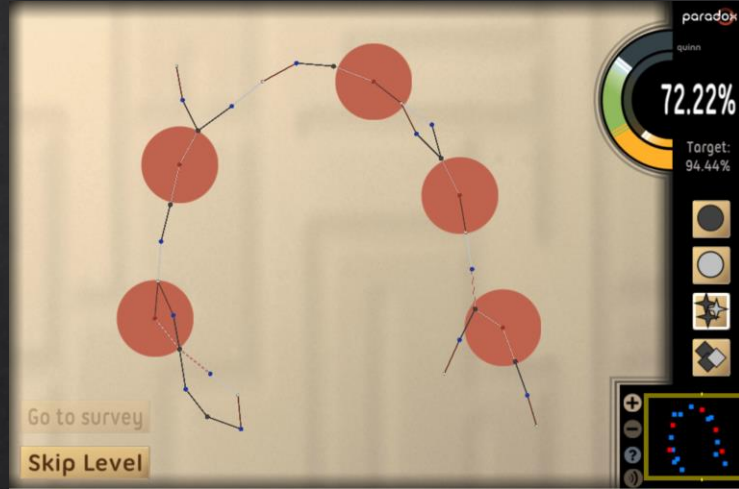# Blind and Ratings Condition
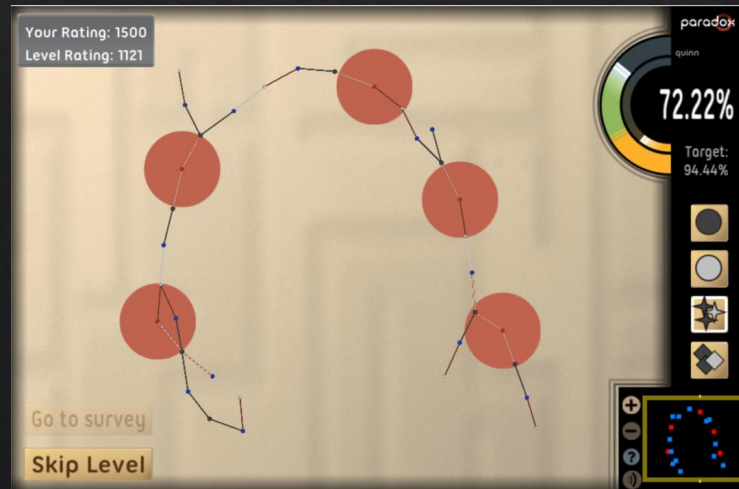

BLIND

RATINGS

Compute desired win rate using player's rating

Compute player's win expectancy versus each remaining level

Serve level with win expectancy closest to desired win rate
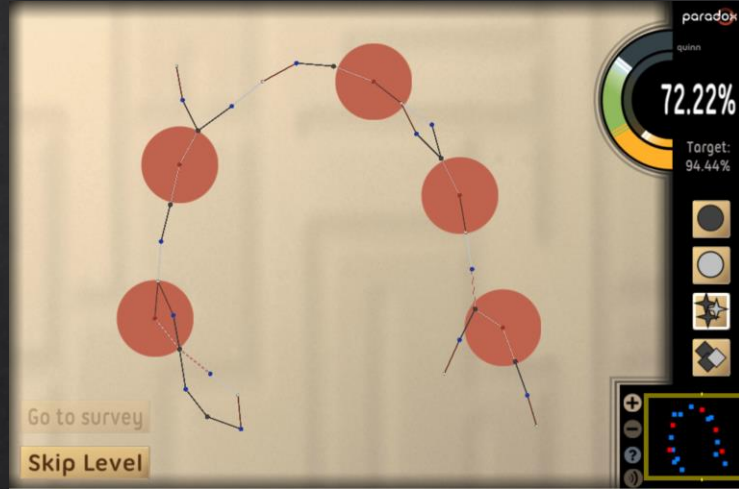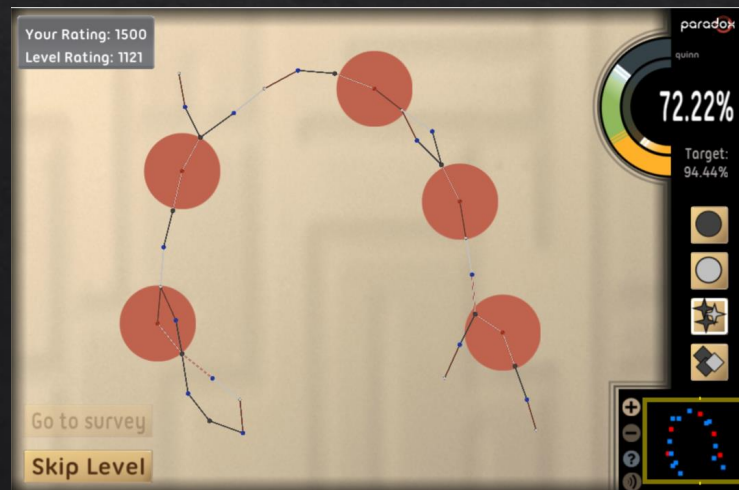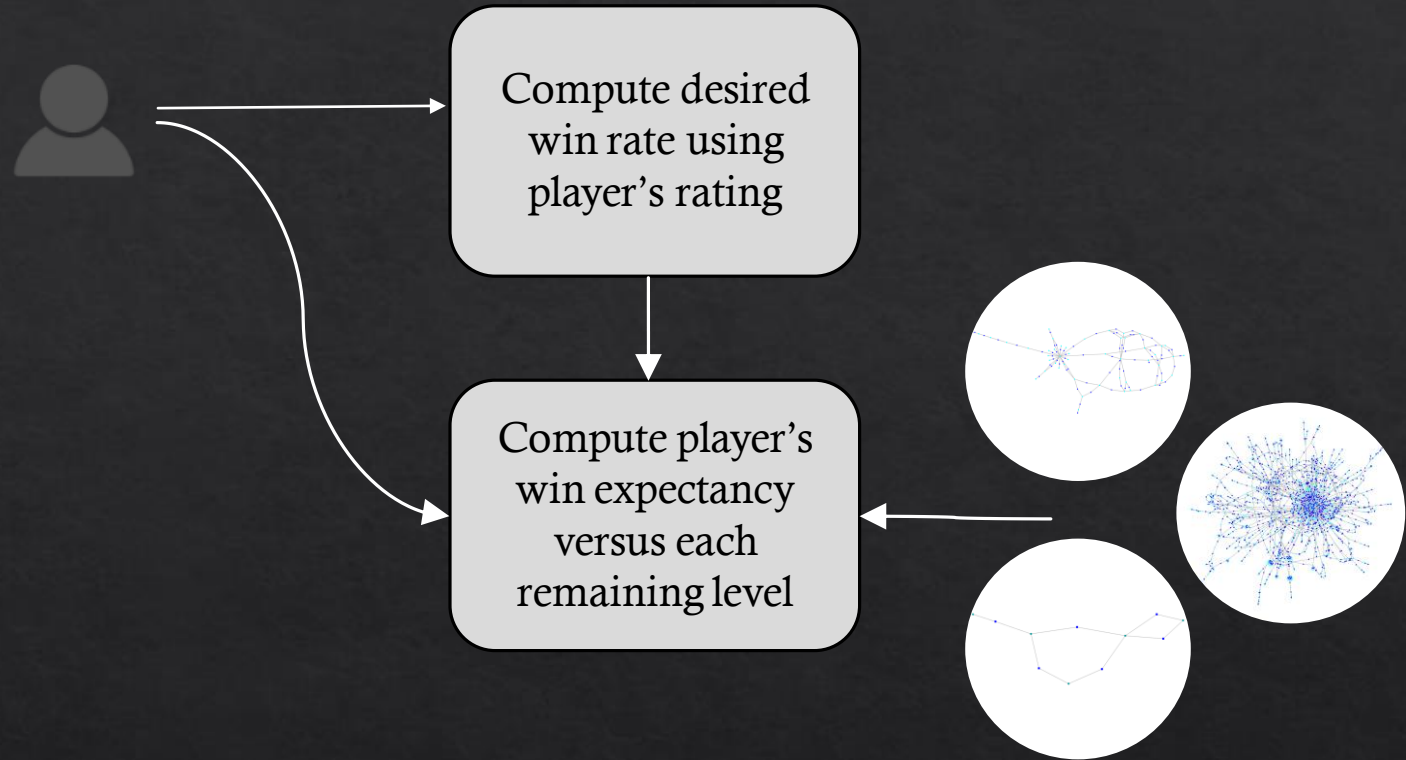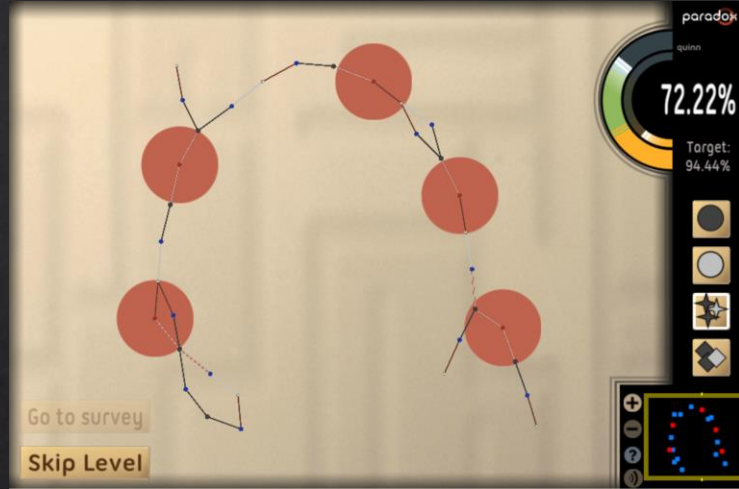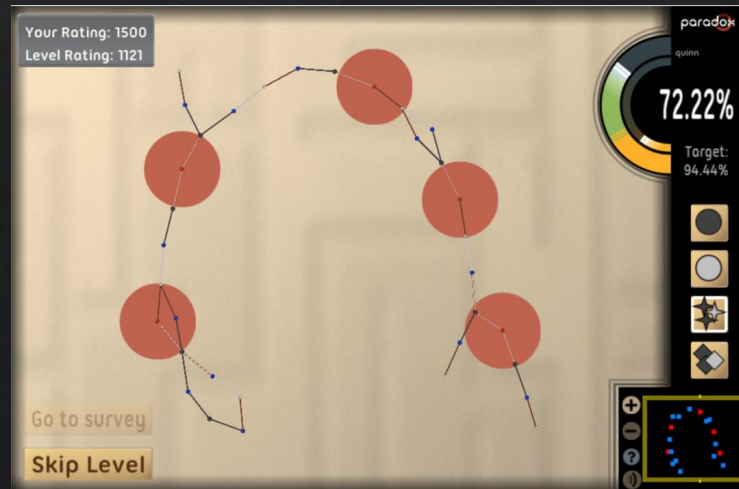
vs.

Choice Condition

CHOICE

Compute desired win rate using (player's rating – 400)

Compute desired win rate using player's rating

Compute desired win rate using (player's rating + 400)

Choice Condition

CHOICE

Compute desired win rate using (player's rating – 400)

Compute desired win rate using player's rating

Compute desired win rate using (player's rating + 400)

Compute player's win expectancy versus each remaining level

Compute player's win expectancy versus each remaining level

Compute player's win expectancy versus each remaining level

Choice Condition

CHOICE

vs.

Compute desired win rate using
(player's rating – 400)

Compute desired win rate using
player's rating

Compute desired win rate using
(player's rating + 400)

Compute player's win expectancy versus each remaining level

Compute player's win expectancy versus each remaining level

Compute player's win expectancy versus each remaining level

EASY

RECOMMENDED

HARD

Measures

◈ 278 workers randomly assigned to one of the three conditions (BLIND – 111, RATINGS – 96, CHOICE – 71)

Measures

- 278 workers randomly assigned to one of the three conditions (BLIND – 111, RATINGS – 96, CHOICE – 71)

- Behavioral Engagement

  - *Challenge Time*

  - *Levels Attempted*

  - *Levels Completed*

  - *Player Rating (Player's Glicko-2 rating after completing the game)*

  - *Highest Level Rating (Highest Glicko-2 rating of any level completed by the player)*
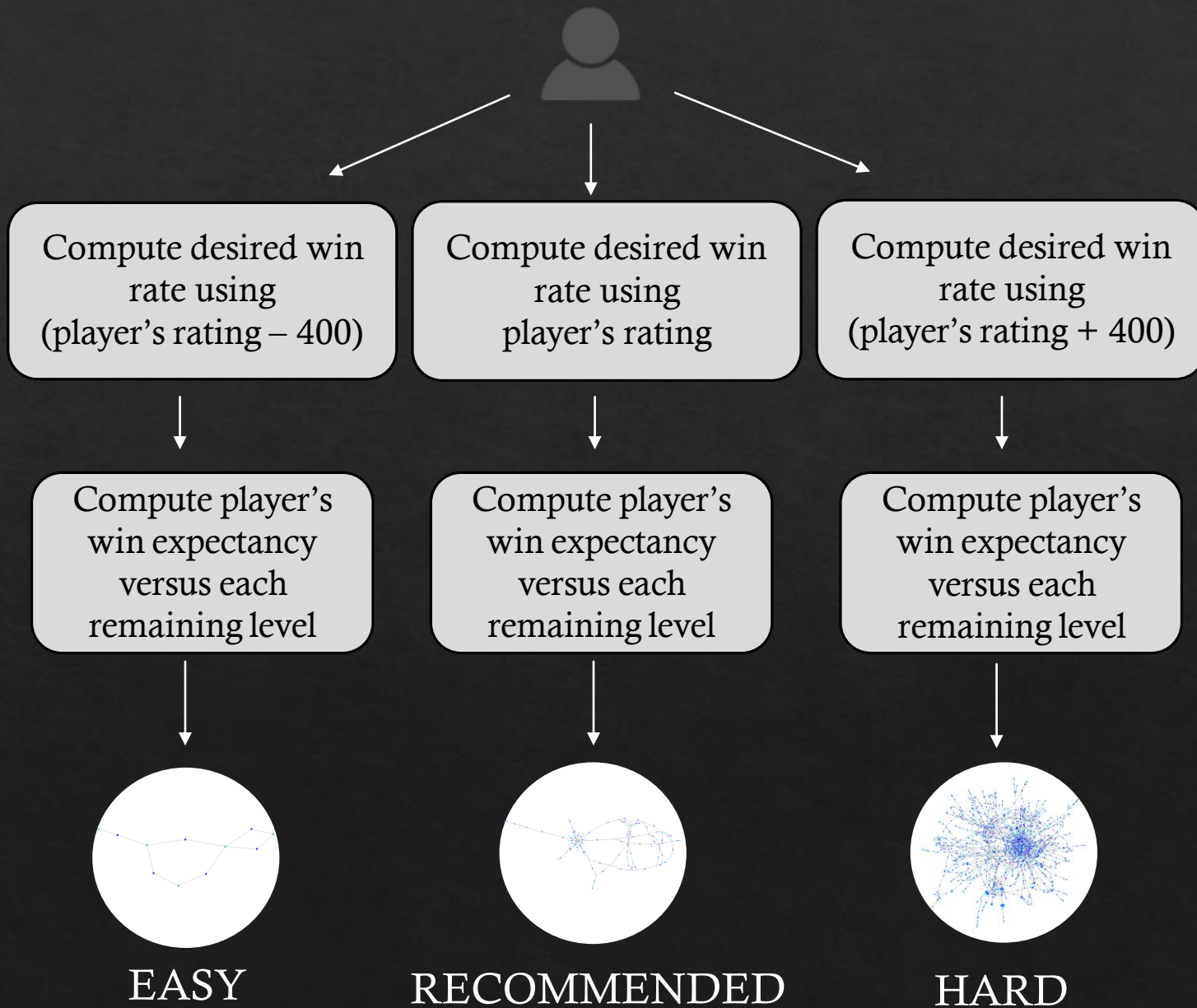
Measures

◇ 278 workers randomly assigned to one of the three conditions (BLIND – 111, RATINGS – 96, CHOICE – 71)

◇ Behavioral Engagement

　◇ *Challenge Time*

　◇ *Levels Attempted*

　◇ *Levels Completed*

　◇ *Player Rating (Player's Glicko-2 rating after completing the game)*

　◇ *Highest Level Rating (Highest Glicko-2 rating of any level completed by the player)*

◇ Intrinsic Motivation Inventory

　◇ *Interest/Enjoyment*

　◇ *Perceived Competence*

　◇ *Perceived Choice*

　◇ *Effort/Importance*

Results

Variable	BLIND	RATINGS	CHOICE
Challenge Time	**515**	**791**	**897**
Levels Attempted	**7**	**10**	**12**
Levels Completed	**5**	**7**	**8**

Statistical Tests: Omnibus Kruskal-Wallis Test, post-hoc Wilcoxon Rank-Sum Test

◇ No significant difference across conditions for *Player Rating* and *Highest Level Rating*

Results

Variable	BLIND	RATINGS	CHOICE
Interest/Enjoyment	63%	65%	63%
Perceived Competence	57%	52%	57%
Perceived Choice	78%	80%	82%
Effort/Importance	83%	86%	83%

Statistical Tests: Omnibus Kruskal-Wallis Test, post-hoc Wilcoxon Rank-Sum Test

◈ No significant difference across conditions for any survey variable

Discussion

◈ *H1 is partially supported*

  ◈ Players did better in terms of *Challenge Time*, *Levels Attempted* and *Levels Completed* under RATINGS as compared to BLIND

  ◈ No improvement observed between two conditions in terms of *Player Rating*, *Highest Level Rating* or any survey variable

Discussion

- *H1 is partially supported*

  - Players did better in terms of *Challenge Time*, *Levels Attempted* and *Levels Completed* under RATINGS as compared to BLIND

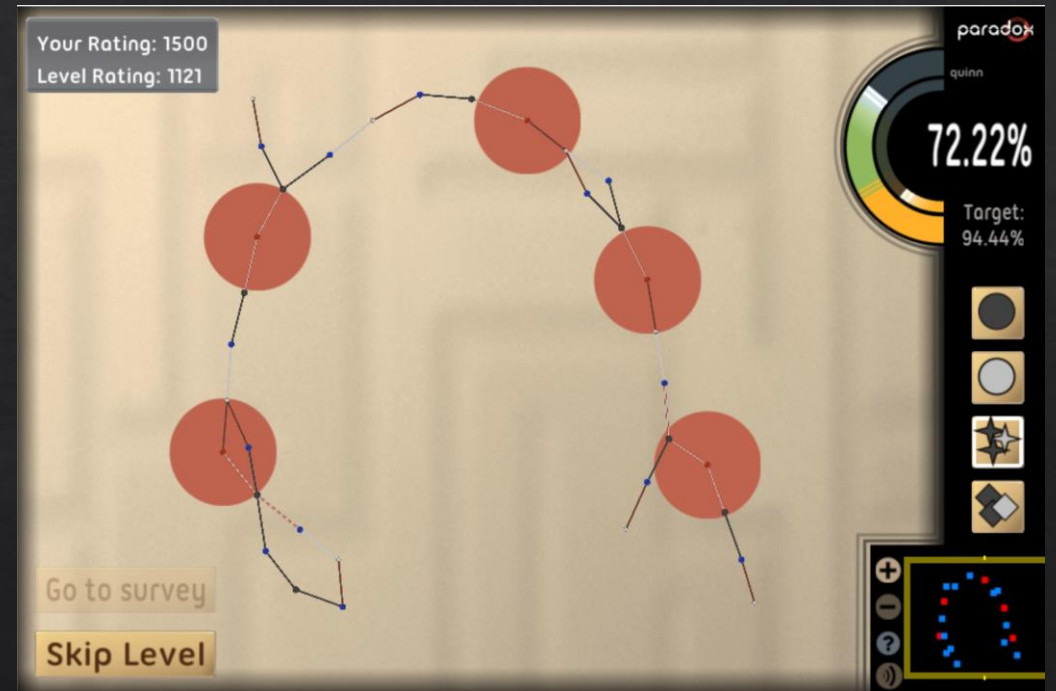 - No improvement observed between two conditions in terms of *Player Rating*, *Highest Level Rating* or any survey variable

- *H2 is rejected*

  - No significant improvement in CHOICE condition for any measured variables as compared to RATINGS

Ratings Feedback Discussion
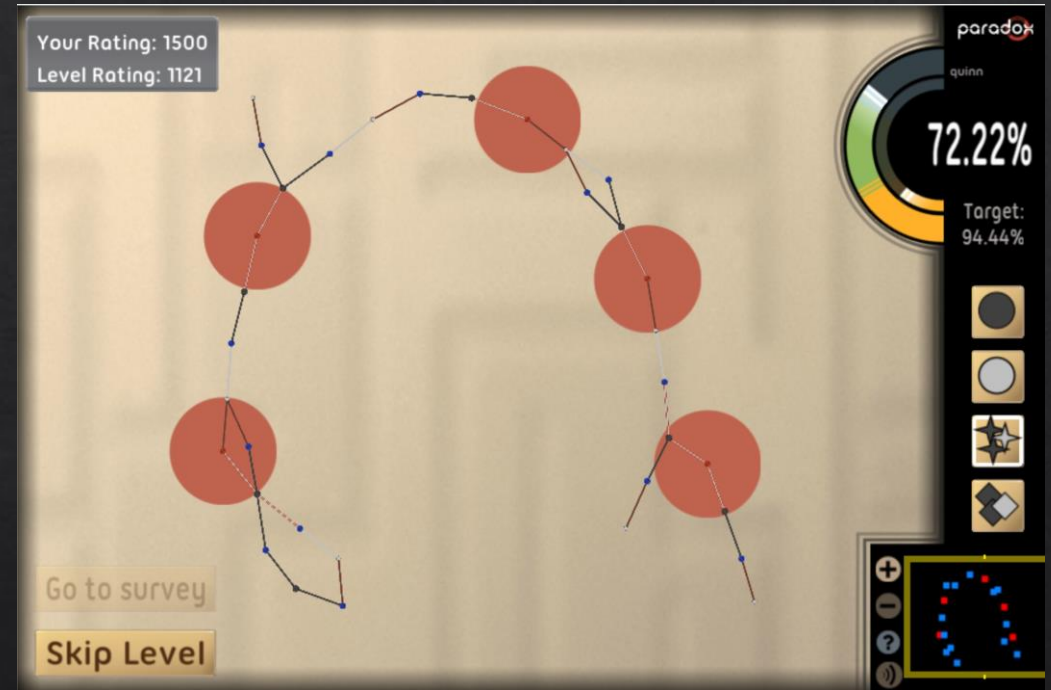
◈ Informed players played longer and completed more levels

RATINGS

Ratings Feedback Discussion

◈ Informed players played longer and completed more levels

◈ But did not gain higher ratings or complete more difficult levels

RATINGS

Ratings Feedback Discussion
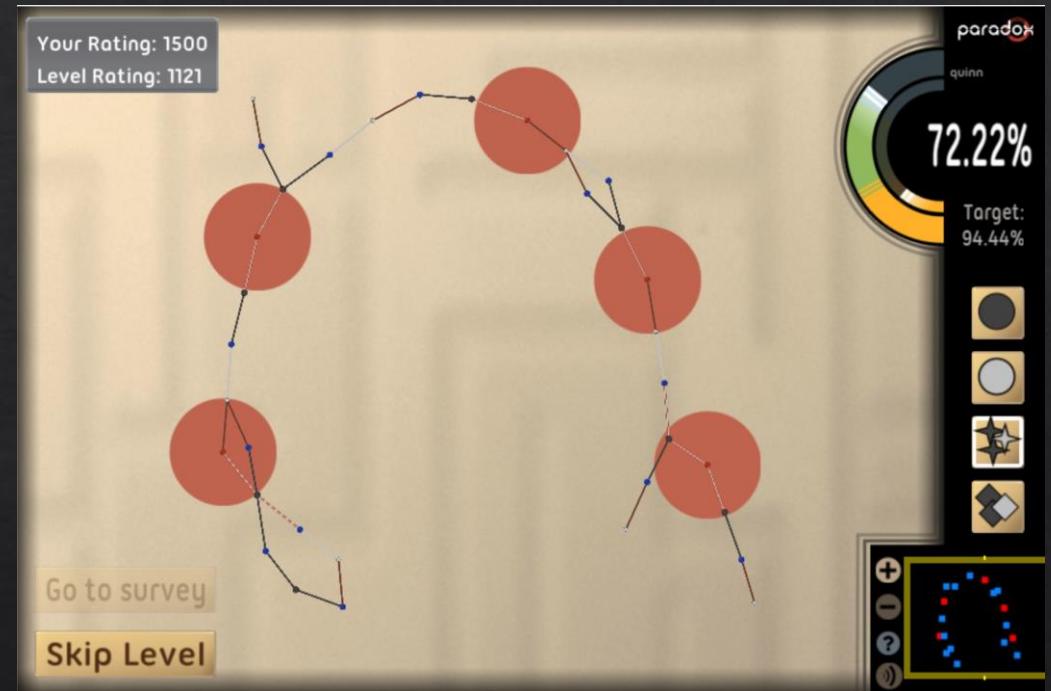
◈ Informed players played longer and completed more levels

◈ But did not gain higher ratings or complete more difficult levels

◈ Informed players didn't try to game the system by attempting fewer levels to hold onto current rating

RATINGS

Choice Discussion

◈ Choice of difficulty impacted neither engagement metrics nor experience measures in the survey

CHOICE

Choice Discussion

◈ Choice of difficulty impacted neither engagement metrics nor experience measures in the survey

◈ *Perceived Choice* not significantly increased under CHOICE

◈ CHOICE offered explicit choices, but implicit choices in other conditions may have been meaningful enough

◈ Players could skip levels and stop playing whenever they wanted to in all conditions

CHOICE

Choice of Level Difficulty

◈ Choice of level difficulty often impacted by previous match outcome

Previous Result	Easy	Recommended	Hard
Complete (Win)	40%	49%	11%
Forfeit (Loss)	41%	36%	23%
Skip	57%	32%	11%

Percentage of times each option selected given last outcome

$x^2(4)=37.3$, $p<0.001$

Choice of Level Difficulty

◈ Choice of level difficulty often impacted by previous match outcome

◈ Win → Recommended

Previous Result	Easy	Recommended	Hard
Complete (Win)	40%	**49%**	11%
Forfeit (Loss)	41%	36%	23%
Skip	57%	32%	11%

Percentage of times each option selected given last outcome

$x^2(4)=37.3$, $p<0.001$

Choice of Level Difficulty

◈ Choice of level difficulty often impacted by previous match outcome

◈ Win → Recommended

◈ Skip → Easy

Previous Result	Easy	Recommended	Hard
Complete (Win)	40%	49%	11%
Forfeit (Loss)	41%	36%	23%
Skip	**57%**	32%	11%

Percentage of times each option selected given last outcome

$x^2(4)=37.3$, $p<0.001$

Choice of Level Difficulty

◈ Choice of level difficulty often impacted by previous match outcome

◈ Win → Recommended

◈ Skip → Easy

◈ Forfeit → Hard more often than after a win or a skip

Previous Result	Easy	Recommended	Hard
Complete (Win)	40%	49%	11%
Forfeit (Loss)	41%	36%	**23%**
Skip	57%	32%	11%

Percentage of times each option selected given last outcome

$x^2(4)=37.3$, $p<0.001$

Conclusion

◈ Skill feedback increased player engagement in terms of number of levels attempted and completed and time spent playing

Conclusion

◇ Skill feedback increased player engagement in terms of number of levels attempted and completed and time spent playing

◇ Offering choice of difficulty improved player engagement but not significantly and the choice made by players was impacted by previous match outcome

Future Work

◇ Examination of how meaningful different choices are

Future Work

◈ Examination of how meaningful different choices are

◈ Effects of previous match outcomes on player choice

Contact

Anurag Sarkar
Northeastern University
sarkar.an@husky.neu.edu

Acknowledgments