Collaborative solving in a human computing game

Olivier Tremblay-Savard[1,2], Alexander Butyaev[1]
and Jérôme Waldispühl[1]

[1] School of Computer Science, McGill University, Montreal
[2] Department of Computer Science, University of Manitoba, Winnipeg

McGill School of Computer Science Seminar, October 11th

Outline

- Introduction to human computing
- Collaborative solving with a market, skills and challenges

Human computing

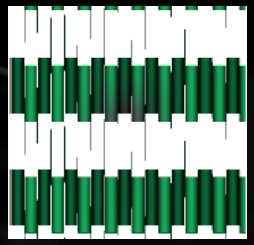
- Human computing is about combining the strengths of the human brain and those of the computer

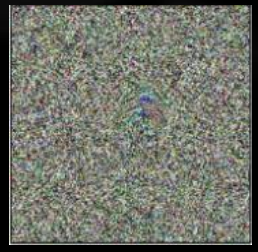
- Can be used to solve specific problems
  - Perception (image labeling, sound recognition)
  - Knowledge (collecting common-sense facts)
  - Reasoning (puzzle solving)
  - Coordination (robotics)

Human computing

- Deep Neural Networks are Easily Fooled (Nguyen *et al.* 2014)





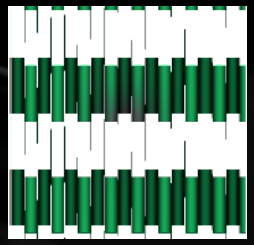



Human computing

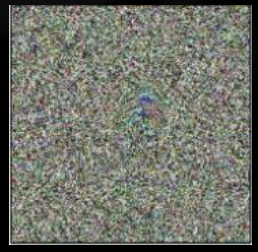
- Deep Neural Networks are Easily Fooled (Nguyen *et al.* 2014)


Green snake


Volcano


Screwdriver


Ski mask

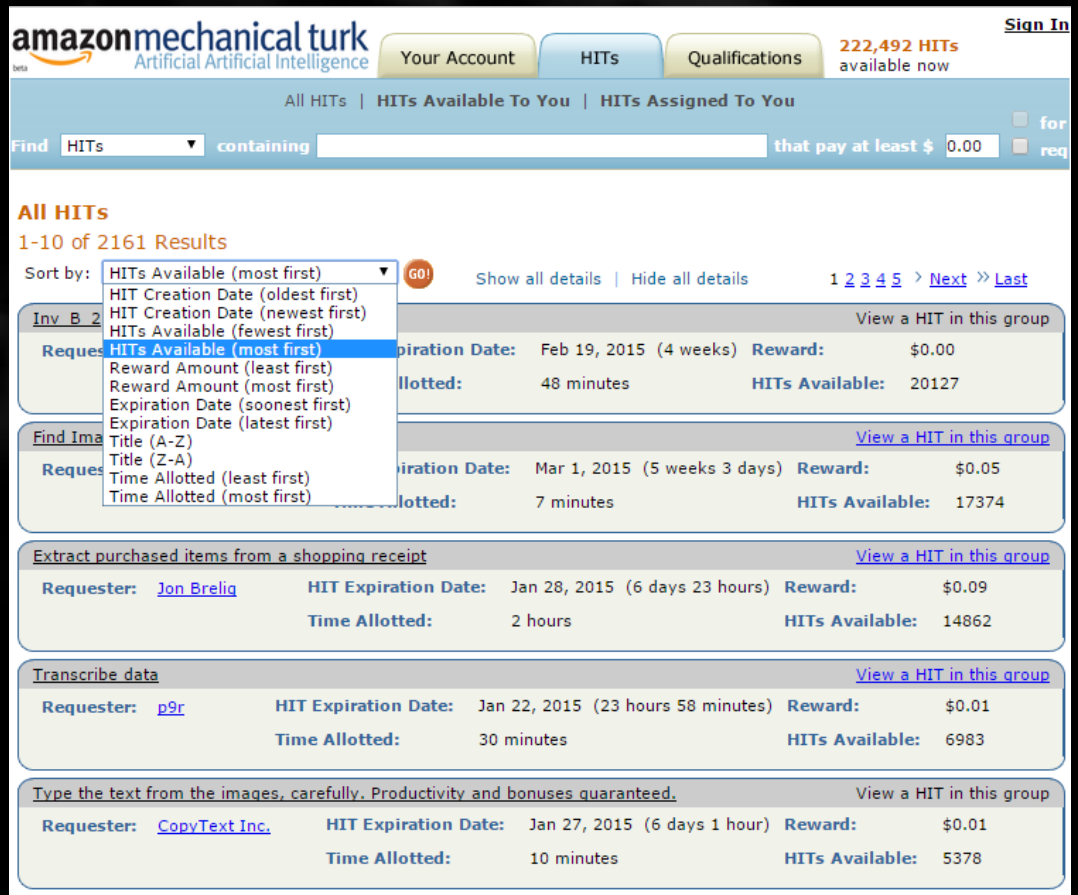

Peacock


Lesser panda

Confidence score > 99%

Human computing

- Typically, participants (usually called “workers”) perform the task voluntarily or in exchange of money
- Amazon Mechanical Turk :



Human computing games

- It is also possible to embed the tasks into games

- Human computing games (games with a purpose):
  - ESP
  - Foldit
  - Phylo
  - Ribo

Human brain power available

Top Games By Current Players — Next page

	Name	Current Players	Last 30 Days	Peak Players	Hours Played
1.	Dota 2	531,393		1,048,369	446,843,777
2.	Counter-Strike: Global Offensive	325,786		638,360	232,411,115
3.	Sid Meier's Civilization V	47,907		72,026	27,741,723
4.	Team Fortress 2	43,942		74,797	33,366,726
5.	Rocket League	42,995		79,304	24,558,106
6.	Grand Theft Auto V	38,181		67,892	24,710,445
7.	ARK: Survival Evolved	35,715		67,157	26,593,375
8.	Rust	34,233		45,454	17,696,477
9.	Garry's Mod	29,275		53,266	18,170,124
10.	Paladins	27,434		56,034	12,781,680
11.	Mafia III	26,180		38,727	607,440
12.	Football Manager 2016	25,550		58,138	22,824,817
13.	Arma 3	24,770		41,168	14,944,724
14.	The Elder Scrolls V: Skyrim	22,523		38,040	15,194,066
15.	Fallout 4	22,051		44,004	15,937,087
16.	Warframe	21,586		36,009	15,765,819
17.	Unturned	21,162		48,001	16,590,465
18.	H1Z1: King of the Kill	18,361		24,677	8,098,585
19.	Terraria	14,783		24,473	9,370,711
20.	Path of Exile	14,727		30,706	11,398,165

In the last 30 days, for the 20 most popular PC games on Steam alone:

- At least 2,586,602 distinct players in total
- A total of 995,605,427 hours played

http://steamcharts.com/

Human brain power available

Top Games By Current Players

Next page

	Name	Current Players	Last 30 Days	Peak Players	Hours Played
1.	Dota 2	531,393		1,048,369	446,843,777
2.	Counter-Strike: Global Offensive	325,786		638,360	232,411,115
3.	Sid Meier's Civilization V	47,907		72,026	27,741,723
4.	Team Fortress 2	43,942		74,797	33,366,726
5.	Rocket League	42,995		79,304	24,558,106
6.	Grand Theft Auto V	38,181		67,892	24,710,445
7.	ARK: Survival Evolved	35,715		67,157	26,593,375
8.	Rust	34,233		45,454	17,696,477
9.	Garry's Mod	29,275		53,266	18,170,124
10.	Paladins	27,434		56,034	12,781,680
11.	Mafia III	26,180		38,727	607,440
12.	Football Manager 2016	25,550		58,138	22,824,817
13.	Arma 3	24,770		41,168	14,944,724
14.	The Elder Scrolls V: Skyrim	22,523		38,040	15,194,066
15.	Fallout 4	22,051		44,004	15,937,087
16.	Warframe	21,586		36,009	15,765,819
17.	Unturned	21,162		48,001	16,590,465
18.	H1Z1: King of the Kill	18,361		24,677	8,098,585
19.	Terraria	14,783		24,473	9,370,711
20.	Path of Exile	14,727		30,706	11,398,165

995,605,427 hours
played in 30 days
x
12 months
=

http://steamcharts.com/

Human brain power available

Top Games By Current Players — Next page

	Name	Current Players	Last 30 Days	Peak Players	Hours Played
1.	Dota 2	531,393		1,048,369	446,843,777
2.	Counter-Strike: Global Offensive	325,786		638,360	232,411,115
3.	Sid Meier's Civilization V	47,907		72,026	27,741,723
4.	Team Fortress 2	43,942		74,797	33,366,726
5.	Rocket League	42,995		79,304	24,558,106
6.	Grand Theft Auto V	38,181		67,892	24,710,445
7.	ARK: Survival Evolved	35,715		67,157	26,593,375
8.	Rust	34,233		45,454	17,696,477
9.	Garry's Mod	29,275		53,266	18,170,124
10.	Paladins	27,434		56,034	12,781,680
11.	Mafia III	26,180		38,727	607,440
12.	Football Manager 2016	25,550		58,138	22,824,817
13.	Arma 3	24,770		41,168	14,944,724
14.	The Elder Scrolls V: Skyrim	22,523		38,040	15,194,066
15.	Fallout 4	22,051		44,004	15,937,087
16.	Warframe	21,586		36,009	15,765,819
17.	Unturned	21,162		48,001	16,590,465
18.	H1Z1: King of the Kill	18,361		24,677	8,098,585
19.	Terraria	14,783		24,473	9,370,711
20.	Path of Exile	14,727		30,706	11,398,165

995,605,427 hours
played in 30 days
x
12 months
=
11,947,265,124

12 billion hours
per year
of brain power
available

http://steamcharts.com/

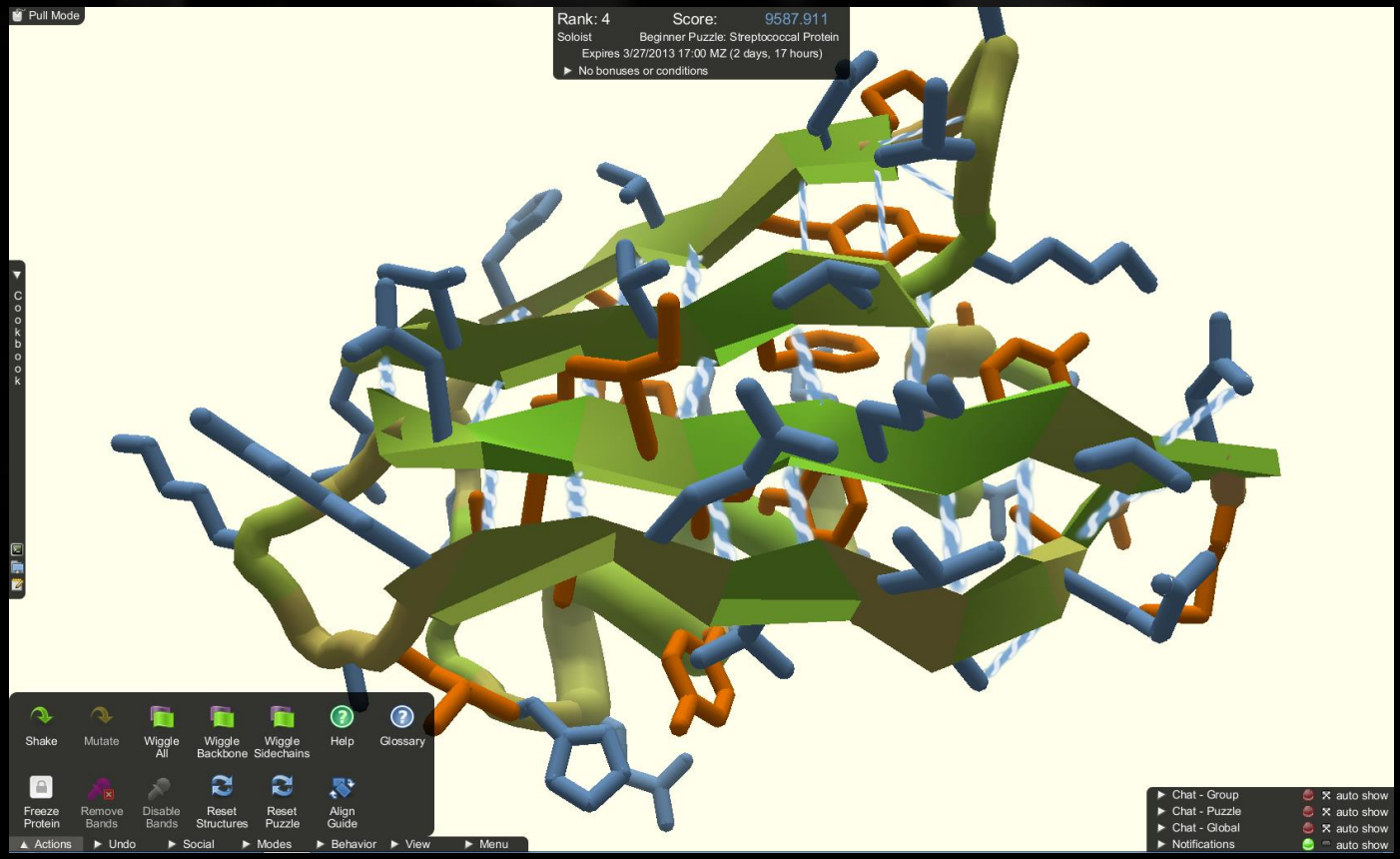
ESP game

"It turns out that the string on which the two players agree is typically a good label for the image"

L. von Ahn and L. Dabbish, *Labeling Images with a Computer Game*, CHI 2004

Foldit

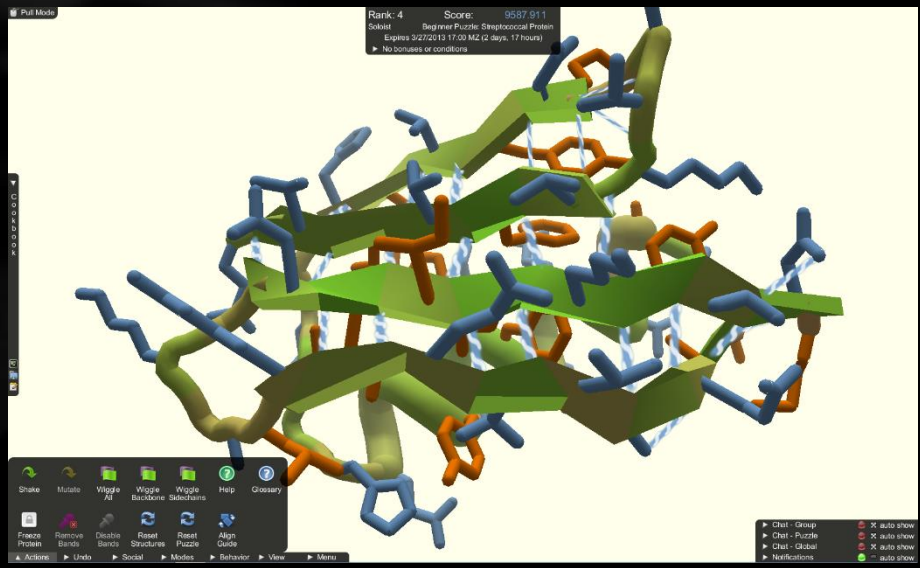
- First human computing game on molecular biology



http://fold.it/portal/

Foldit

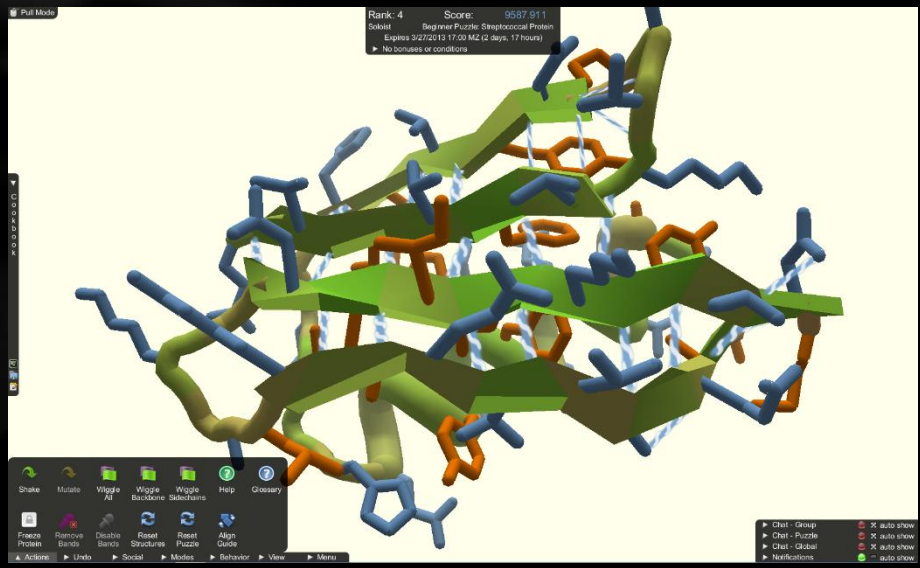
- Has been used to solve the crystal structure of a retroviral protease (Khatib *et al.* 2011)



http://fold.it/portal/

Foldit

- Has been used to solve the crystal structure of a retroviral protease (Khatib *et al.* 2011)



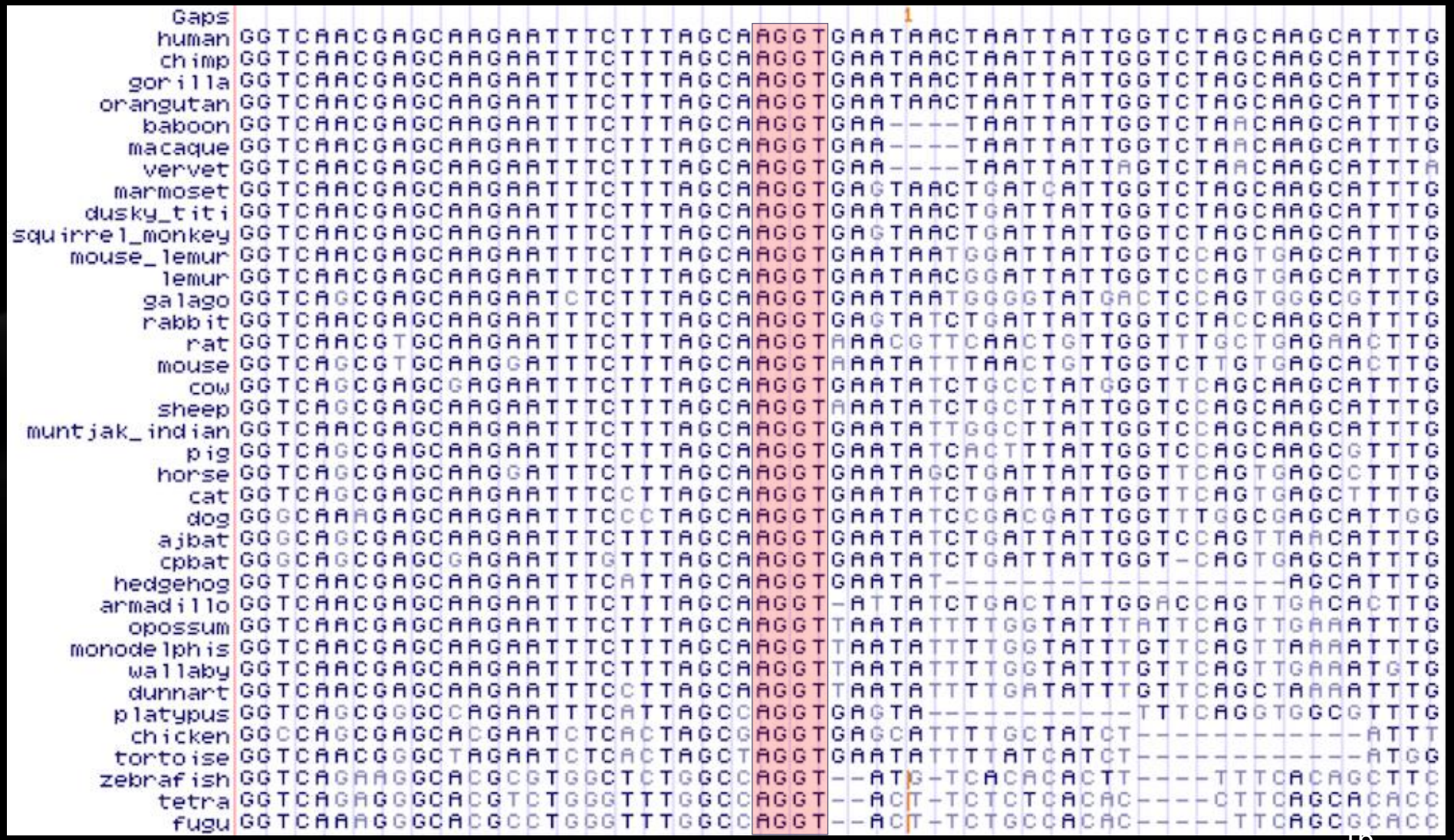
- It is not a casual game:
  - Complex interface
  - Requires a certain level of scientific knowledge

http://fold.it/portal/

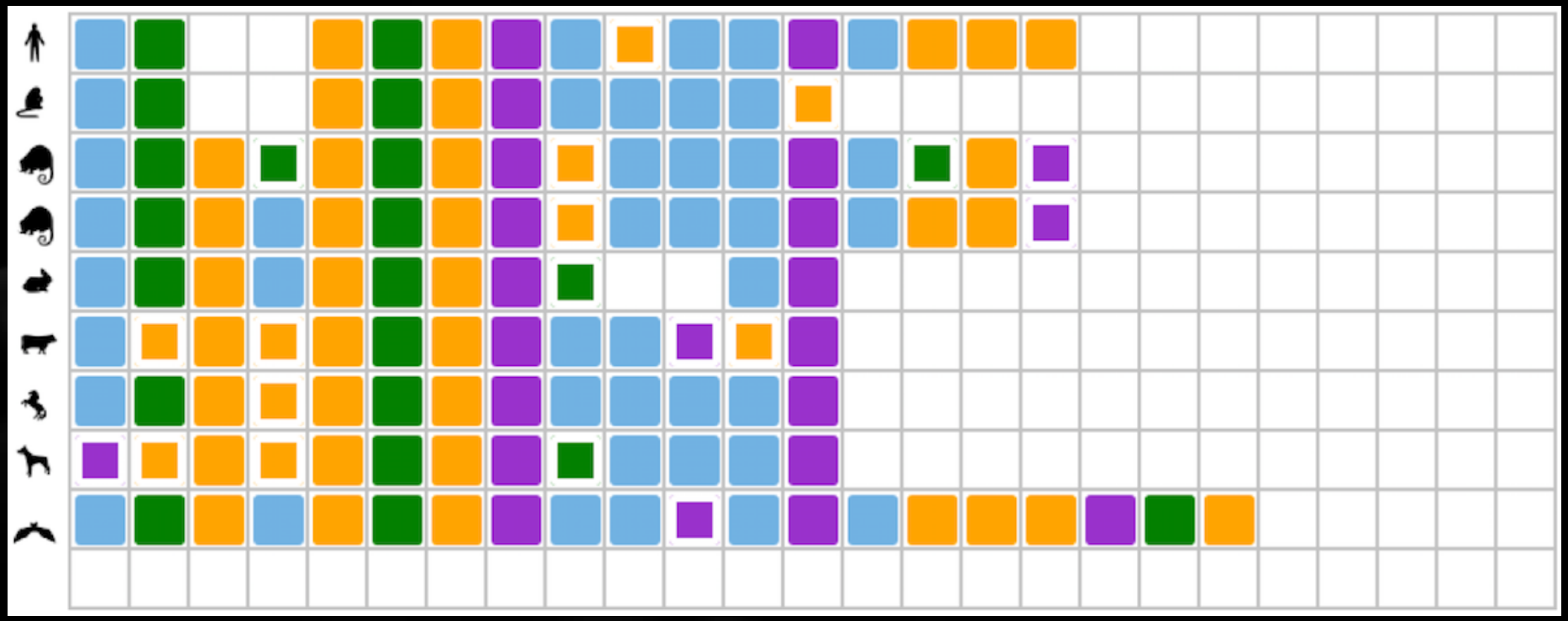
Phylo

- In 2010, Phylo was released
- It is tackling a fundamental problem in comparative genomics: Multiple Sequence Alignment (MSA)
- It converts the MSA problem into a casual game that can be played by ordinary web users

Comparative Genomics - MSA



Phylo DNA Puzzles



Turn the multiple sequence alignment problem into a casual tile-matching game.

The 3 Pillars of Phylo

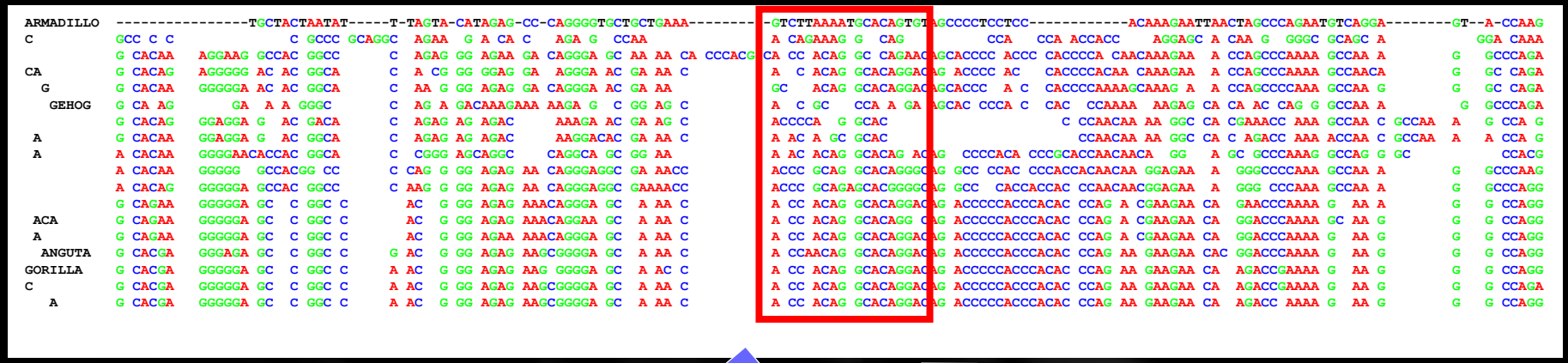
Why? Fundamental problem in molecular biology.

What? Well-defined problem proven difficult for computers.

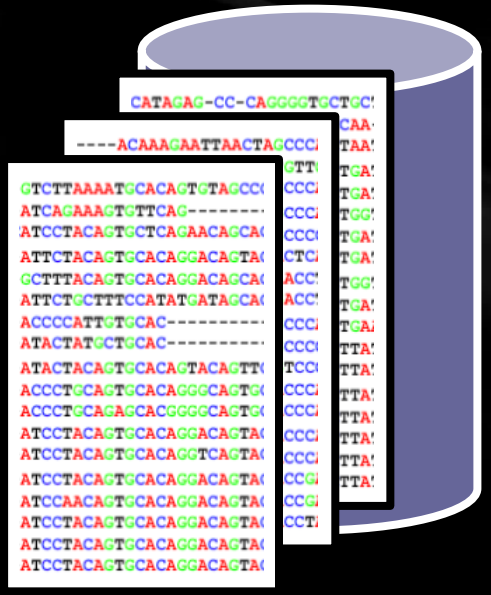
How? Humans are good at matching colors.

http://thegoddamn90s.com

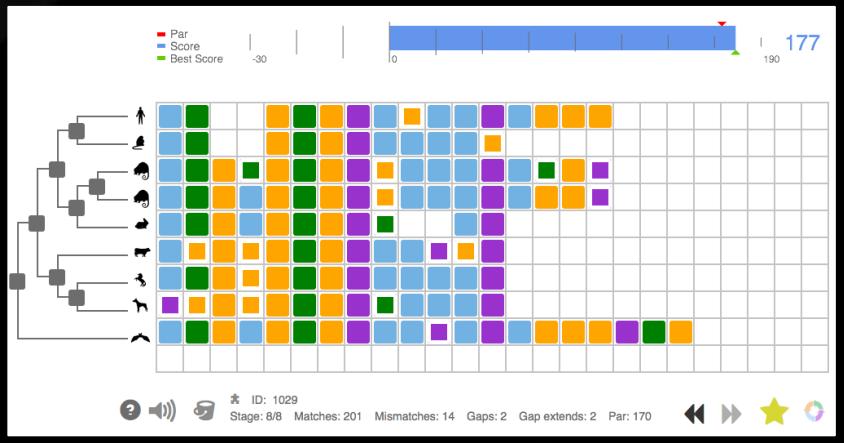
Whole-genome multiple alignment calculated with computers



Extract dubious alignment region



Video game:
- Computers
- Tablets



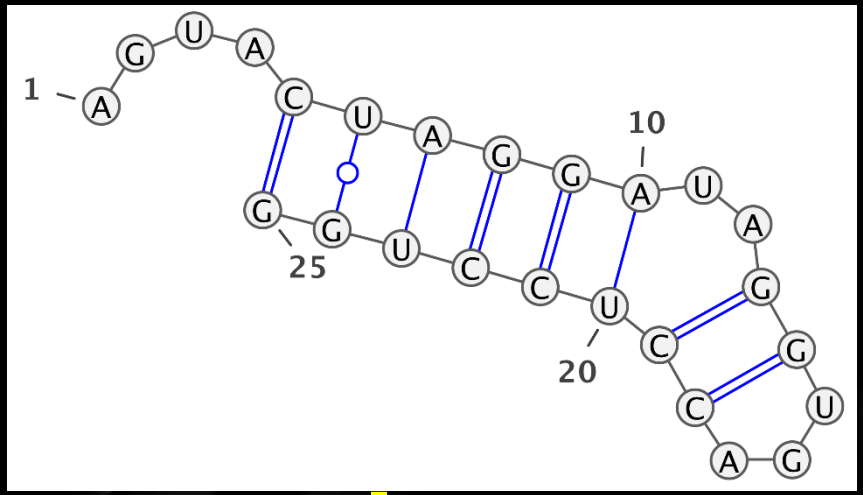
Reinsertion into original alignment + Evaluation

Database of interesting puzzles

http://phylo.cs.mcgill.ca

RNA sequence & structure

2D structure:



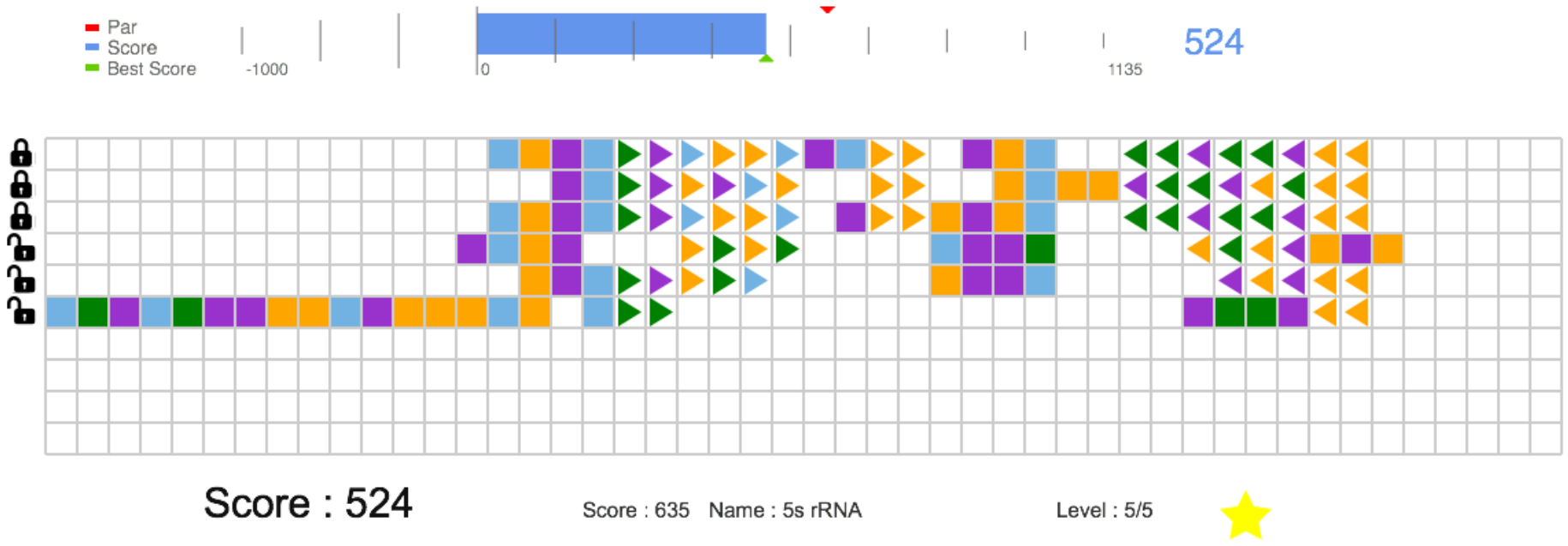
Bracket notation:

AGUACUAGGAUAGGUGACCUCCUGG

. . . . ((((((. . ((. . .))))))))

Game representation:

Ribo



http://ribo.cs.mcgill.ca

Still not convinced about the potential of human computing?

Still not convinced about the potential of human computing?

Collaborative solving

- Using a human computing game to solve a problem that has a large search space is not straightforward:

1. Overwhelming for a single player

2. Impossible to find an optimal solution without considering all the data

Collaborative solving

- How to deal with a large search space?

Collaborative solving

- How to deal with a large search space?

A. Decompose the problem, distribute small tasks, aggregate the answers (AMT, Crowdcrafting):
   - No interactions
   - Trying to limit groupthink
   - Cannot benefit from collective intelligence

Collaborative solving

- How to deal with a large search space?

A. Decompose the problem, distribute small tasks, aggregate the answers (AMT, Crowdcrafting):
   - No interactions
   - Trying to limit groupthink
   - Cannot benefit from collective intelligence

B. Build a collaborative environment:
   - Interactions allowed but controlled
   - Promote cooperation
   - Allow exchange of information
   - Improve the solutions of others

Collaborative solving

- Before transitioning to collaborative models in games, we need to:
  - Estimate the potential gains in productivity
  - Quantify the usefulness of different mechanisms

Collaborative solving

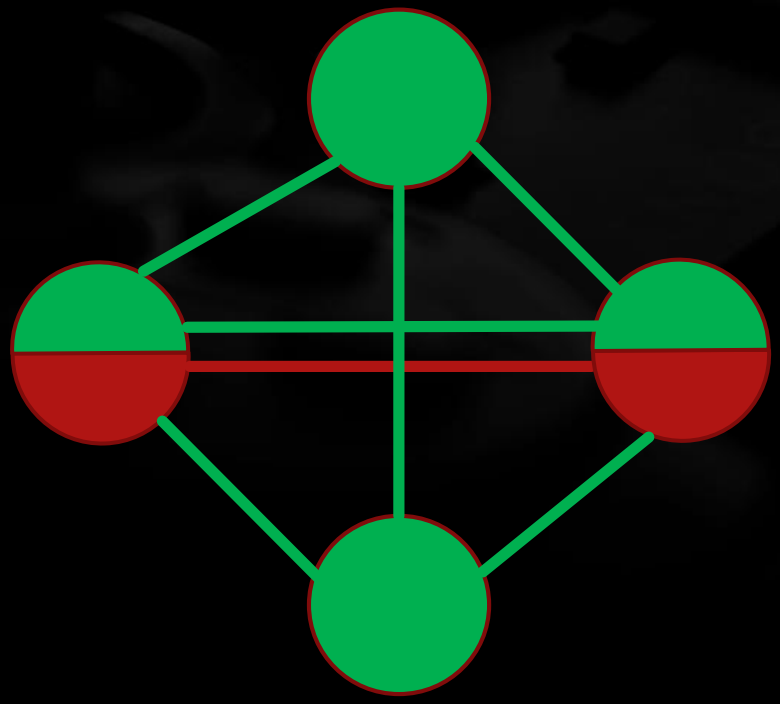
- Before transitioning to collaborative models in games, we need to:
  - Estimate the potential gains in productivity
  - Quantify the usefulness of different mechanisms
- We propose a formal framework to study human collaborative solving in a video game with:
  - A market
  - Skills
  - Challenges

Hypotheses

1. A market system will help players build longer solutions
2. Skills are useful to guide the players
3. A challenge system can encourage players to do certain actions
4. Collected solutions are better when all the three features are present

Problem

- Problem we want to solve is equivalent to finding maximal cliques in a colored multigraph



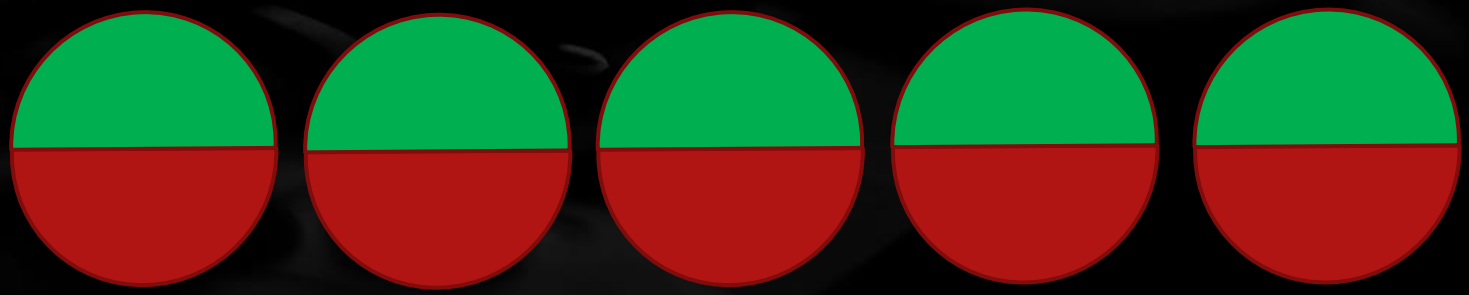
Exact solution:
$O(|V|*2^{|C|})$

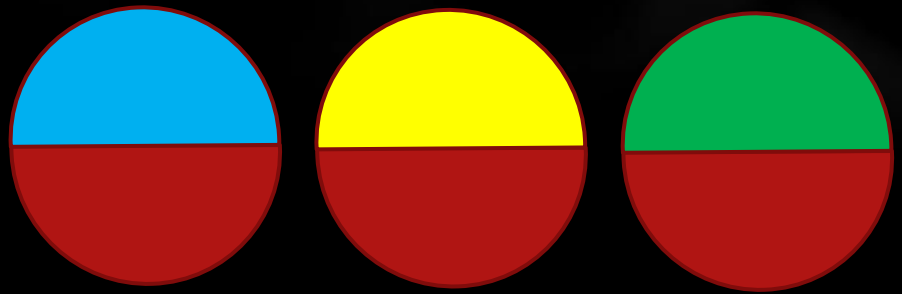
Problem

- The problem can actually be solved quickly by algorithms
- Allows us to find the exact solutions and evaluate the performance of players

Goal of the game

- Goal: build long sequences with many colors in common



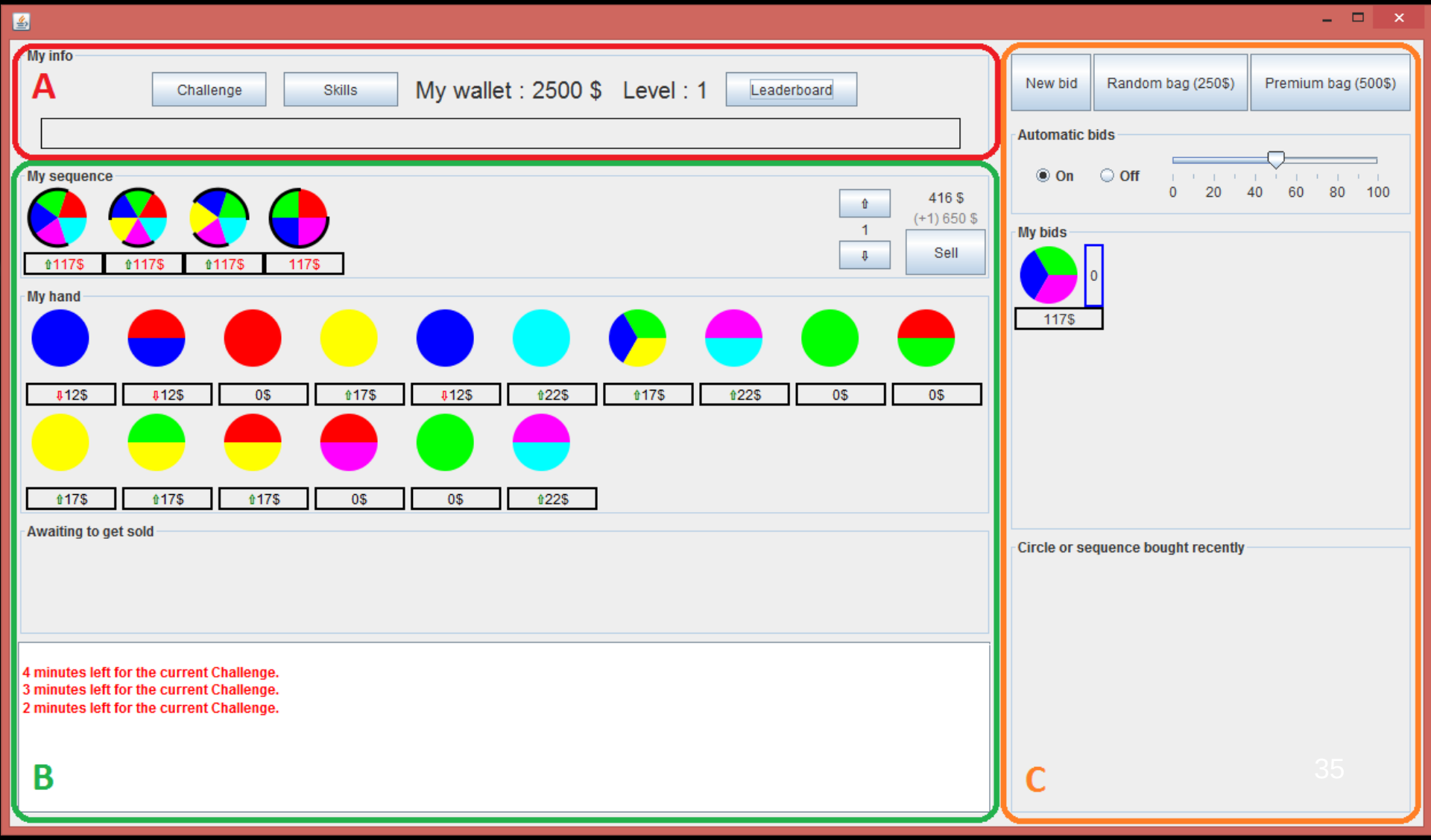


Parameters

- Scoring function: $baseScore_n$ * $seqLength^2$

Number of colors	$baseScore_n$
0	0
1	5
2	14
3	26
4	40
5	55
6	72

Game interface



Features

- Market:
  - Selling/buying individual circles
  - Buying sequences that have been sold to the system (buyout)

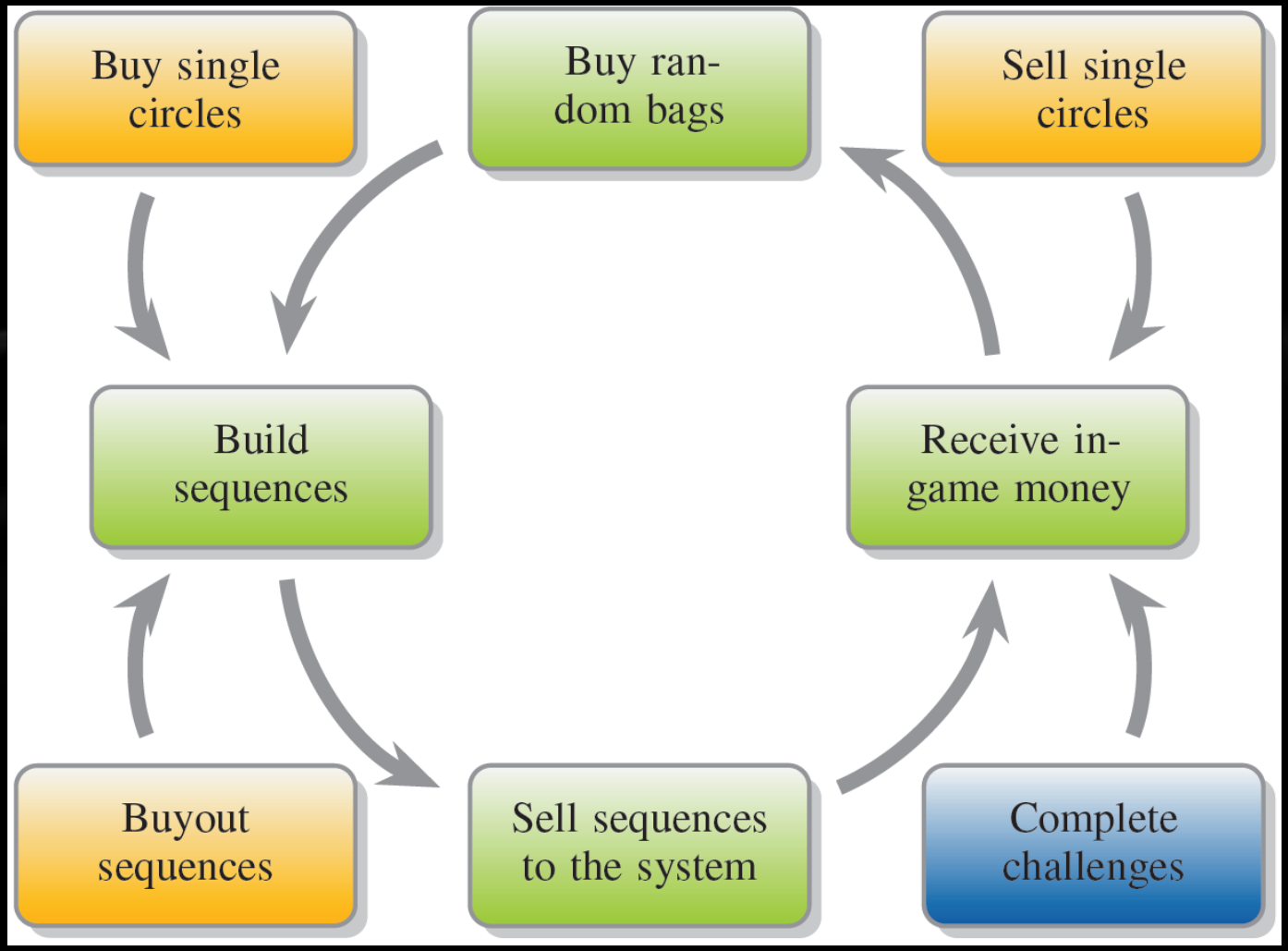
Features

- Skills:
  - Buyout King: lowers the price of buying a sequence from another player
  - Color Expert: gives a bonus to selling sequences that have more than one color in common
  - Sequence Collector: gives an additional sequence slot
  - Master Trader: gives a bonus to selling circles to other players

Features

- Challenges:
  - Sell/buy circles
  - Buyout sequences
  - Minimum number of colors
  - Minimum sequence length
  - Specific colors in common

Gameplay loop



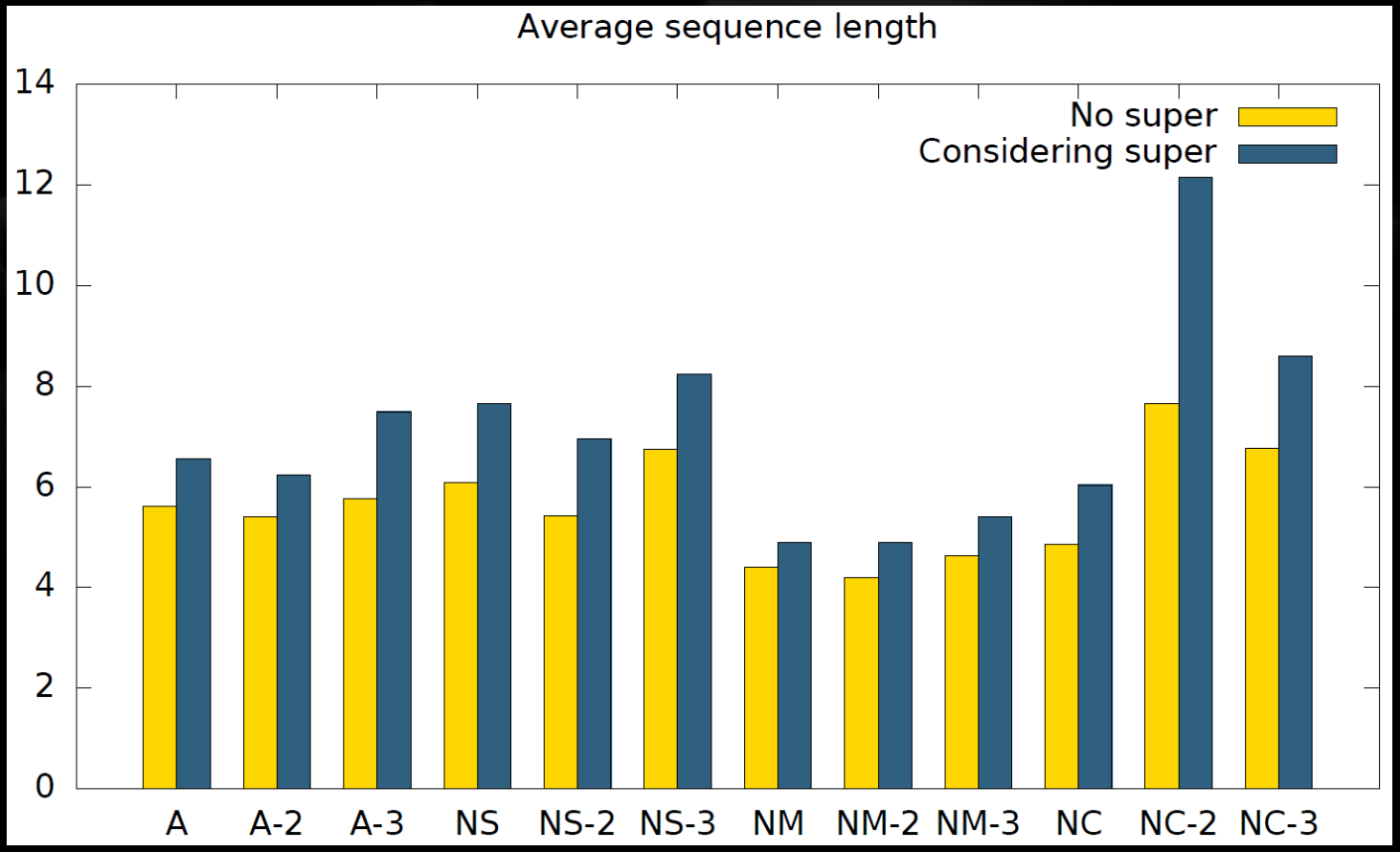
- Normal actions
- Market-related actions
- Challenges

Experiments

- Generated a graph of 300 vertices and 6 colors
- We recruited 12 groups of 10 people
- We tested 4 conditions (3 times each):
  1) All (skills + market + challenges)
  2) No skills (market + challenges)
  3) No market (skills + challenges)
  4) No challenges (skills + market)

Testing hypothesis 1 (market)

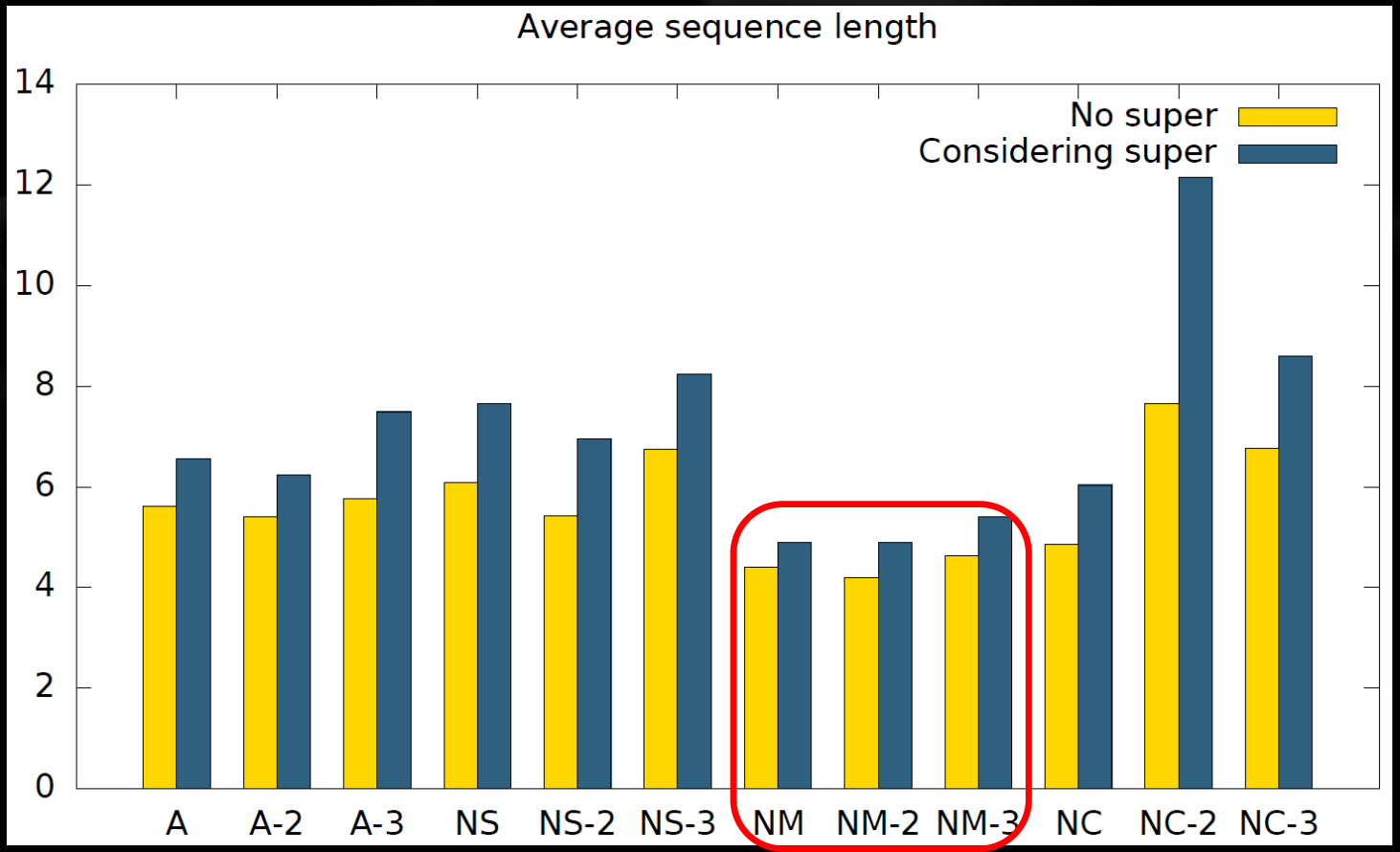
- Presence of the market and sequence length



Kruskal-Wallis test
$p < 2.2E-16$

Testing hypothesis 1 (market)

- Presence of the market and sequence length



Kruskal-Wallis test
$p < 2.2E-16$

Testing hypothesis 1 (market)

- Presence of the market and sequence length

Similar groups (Dunn's test):

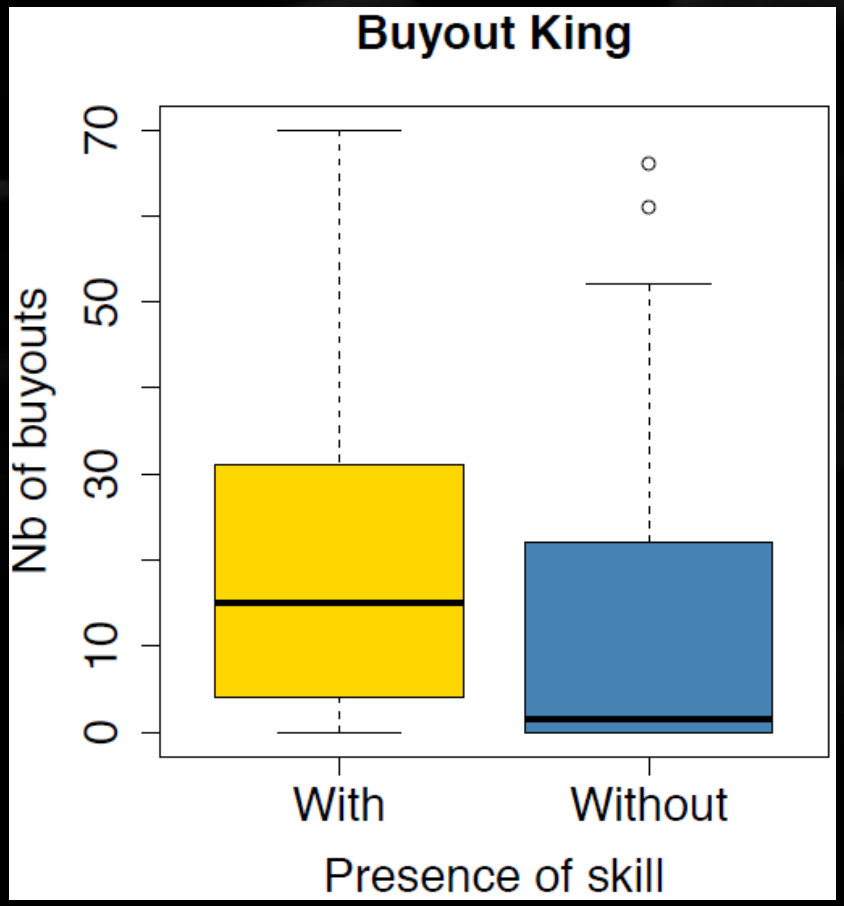
	A-2	A-3	NS	NS-2	NS-3	NM	NM-2	NM-3
A	n/s	n		n/s				
A-2		n		n/s				
A-3			n/s					
NC						n		n/s
NC-3					n/s			
NM							n/s	n

Testing hypothesis 2 (skills)

- The skills have an effect on the players' strategies:
  - Buyout King (affects nb buyouts)
  - Master Trader (increases the nb of circles sold)
  - Color Expert (increases the proportion of sequences with many colors)
  - Sequence collector (increases the sequence length and nb of colors)

Testing hypothesis 2 (skills)

- Buyout King skill and number of buyouts

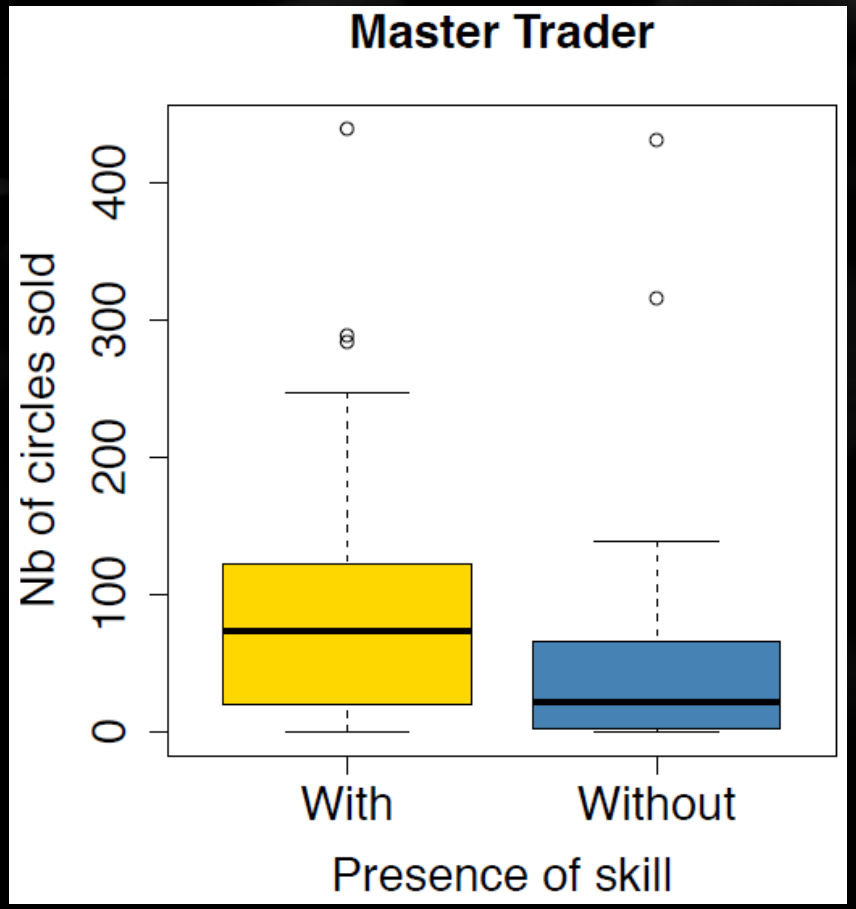


Median values:
With = 15
Without = 1.5

Mann-Whitney's U test:
p = 0.004, effect size r = 0.28

Testing hypothesis 2 (skills)

- Master Trader skill and number of circles sold

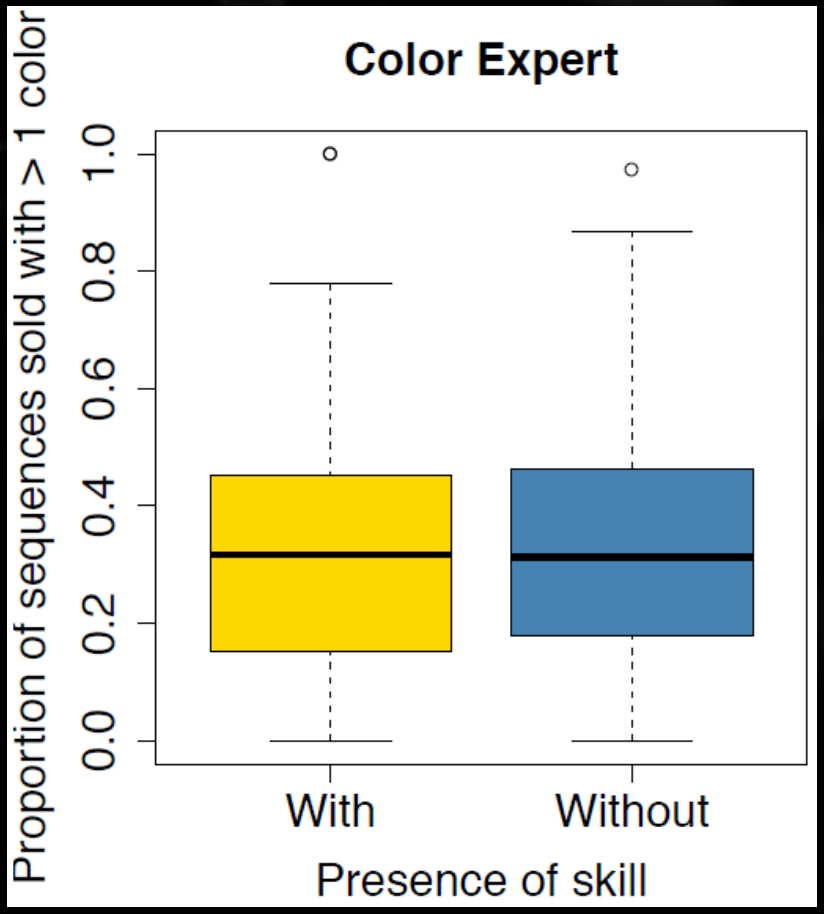


Median values:
With = 73
Without = 21.5

Mann-Whitney's U test:
p = 7.2E-4, effect size r = 0.33

Testing hypothesis 2 (skills)

- Color Expert skill and proportion of multicolored sequences

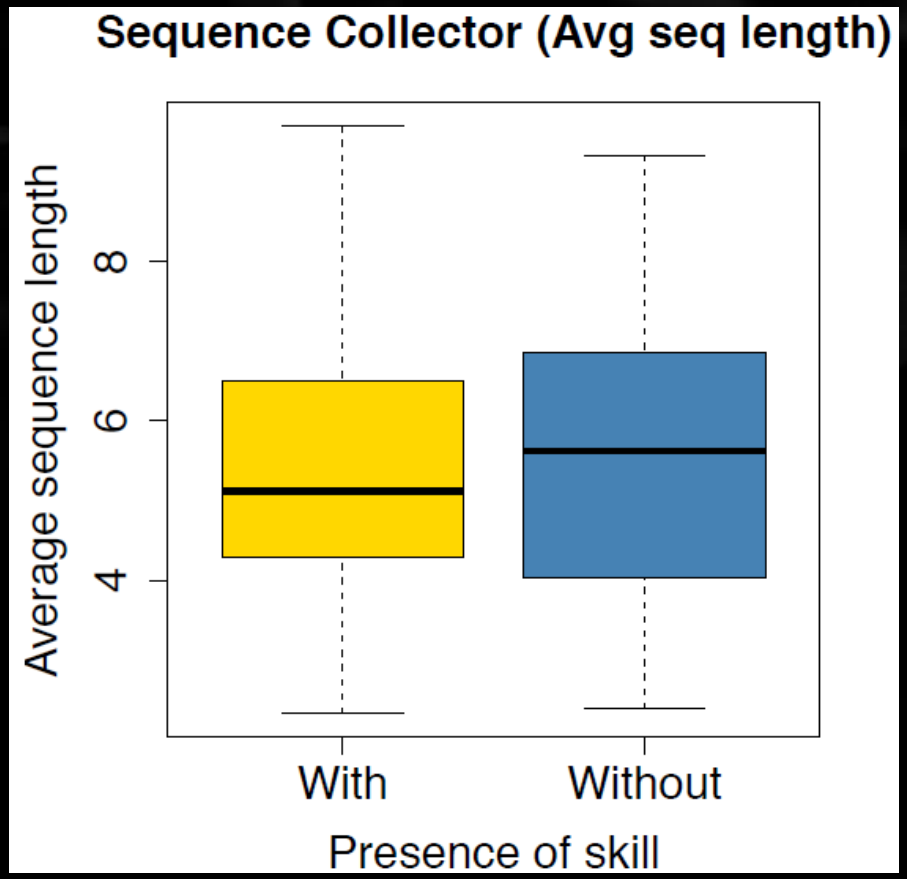


Median values:
With = 0.317
Without = 0.313

Mann-Whitney's U test:
p = 0.89 -> failure to reject the null hypothesis

Testing hypothesis 2 (skills)

- Sequence Collector skill and average sequence length

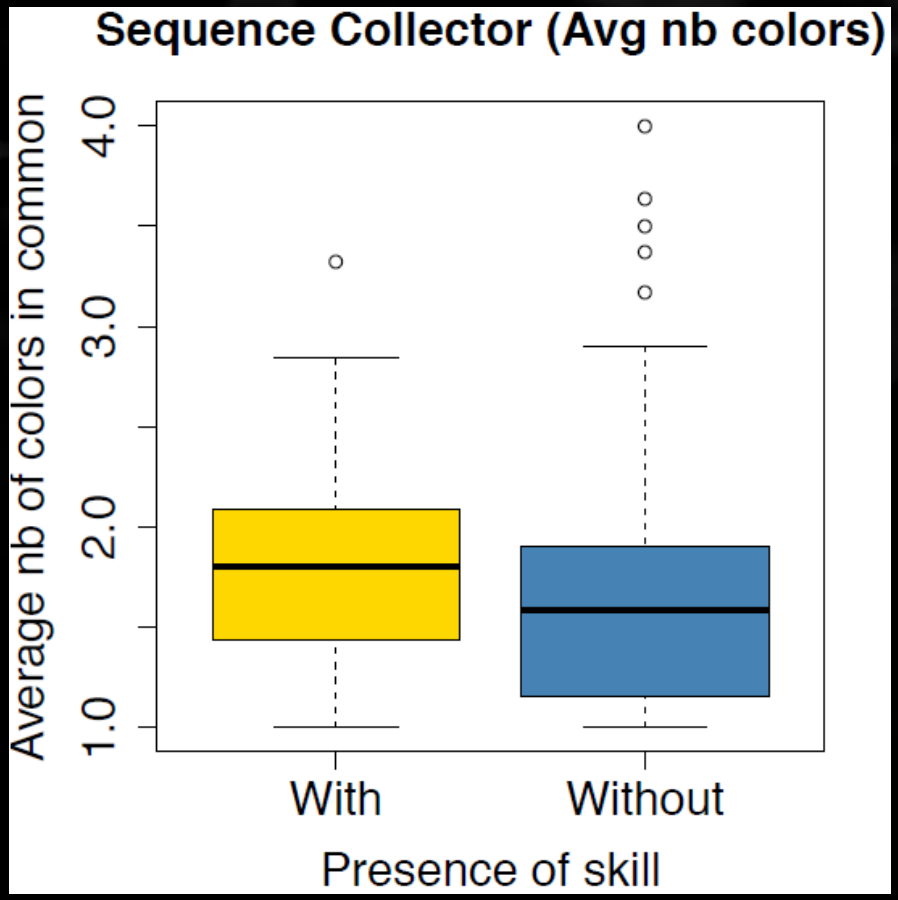


Median values:
With = 5.12
Without = 5.63

Mann-Whitney's U test:
p = 0.69 -> failure to reject the null hypothesis

Testing hypothesis 2 (skills)

- Sequence Collector skill and average number of colors



Median values:
With = 1.80
Without = 1.58

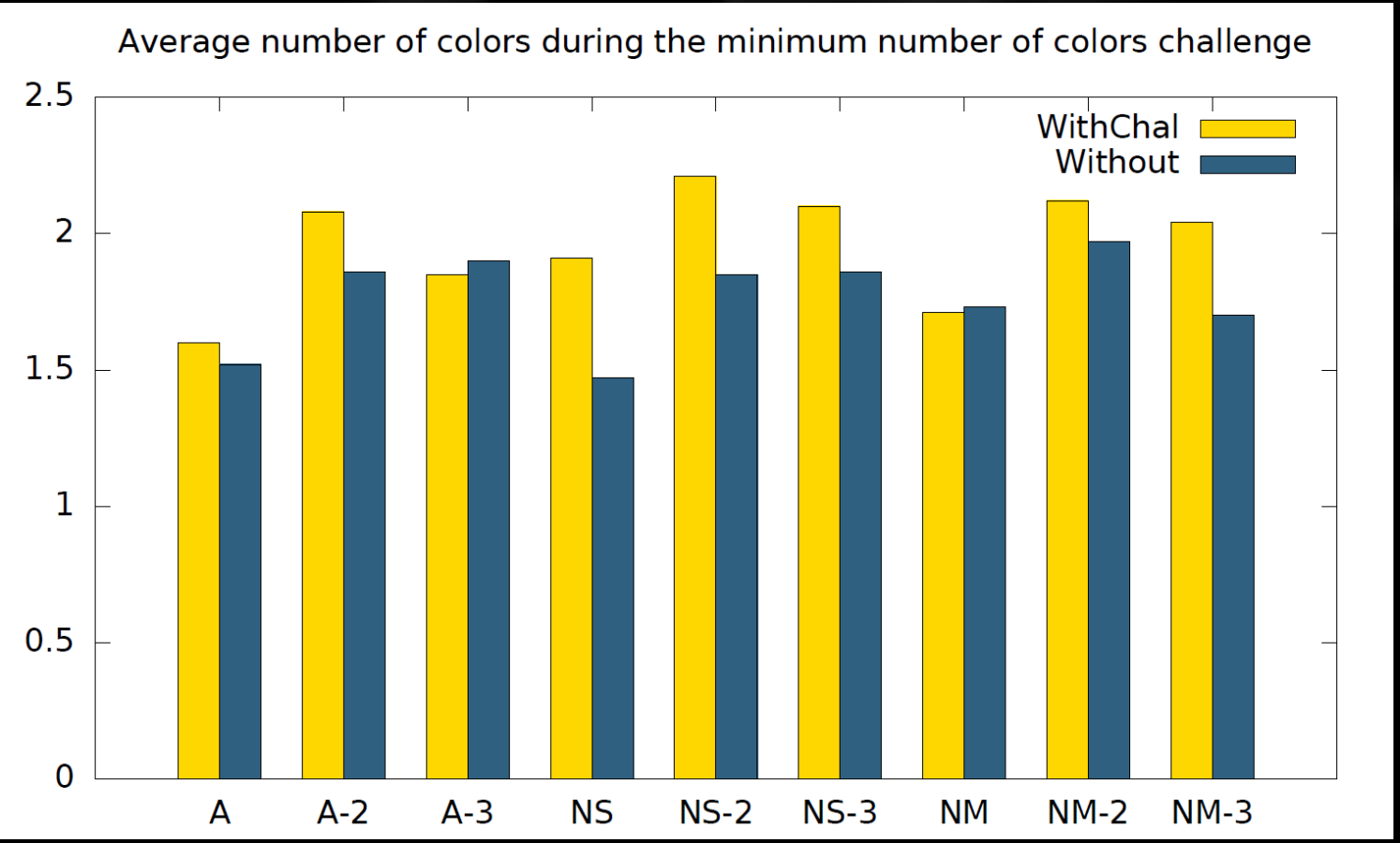
Mann-Whitney's U test:
p = 0.01, effect size r = 0.21

Testing hypothesis 3 (challenges)

- Challenges encourage players to do a certain action
  - Minimum number of colors challenge
  - Minimum sequence length challenge
  - Sell/buy challenge
  - Buyout challenge
  - Specific colors in common challenge

Testing hypothesis 3 (challenges)

- Minimum number of colors challenge

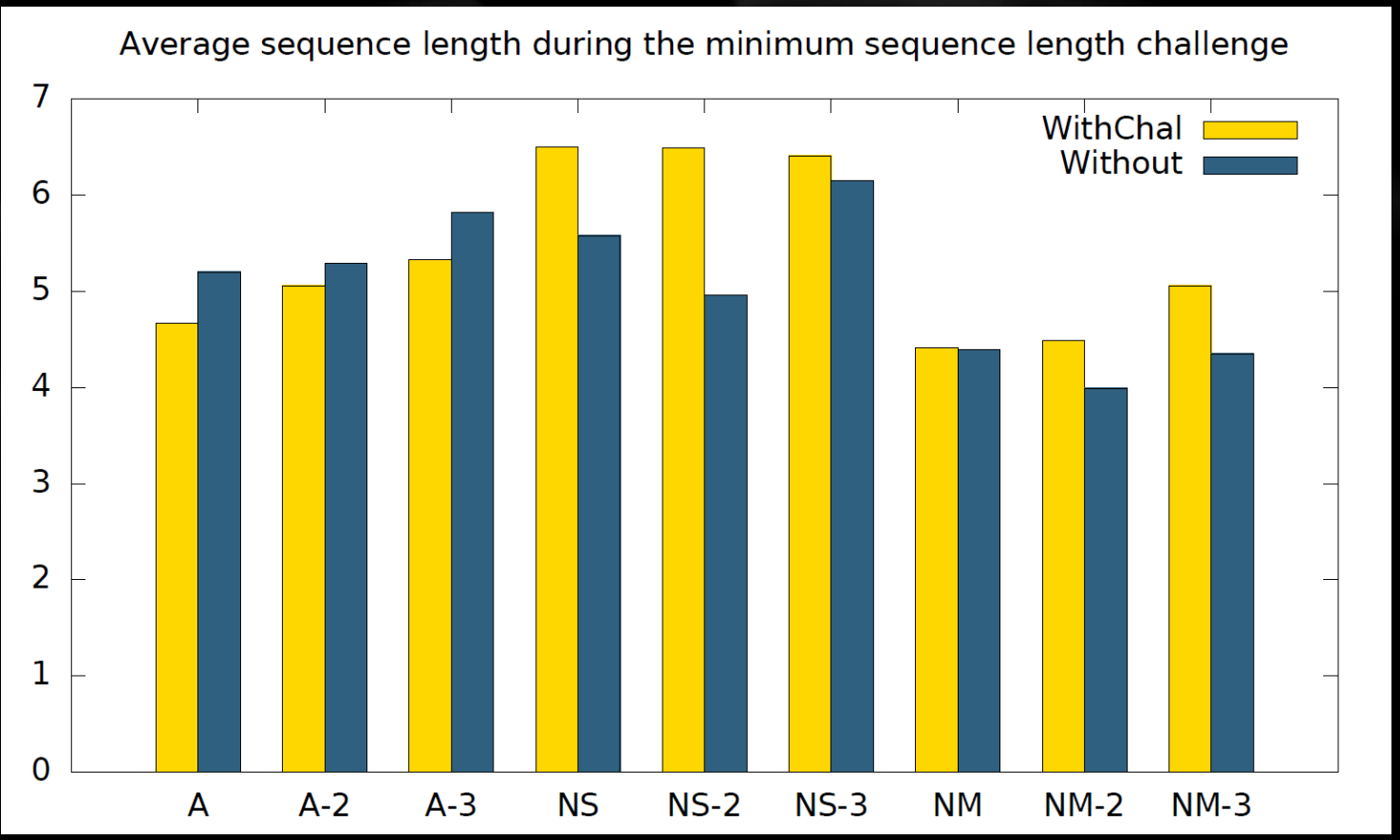


Means:
During chal. = 1.96
Rest of time = 1.76

Welch t-test:
p = 0.04,
Cohen's d = 1.03

Testing hypothesis 3 (challenges)

- Minimum sequence length challenge



Means:

During chal. = 5.38

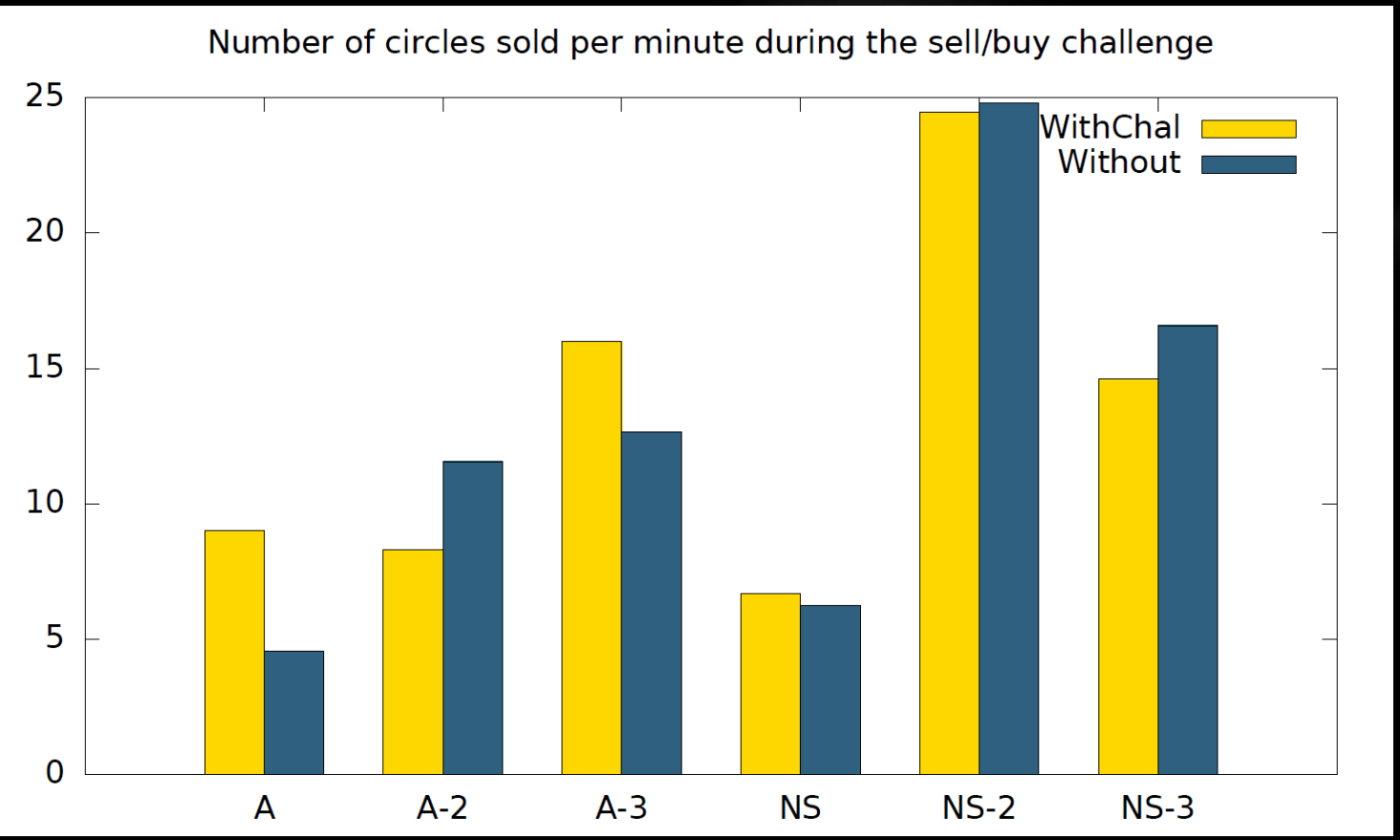
Rest of time = 5.08

Welch t-test:

p = 0.44 -> failure to reject the null hypothesis

Testing hypothesis 3 (challenges)

- Sell/buy challenge



Means:
During chal. = 13.18
Rest of time = 12.73

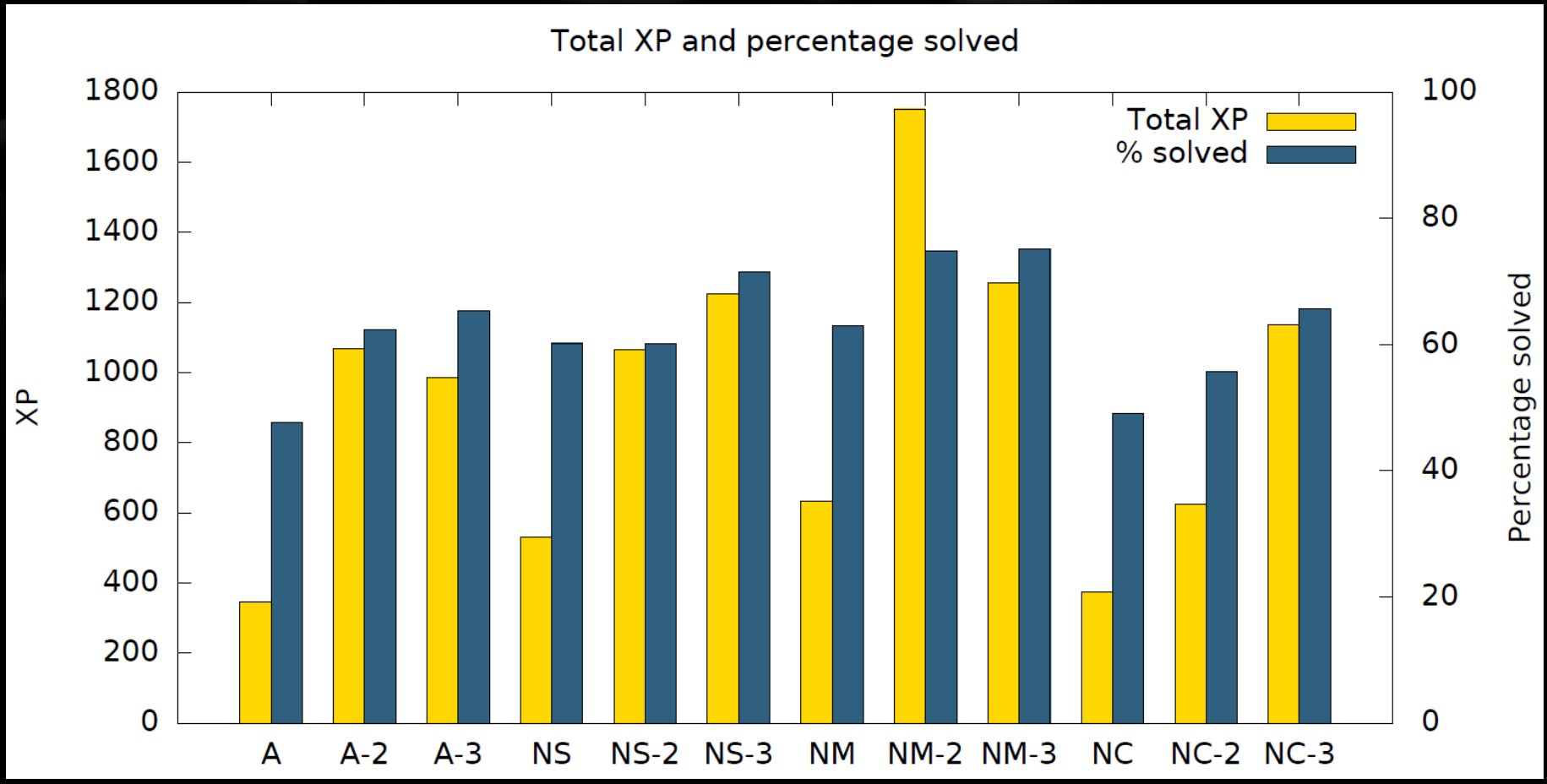
Welch t-test:
p = 0.91 -> failure to reject the null hypothesis

Testing hypothesis 3 (challenges)

- Buyout challenge: appeared only once in all the game sessions
- Specific set of colors in common challenge:
  - Appeared 11 times in total
  - Completed only 8 times

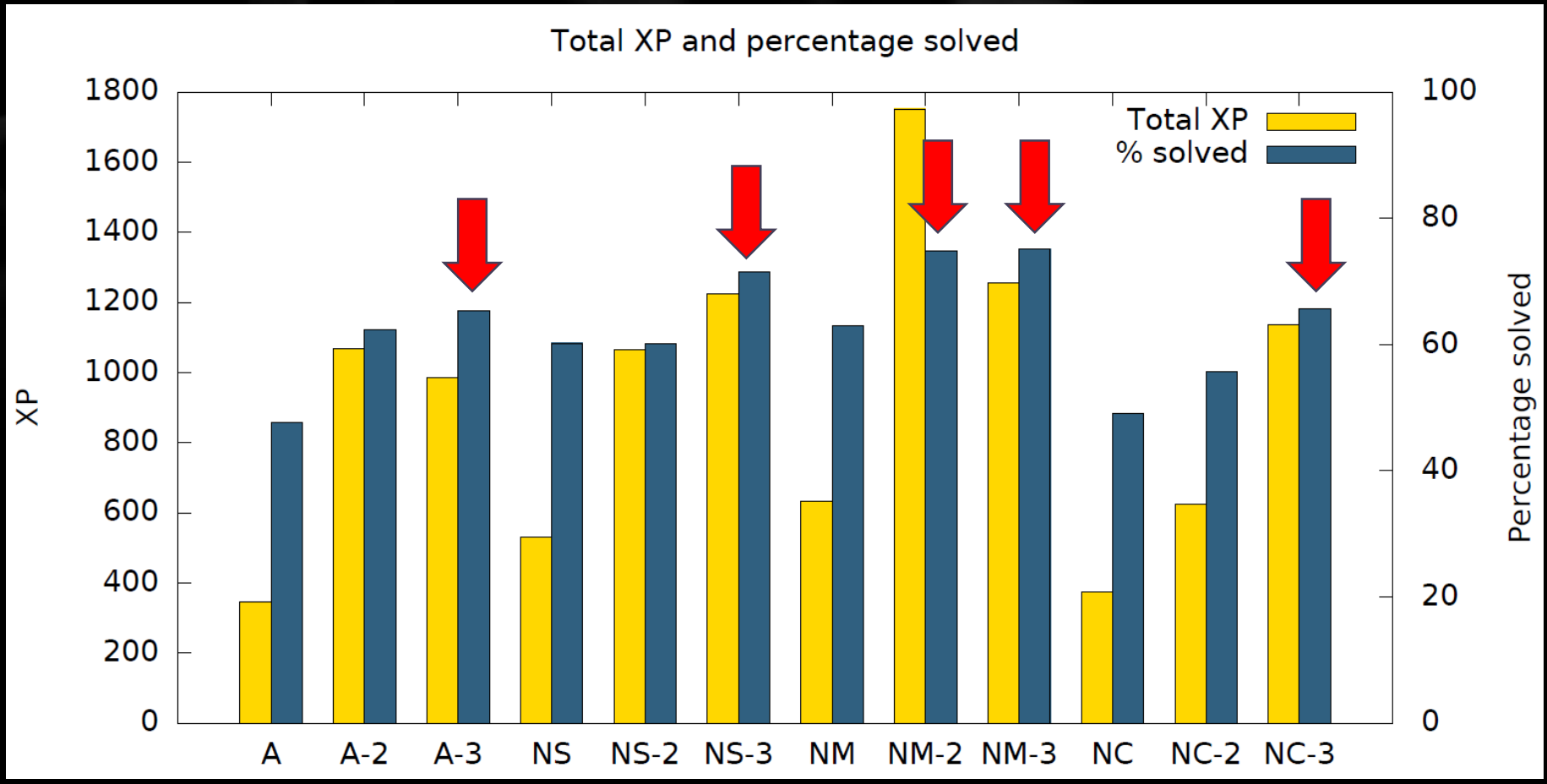
Testing hypothesis 4 (% solved)

- Is the percentage solved better when all the features are present?



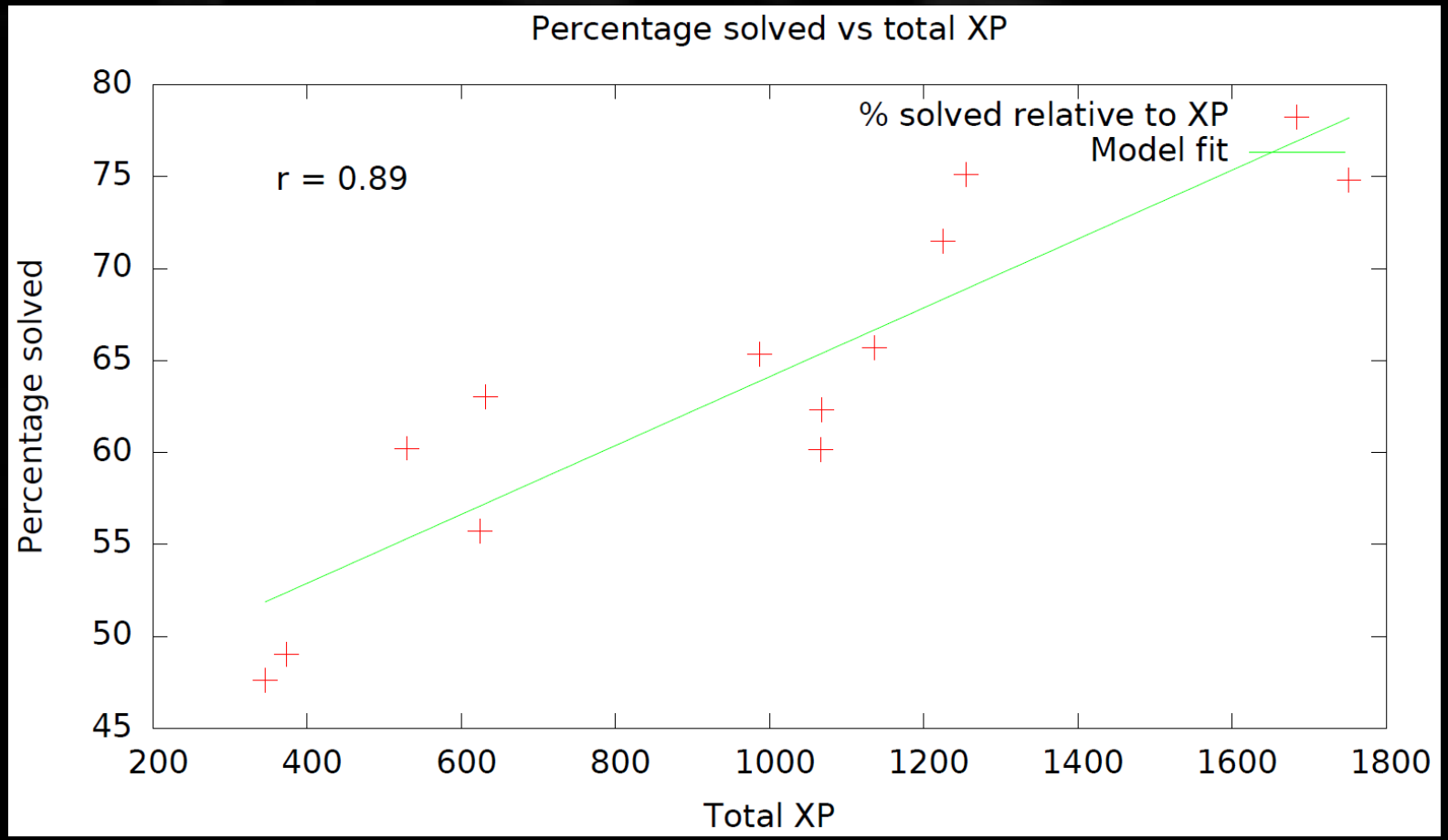
Testing hypothesis 4 (% solved)

- Is the percentage solved better when all the features are present?



Testing hypothesis 4 (% solved)

- Is the percentage solved better when all the features are present?

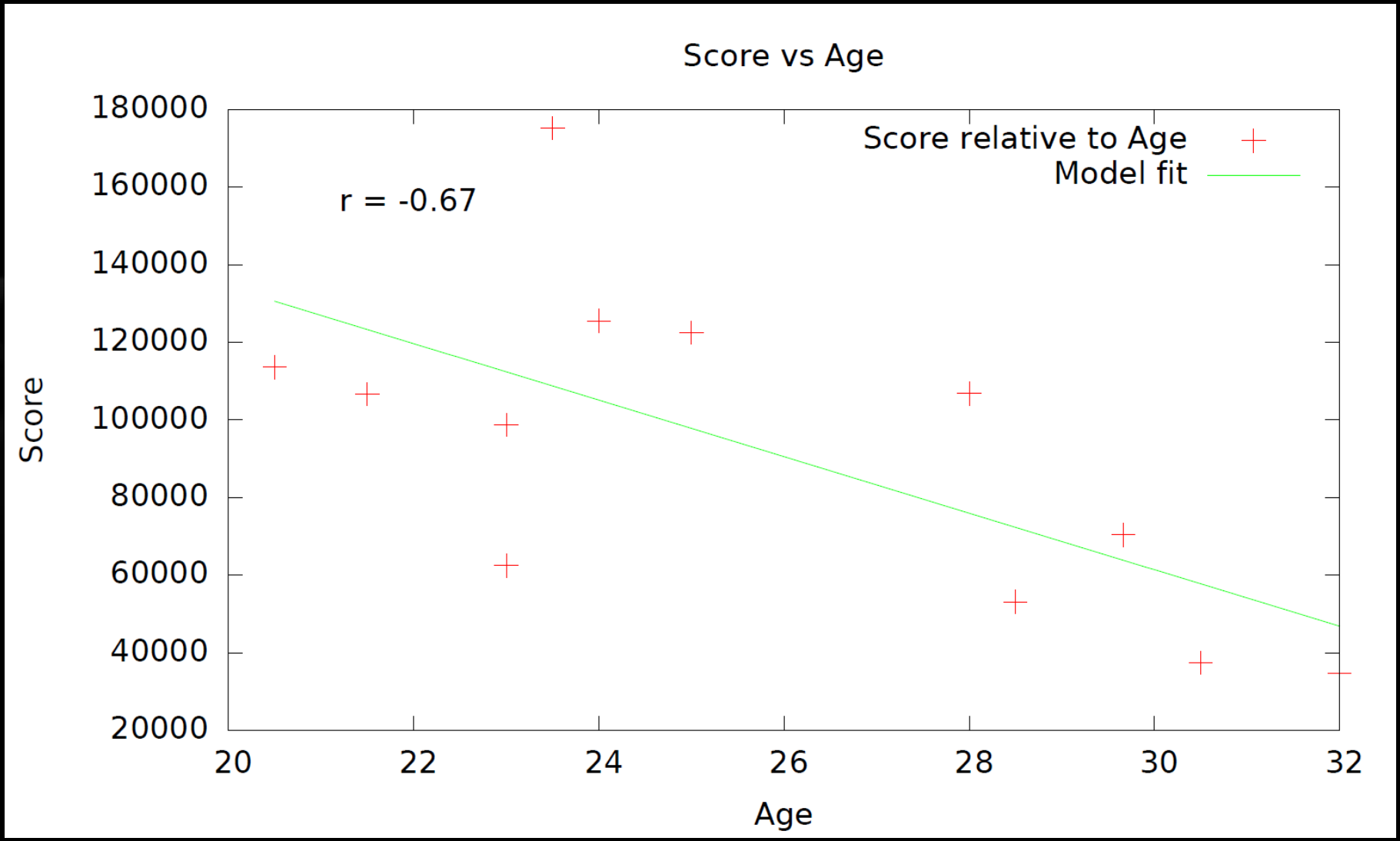


What makes a good player

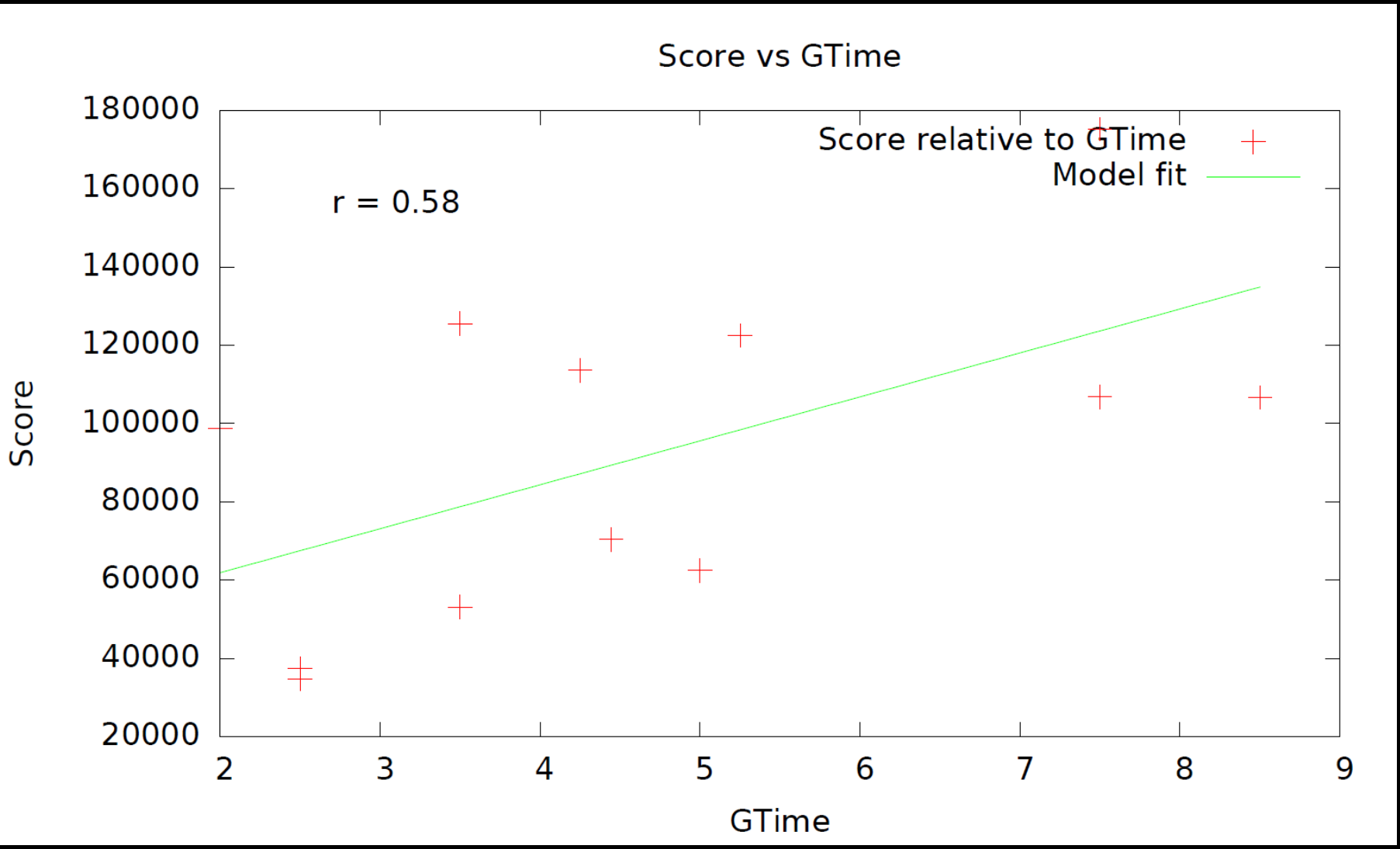
- Top 12 players VS the others:

	Top 12	Others
Age	23.42	25.99
Self evaluation	3.67	2.90
Game time / week	10.00	4.11

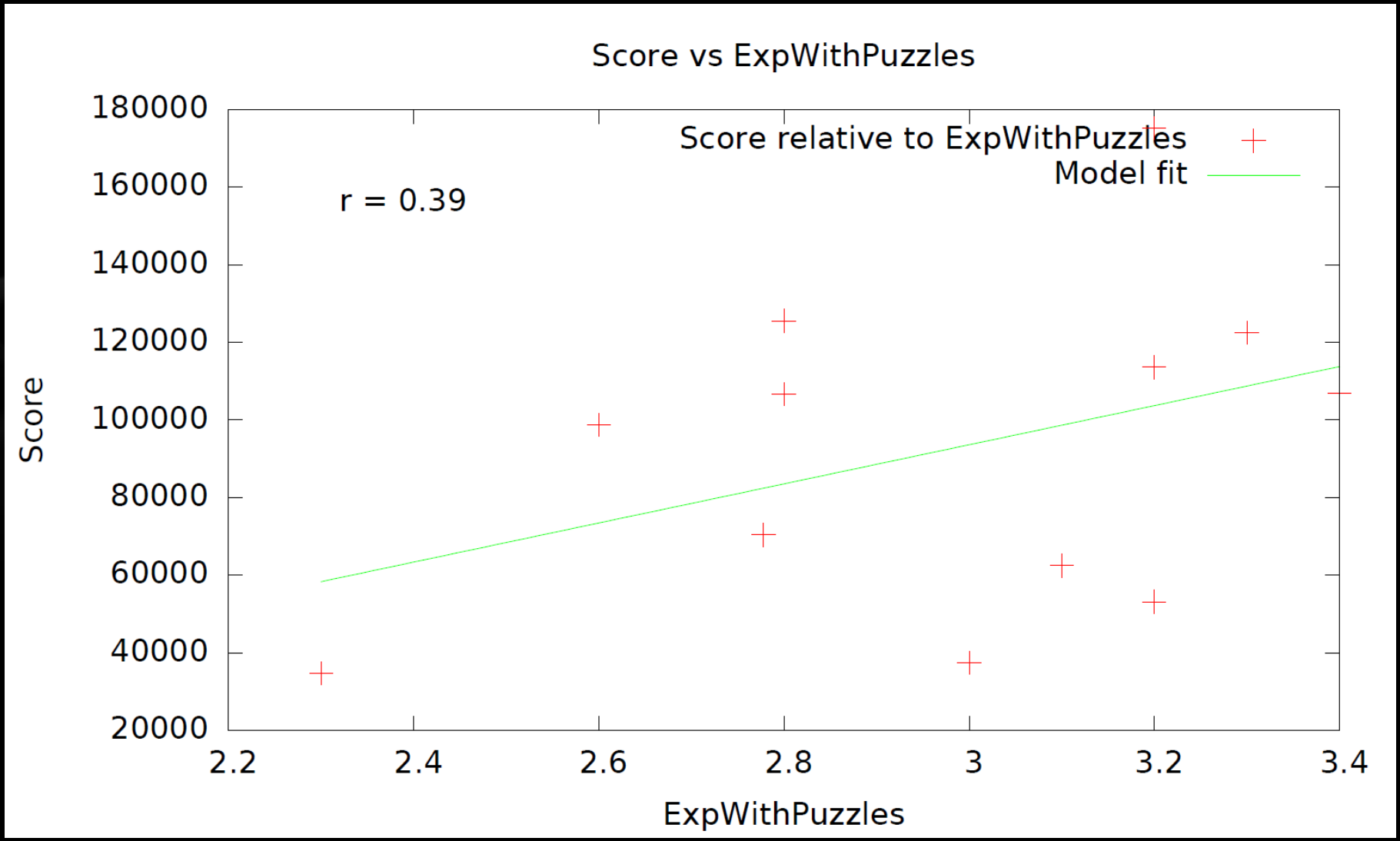
What makes a good player



What makes a good player



What makes a good player



Conclusions

- A market system is a useful tool to help players build longer solutions
- Skills are helpful to guide the players into doing specific actions in the game
- Well-balanced challenges in terms of difficulty can promote a certain action from the players
- Percentage of the problem solved depends not only on the features present in the game, but also on the players' skills

Conclusions

- Players' satisfaction:
  - Average score of 7.16/10
  - Very addictive
  - Leaderboard = powerful motivation
  - Some participants found the game too easy / hard

Conclusions

- Scalability:
  - Would we get similar results with more players?
  - Probably: in the tests during development, we noticed that having more players only helped
  - Future work: development of an online version that would be available 24/7

More future work

- Verify if the results are task dependent
- Build bots (AI) and test their performance
- Define an optimal crowd of players with different strategies
- Switch to a problem that makes a better use of human skills

Crowdsourcing genomic databases

- Development of a web and mobile crowdsourcing platform for curating genomic databases
- Why: more and more data, manual curation is necessary, but funding is limited

Crowdsourcing genomic databases

1. Develop crowdsourcing systems for curating, maintaining and updating genomic databases

2. Develop learning/teaching interface

3. Explore new technologies: virtual reality (3D genome browser)

Acknowledgments

- Jérôme Waldispühl
- Alexander Butyaev
- Jean-François Bourbeau
- Chris Drogaris
- Mathieu Blanchette
- Derek Ruths
- Edward Newell

- Alexandre Leblanc

Game sessions:

- Silvia Leon
- Shu Hayakawa
- All the players

Financial support: