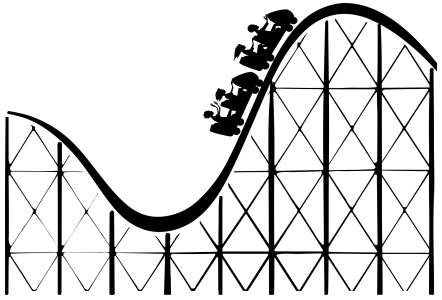
Roller Coaster Track Design Behavioral Experiment



Murtuza Shergadwala
Jitesh H. Panchal

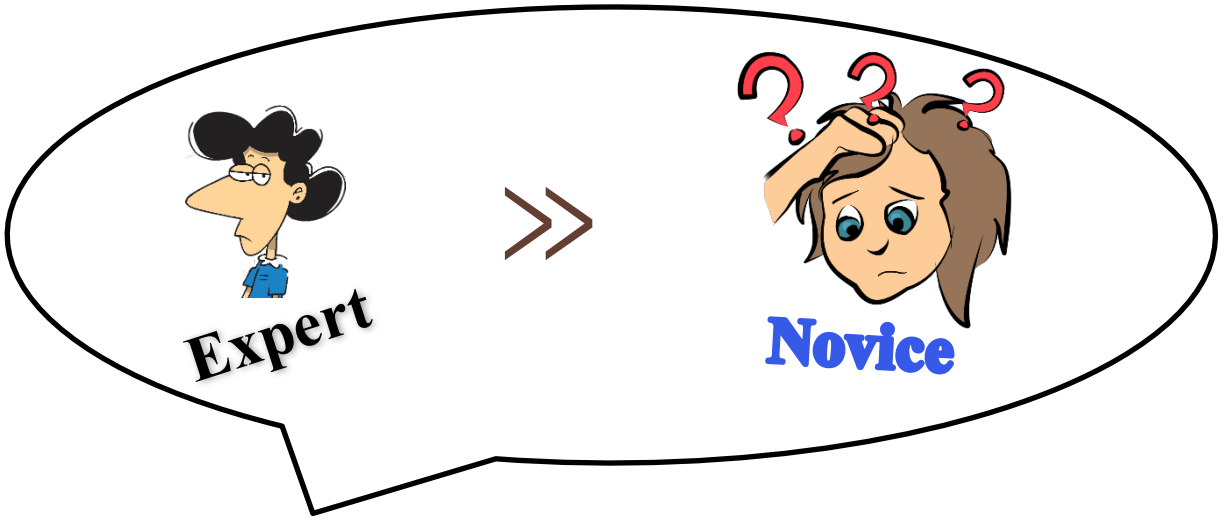
Karthik N. Kannan

School of Mechanical Engineering
Purdue University

Krannert School of Management
Purdue University

Research Question

We know…

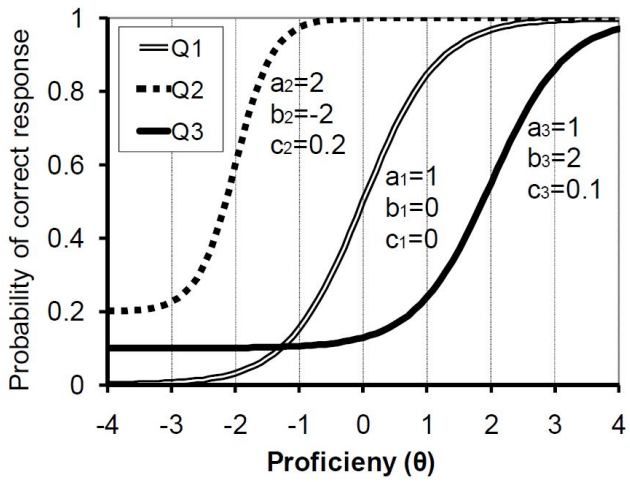


However,

How can we quantify the impact of expertise on solution quality?

Expertise Quantification: State of the Art

Item Response Theory (IRT) is a psychometric assessment model that is based on the idea that "the probability of a correct/keyed response to an item/question is a mathematical function of person and item parameters."



3 Parameter Item Response Model [1]

IRT helps in formulating characteristic curves for a multiple choice question. It gives the probability of a person, with a given proficiency, answering the i^{th}question correctly.

$$P_i(\theta) = c_i + \frac{1 - c_i}{1 + \exp[-1.7a_i(\theta - b_i)]}$$

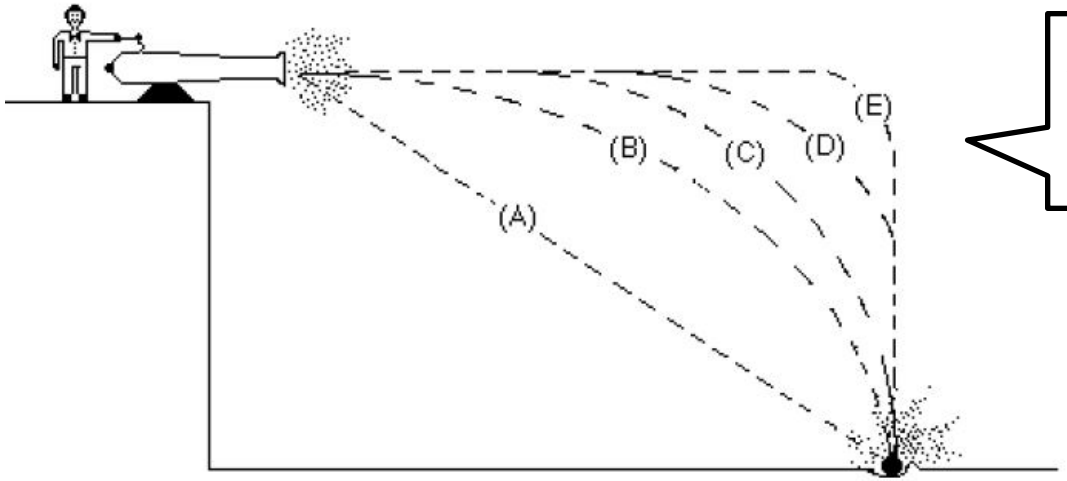
$$a = discrimination\ parameter$$
$$b = difficulty\ parameter$$
$$c = chance\ parameter$$

[1] Wang, J., & Bao, L. (2010). Analyzing force concept inventory with item response theory. *American Journal of Physics*, *78*(10), 1064-1070.

Concept Inventories (CIs)

Concept Inventories (CIs) is a form of test, whose scores provide a quantification of an examinee's knowledge about a set of concepts. The Force Concept Inventory (FCI) [2] is one such CI that tests Newtonian concepts of force.

A ball is fired by a cannon from the top of a cliff as shown in the figure below. Which of the paths would the cannon ball most closely follow?

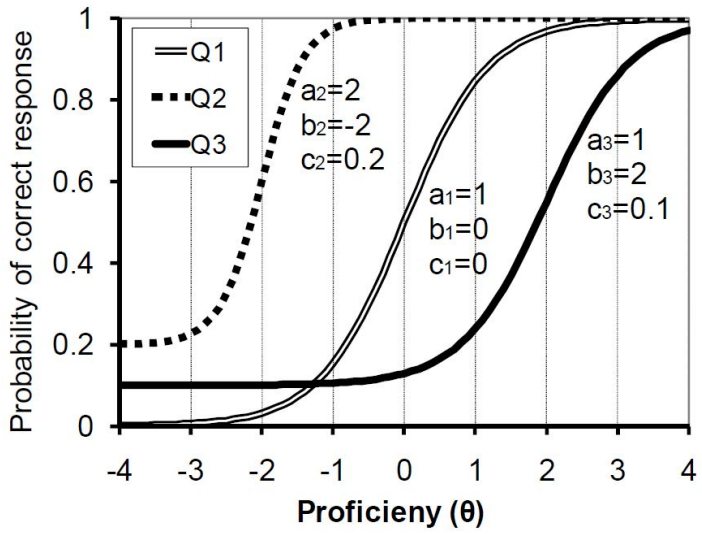


A sample question from the Force Concept Inventory. [2]

[2] Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The physics teacher*, *30*(3), 141-158.

Reliability of the CI Scores as an Expertise Metric

The IRT can be used to validate the CI. The work of Wang and Bao[1] utilize the FCI as an example CI and validate the FCI using IRT.



$$\boldsymbol{\theta} \approx 0.147\boldsymbol{S} - 2.324$$

The empirical equation that quantifies the relationship between proficiency from IRT and the FCI score

[1] Wang, J., & Bao, L. (2010). Analyzing force concept inventory with item response theory. *American Journal of Physics*, *78*(10), 1064-1070.

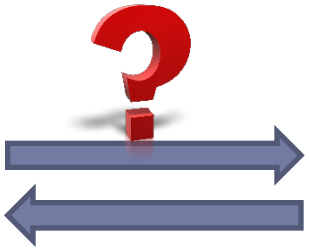
Why a Behavioral Experiment?

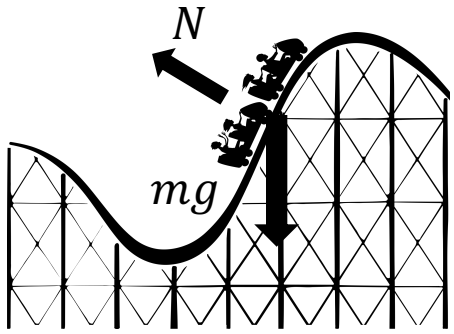
The state of the art theories help validate the expertise metric. However, the behavioral experiment will help verify the utilization of the expertise metric for quantification of expertise impact on quality.

Expertise metric θ evaluated through a test

Correlation between θ and performance

A roller coaster track design behavioral experiment

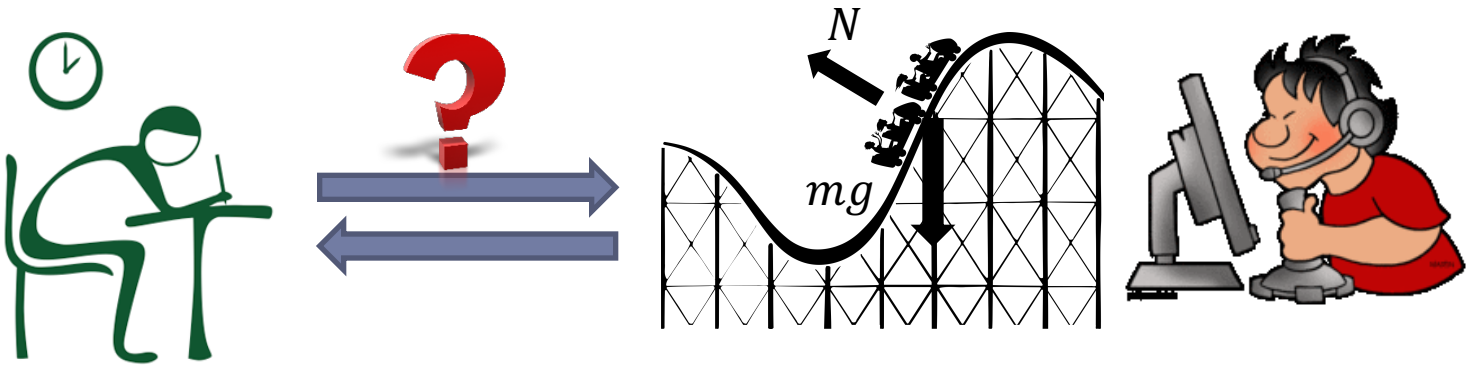




How do we correlate design performance with expertise metric θ?

Why a Track Design Behavioral Experiment?

As the expertise metric is based on the knowledge of Newtonian concepts of Force we **need a design task that encapsulates the tested concepts.** We thus formulate a **track design behavioral experiment**.

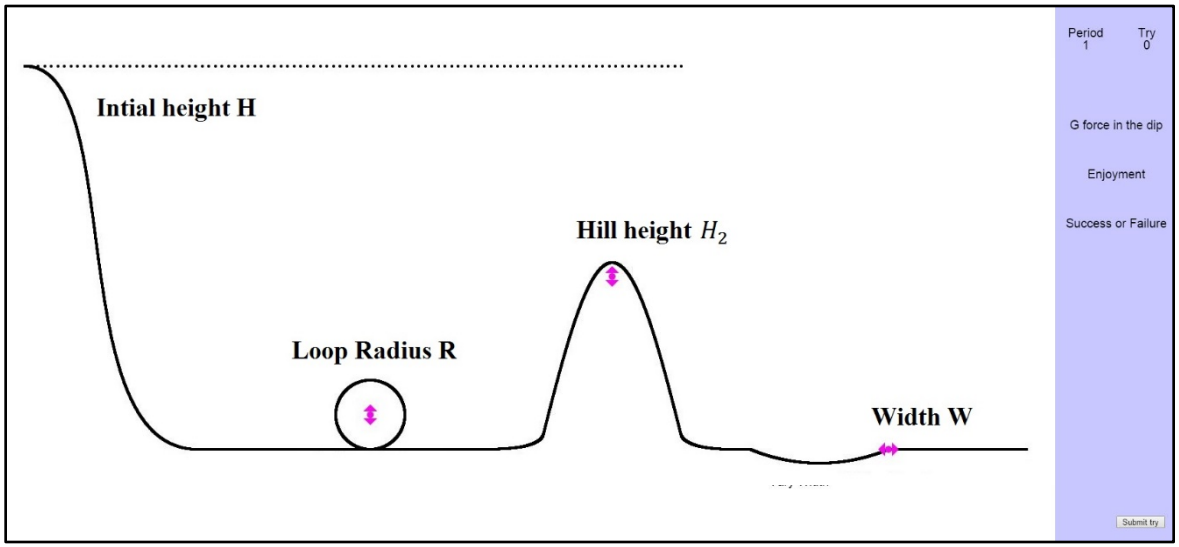


How do we correlate design performance with expertise metric θ?

Behavioral Experiment: Track Design Game

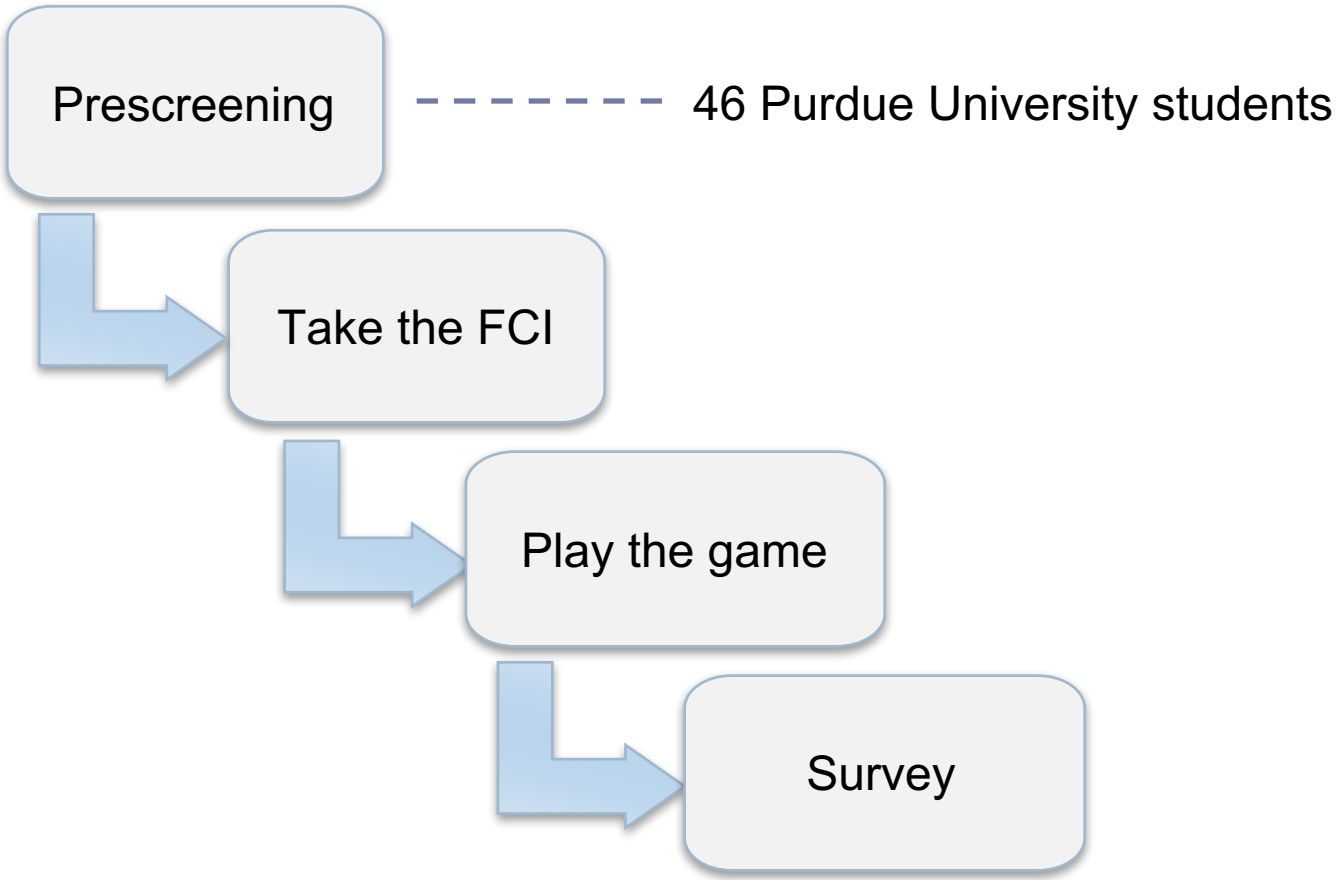
Objective: Maximize Enjoyment (Unknown function)

Constraints: 1) The person should not fly off the track
2) The person should not experience centripetal acceleration greater than $4g$



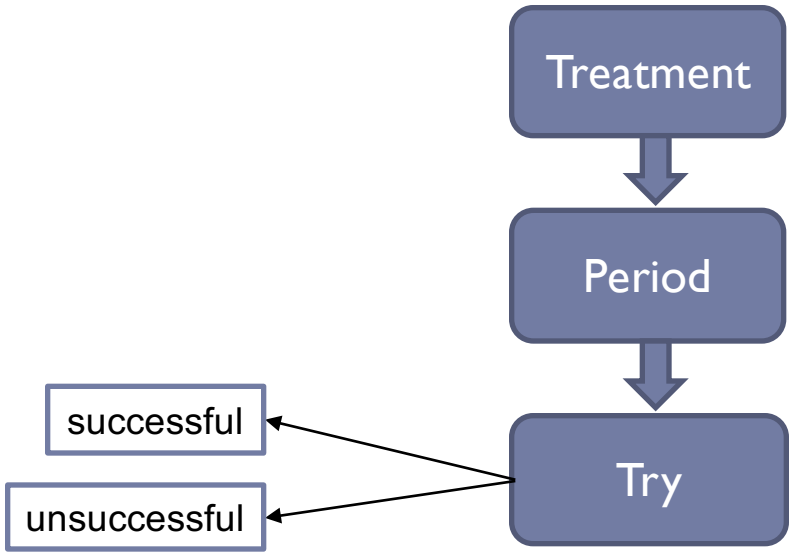
The game interface

Experiment Procedure



Game Details

1. For a given initial height H, the game is said to be in a **period**
2. In each period the participant tries to maximize enjoyment
3. They do so by submitting a set of parameter values (R, H_2, w)
4. Submission of one set of these values in a period is termed as a **try**
    1. If constraints are satisfied, the try is **successful** and the enjoyment value is displayed
    2. If constraints are not satisfied, the try is **unsuccessful** and the enjoyment value is not displayed.
5. A set of 6 periods is termed as a treatment
6. There are 2 treatments with 6 tries and 3 tries for each period respectively



2 treatments

6 periods
Initial Height H randomly generated

3 tries or 6 tries per period

Hypothesis formulation

We **assume** that a participants knowledge lies in their ability to understand constraints.

From the empirical equation generated by Wang and Bao[1]

$$\boldsymbol{\theta} \approx 0.147\boldsymbol{S} - 2.324$$

We term **high scorer** as someone with a positive proficiency θ which equals to $\boldsymbol{S \geq 15}$ and **low scorers** with $\boldsymbol{S \leq 15}$.

We hypothesize that:

Hypothesis 1: *The high scorers will achieve a greater average enjoyment value than the low scorers*

Hypothesis 2*: The high scorers will have a greater percentage of successful tries than the low scorers*

[1] Wang, J., & Bao, L. (2010). Analyzing force concept inventory with item response theory. *American Journal of Physics*, *78*(10), 1064-1070.

Results

We sort out the individuals based on their score into high scorers (HS) and low scorers (LS). We collect the information about their average enjoyment value $\mu_E^{HS\ or\ LS}$ and percentage successful tries $\mu_{\%}^{HS\ or\ LS}$ across all the periods.

We then perform a two sample t-test with unequal variances and get the following results:

Hypothesis 1: The high scorers will achieve a greater average enjoyment value than the low scorers

	Alternate Hypothesis	t stat.	p-value
Two sample t-test	$\mu_E^{HS} > \mu_E^{LS}$	-2.73	0.004

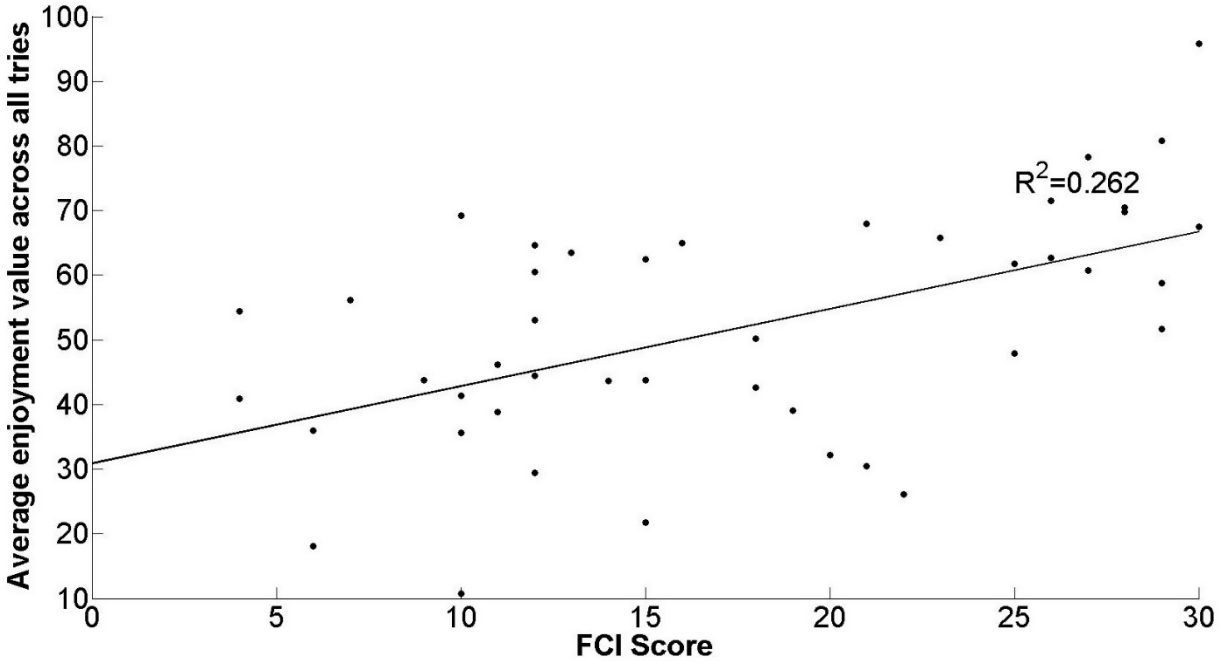
Hypothesis 2: The high scorers will have a greater percentage of successful tries than the low scorers

	Alternate Hypothesis	t stat.	p-value
Two sample t-test	$\mu_{\%}^{HS} > \mu_{\%}^{LS}$	-2.77	0.004

As the p-values are lower than the level of significance ($\alpha = 0.05$) we reject the null for both the hypothesis.

Results

We perform a single predictor linear regression between FCI score S and average enjoyment value achieved.

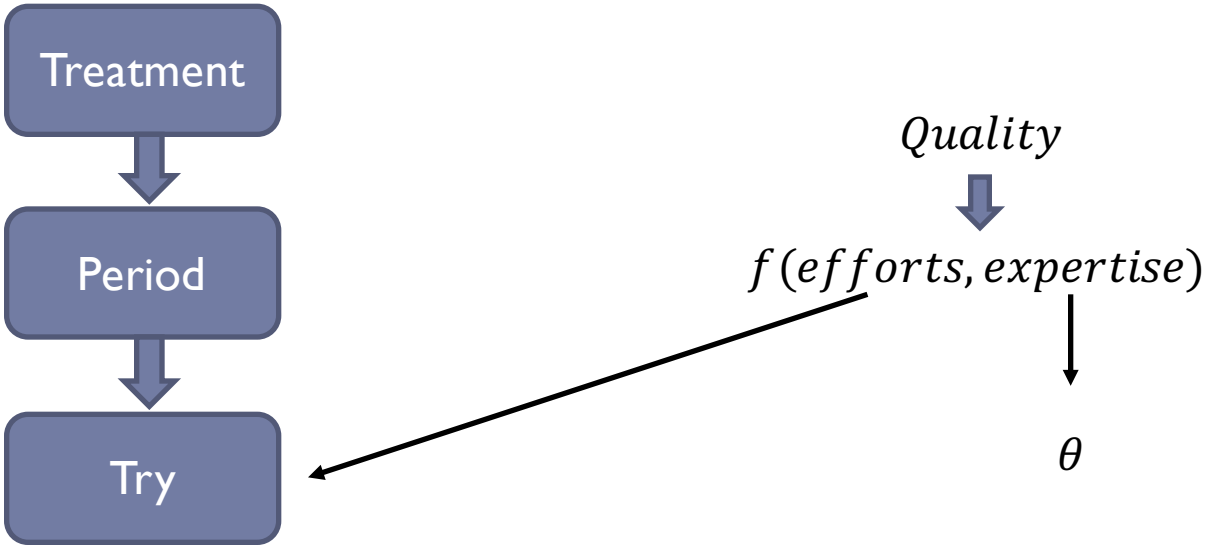


Relationship between Enjoyment (E) and FCI score (S) modeled as

$$E = \beta_0 + \beta_1 S$$

Observations

We had two treatments with 6 tries and 3 tries respectively per period as we wanted to observe behavioral differences if any between the high scorers and low scorers as solution quality is also affected by the number of tries.



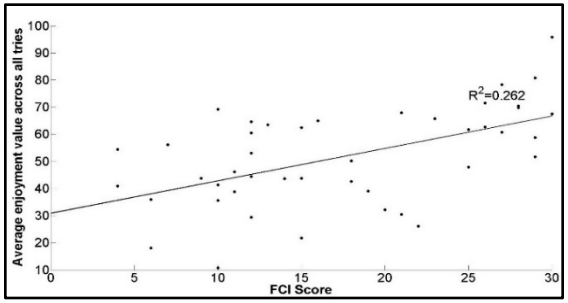
We observed that high scorers maintained their average enjoyment values irrespective of the treatments whereas low scorers had a lower average enjoyment value with lesser tries.

Future Work

What is the impact of expertise in crowdsourcing tournaments? How do we model the strategic decision making process?

Task 1

Quantifying expertise impact on quality



Task 2

Modeling decision making process of the individuals

Task 3

Understanding the impact of strategic decision making

Task 4

Utilizing game theoretic models with behavioral experimentation

Approach

Game Theory

Behavioral Experimentation