A game exists with the following rules:

- You must guess if a weighted coin is H or T.
- You must provide the certainty of your guess.
- You are rewarded based upon a combination of your certainty and if your were correct or incorrect.

Devise a scoring system such that players are incentivized to accurately indicate their certainty in their answer.

Definitions

R_p reward if correct
R_n reward if incorrect
p player's accuracy
c player's reported accuracy
$E = R_p p + R_n(1-p)$ expectation value of player's guess
$M(p)$ optimal strategy for determining c which maximizes expectation

Naive

$$R_p = c, R_n = -c$$

Then

$$E = cp - c(1-p) = c(p-(1-p)) = c(2p-1) = 2pc - c$$

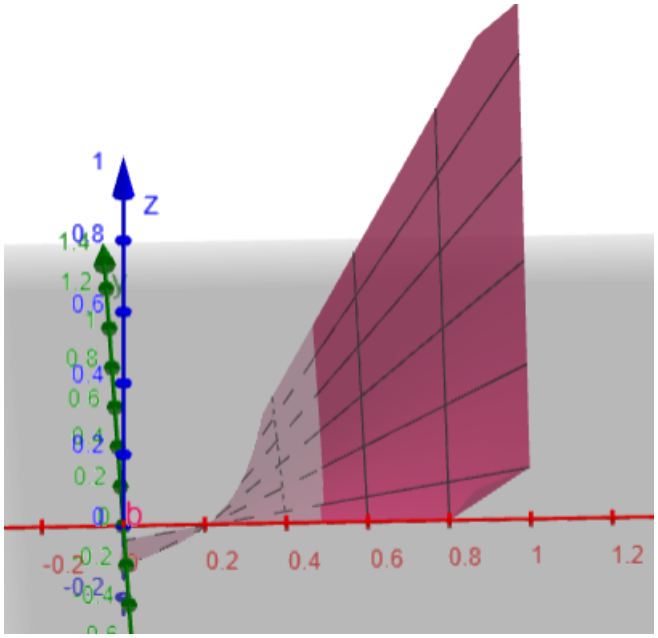

Figure 1: Expectation value for c, p

and so

$$M(p) = \begin{cases} c = 0 \text{ if } p < 0.5 \\ c = 1 \text{ if } p > 0.5 \end{cases}$$

This naive reward system encourages accurately knowing p but lying about it's precise value.

Is there a better reward metric such that the optimal strategy is to not lie i.e. $p = M(p)$?