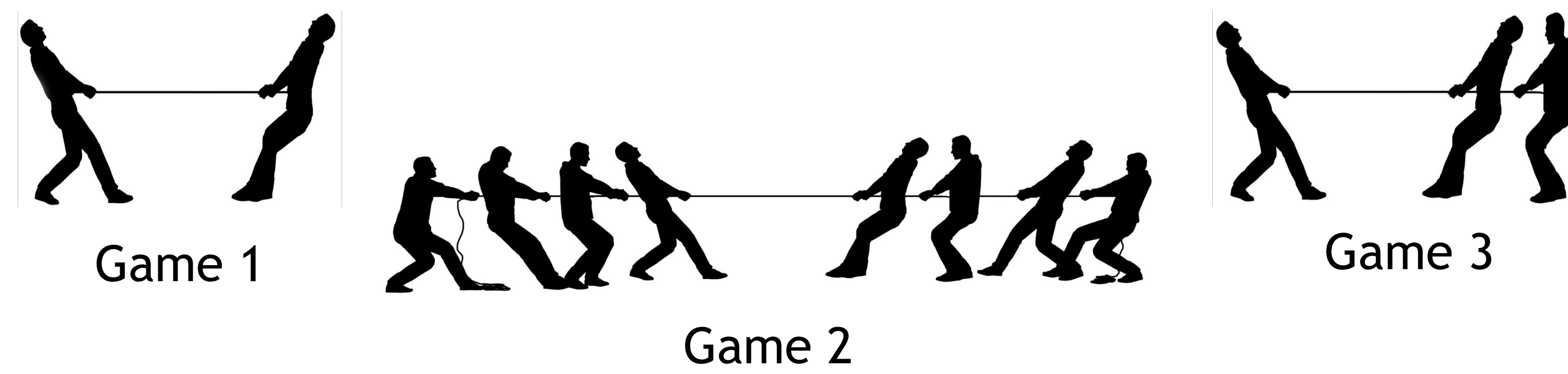


## Imagine a Tug of War Tournament



## The Probabilistic Language of Thought (PLOT)

### Hypothesis 1

*Humans reason about the world using concepts*

strength, winner, team, player, laziness, pulling

### Hypothesis 2

*Concepts are compositional*

A team is made up of players

Team 1 = Alex + Jim + Sarah

### Hypothesis 3

*Concepts are probabilistic*

People aren't always lazy or not lazy

Maybe they are lazy 30% of the time

Concepts in a Probabilistic Language of Thought (Goodman et al. 2015)

## Tug of War PLOT Model

```
var tugOfWarModel = function() {
  var strength = mem(function(person) {
    return sample(DiscretizedGaussian({mu: strengthMean, sigma: strengthSD})))

  var effort = mem(function(person, match) {
    return categorical({ps: [lazyEffortPrior, normalEffortPrior, extraEffortPrior],
      vs: [lazyPulling, normalPulling, extraPulling]}))

  var pulling = function(person, match) {
    return effort(person, match) * strength(person)}

  var totalPulling = function(team, match) {
    return sum(map(function(person) {return pulling(person, match)}, team))}

  var deltaTeamStrength = function(team1, team2, match) {
    return totalPulling(team1, match) - totalPulling(team2, match)}

  var winner = function(team1, team2, match) {
    return deltaTeamStrength(team1, team2, match) > 0 ? 1 : 2}

  var isClose = function(team1, team2, match) {
    return Math.abs(deltaTeamStrength(team1, team2, match)) < ccloseThreshold}

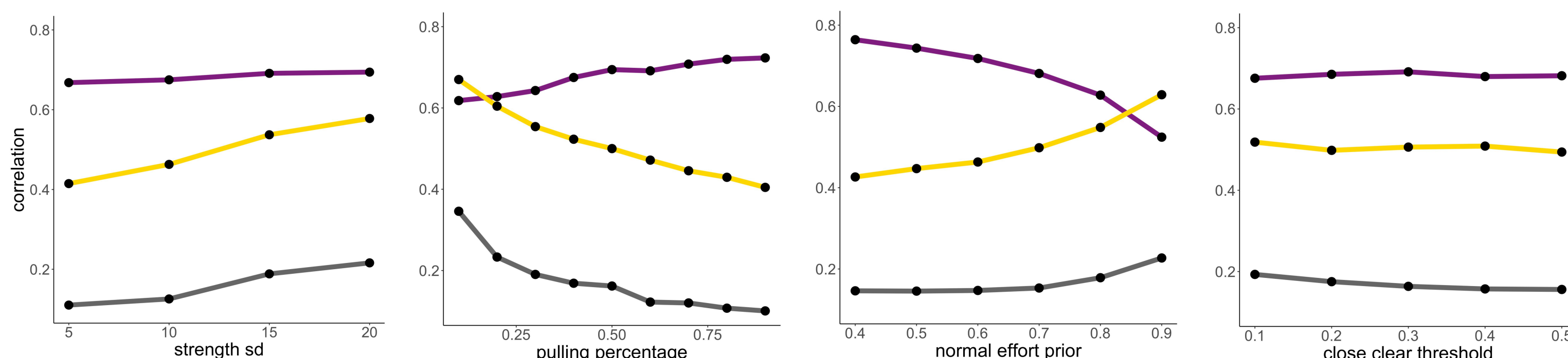
  var isClear = function(team1, team2, match) {
    return Math.abs(deltaTeamStrength(team1, team2, match)) > cclearThreshold}

  evidence {
    condition(winner(['Bob', 'Mary'], ['Tom'], ['match1']))
    condition(winner(['Bob', 'Sue'], ['Tom'], ['match1']))

    query {
      return strength('Bob') }
  }
}
```

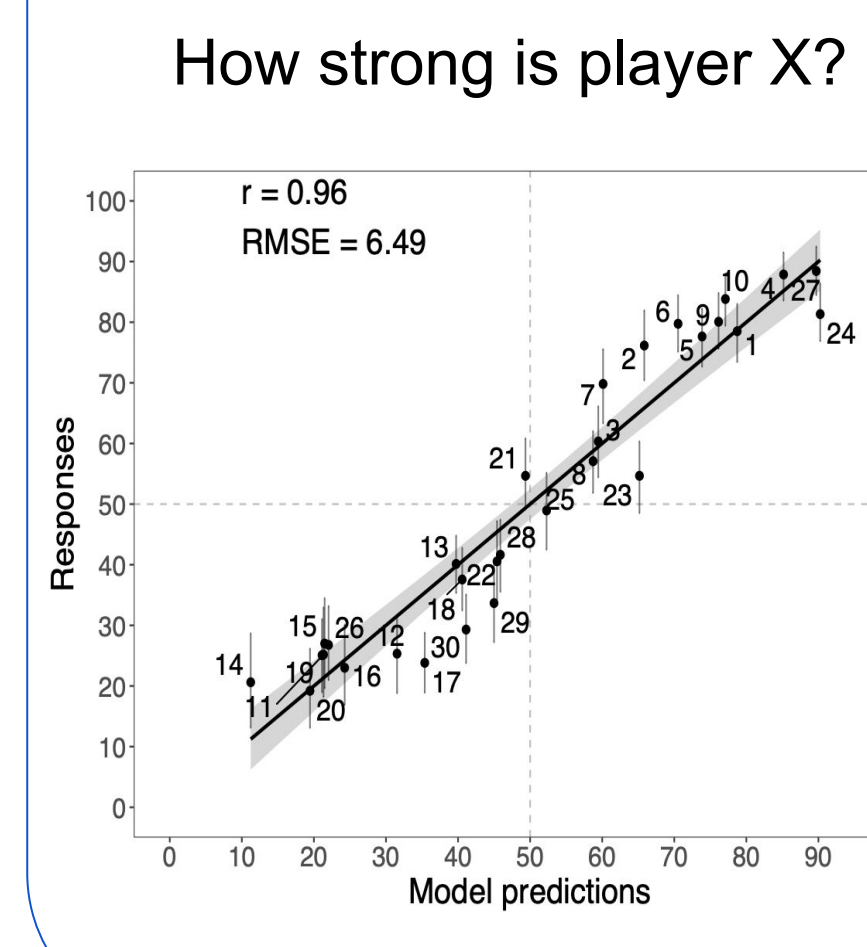
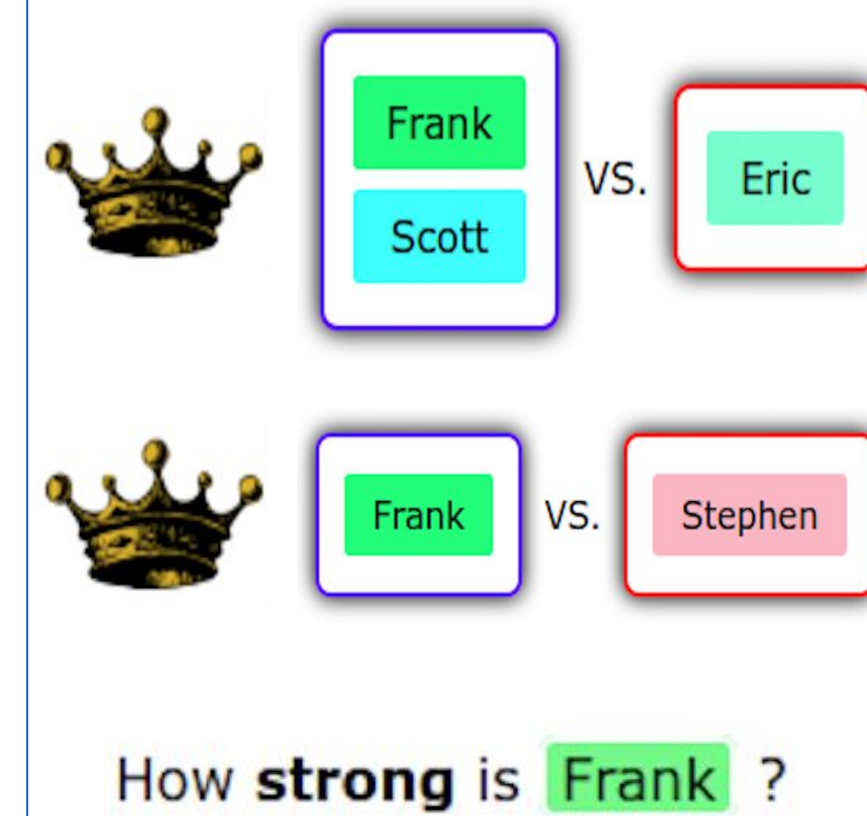
## Marginal Distributions for Experiment 3 by Question

■ effort ■ outcome ■ strength

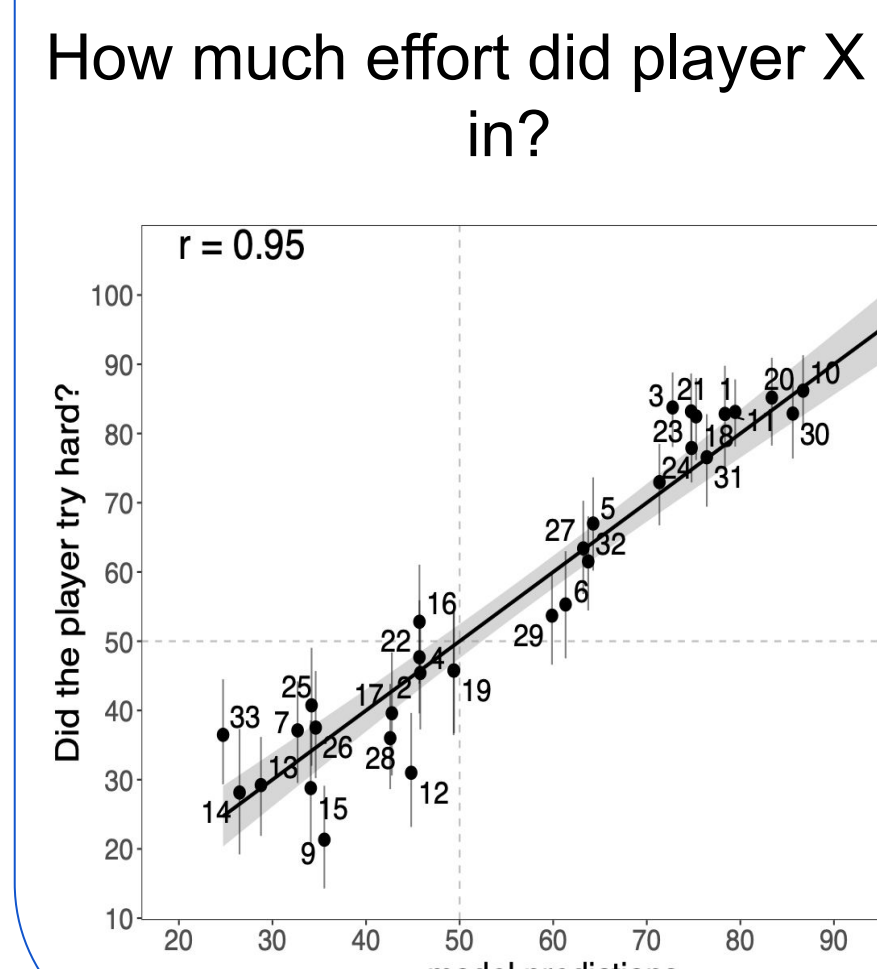
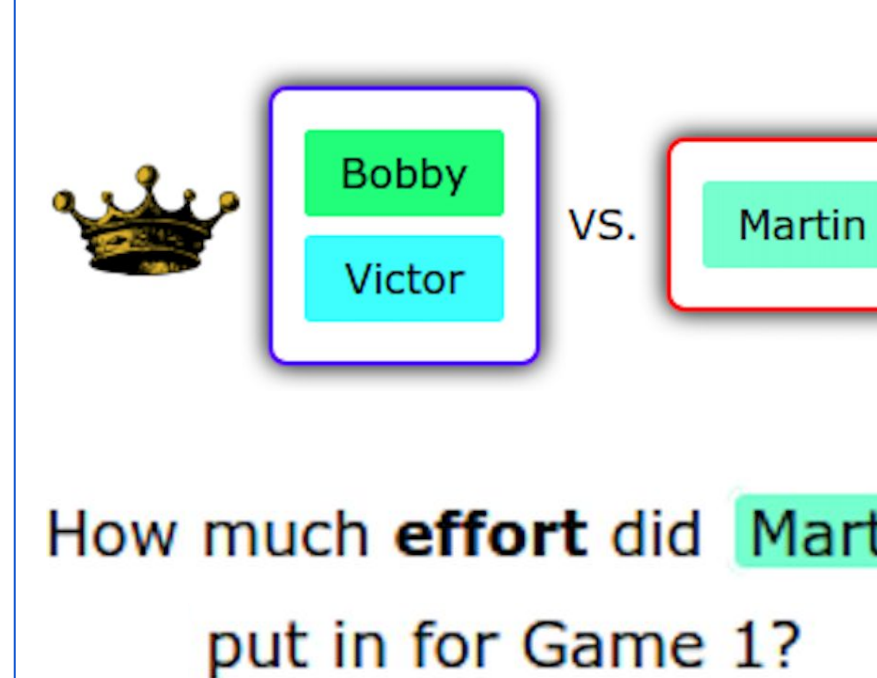


## Tug of war PLOT model matches human intuitions strikingly well

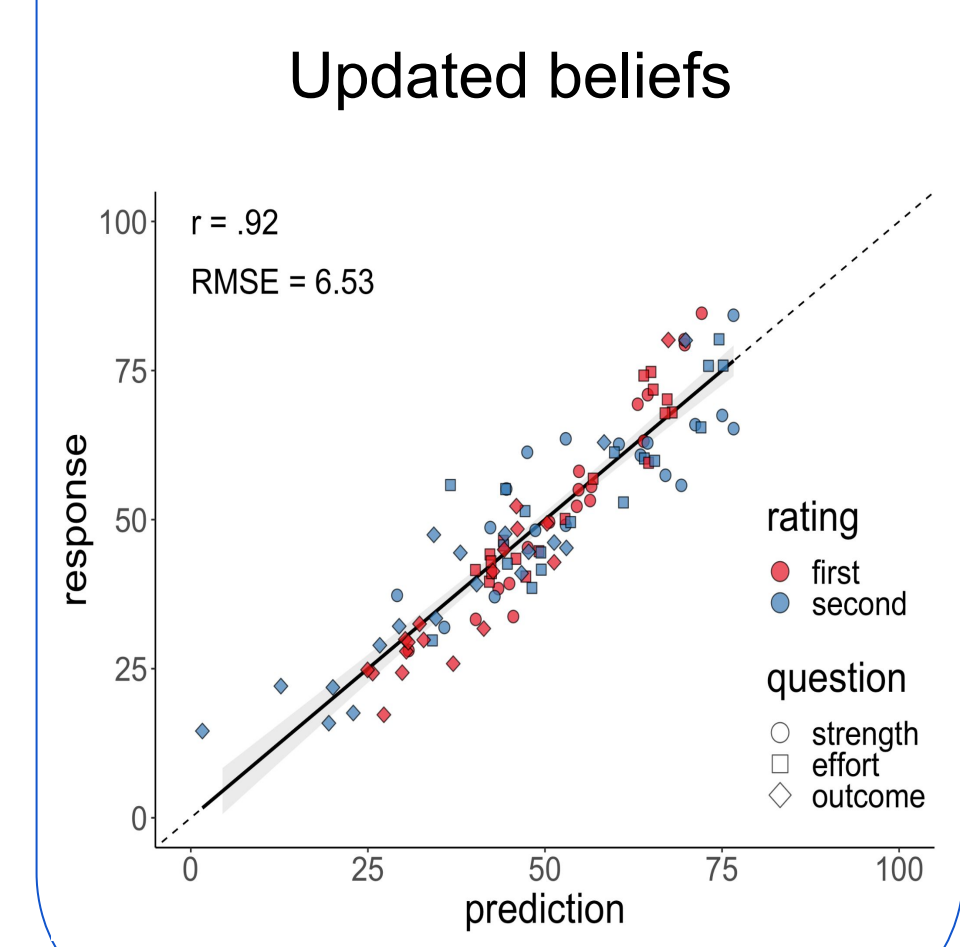
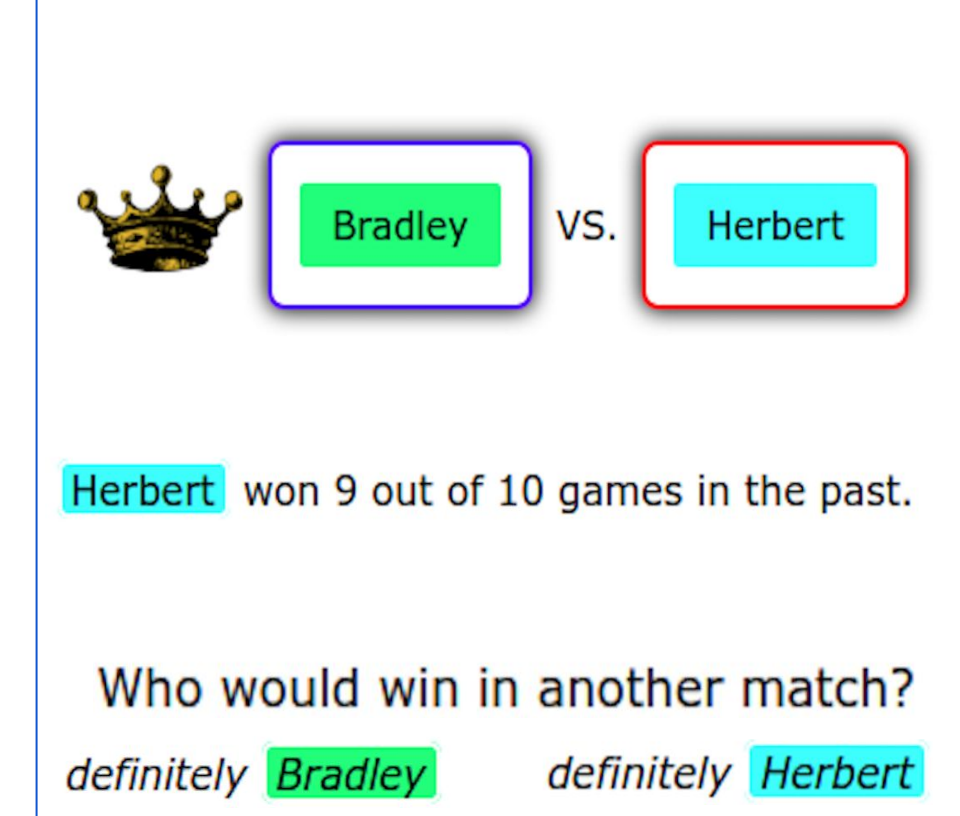
### Experiment 1 Strength



### Experiment 2 Laziness



### Experiment 3 Multiple Questions & Updated Beliefs



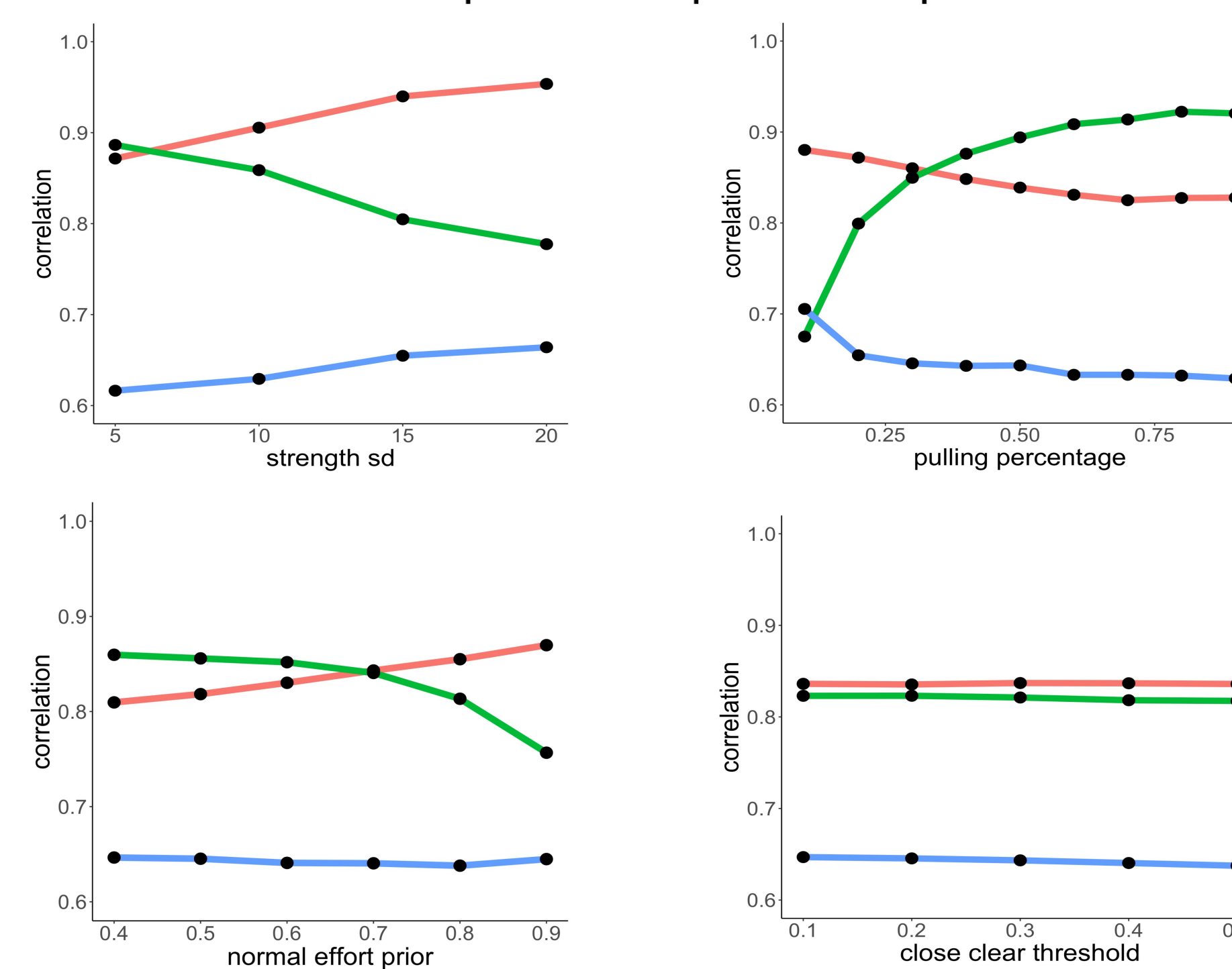
## How sensitive is the PLOT model to parameter choices?

Parameter grid search over four parameters, all other parameters were dependent on one of these four

Parameters	Range
strength standard deviation	[ 5, 10, 15, 20 ]
pulling percentage	[ .1, .2, .3, .4, .5, .6, .7, .8, .9 ]
normal effort prior	[ .4, .5, .6, .7, .8, .9 ]
close clear threshold	[ .1, .2, .3, .4, .5 ]

## Marginal Distributions by Experiment

■ exp 1 ■ exp 2 ■ exp 3



## Alternative Models

### Large Language Models (GPT-3)

Is it possible for a model to have the same kind of flexibility and generalizability of human thought just through large amounts of data?

#### Exp1 trial where GPT-3 closely matches human intuitions

...How strong is player 1 on a scale of 0 to 100 based on the evidence below?

In Game 1, player 1 competed against player 2 and player 1 won.

In Game 2, player 2 competed against player 3 and player 2 won.

In Game 3, player 2 competed against player 4 and player 2 won.

Player 1 is stronger than player 2, who is stronger than players 3 and 4. Therefore, player 1 is at least as strong as player 2, and is likely stronger. I would rate player 1 as a 90.

#### Exp3 trial where GPT-3 does not capture human intuitions

...How strong is player 1 on a scale of 0 to 100 based on the evidence below?

In Game 1, player 1 competed against player 2 and player 1 won

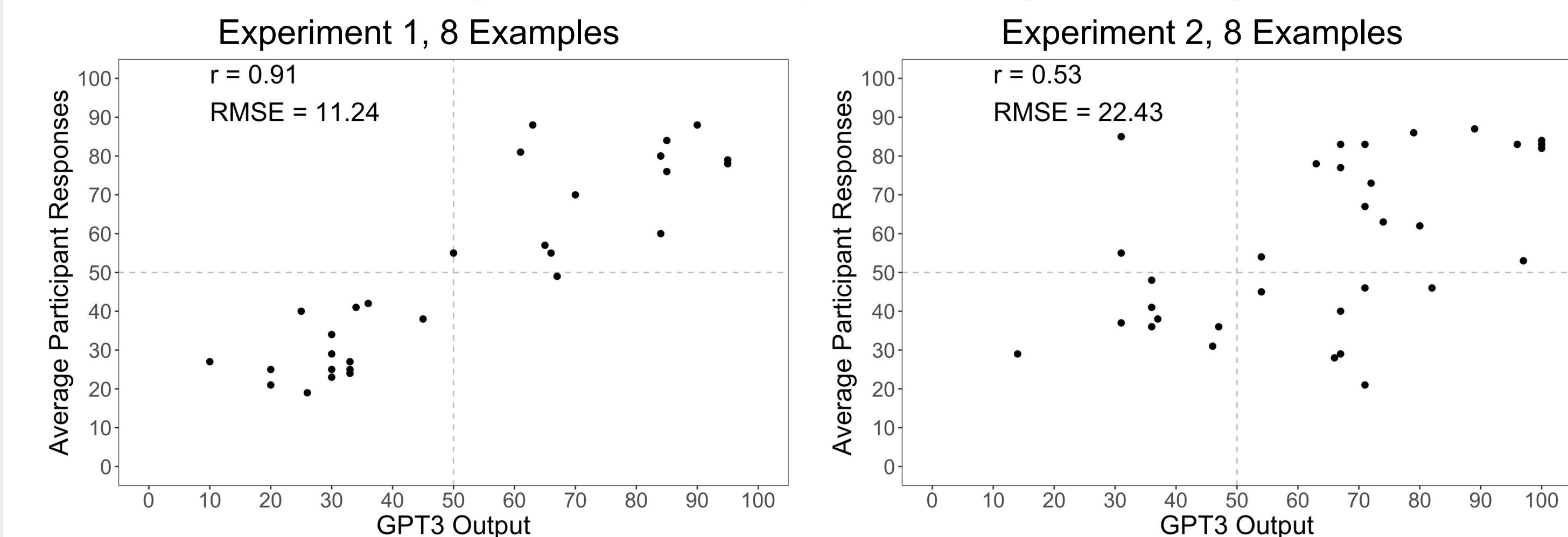
In Game 2, player 1 competed against player 3 and player 3 won.

Player 1 is 50% strong.

Player 1 won 2 out of 10 games in the past. Given this new evidence, on a scale of 0 to 100, how strong is player 1?

Player 1 is 80% strong.

## GPT-3 Output vs. Average Participant Responses



## Feature Based Models

Can features of the tug of war game accurately predict human judgments?

- # games lost
- # games won
- # players on team
- # players on opposite team
- Did player win game being asked about

## Discussions & Future Work

### Parameter Grid Search

- Parameters matter to some degree
- There isn't only one parameter setting that works
- But not the case that all of the parameter settings work

### PLOT & Alternative Models

- PLOT model works well but need to complete evaluation of alternative models