Defining Lying as a function of P and Q

How I Spent My Tuesday Night

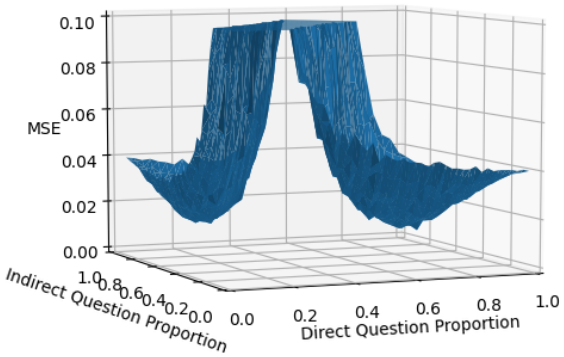
Maxwell Lovig

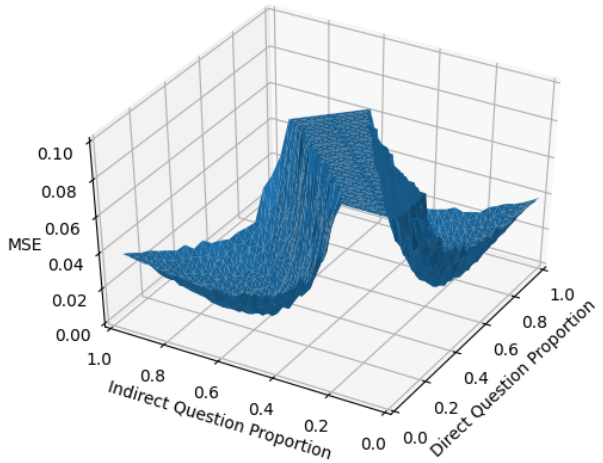
What If Lying Is A Function of the Direct Question

Could it be the case that people do not lie based if the question is sensitive but how well scrambled there response is.

For example would it be the case that the same amount of people would lie when p = .9 versus p = .5 for Greenberg or Warners model

Without Incorporating Lies

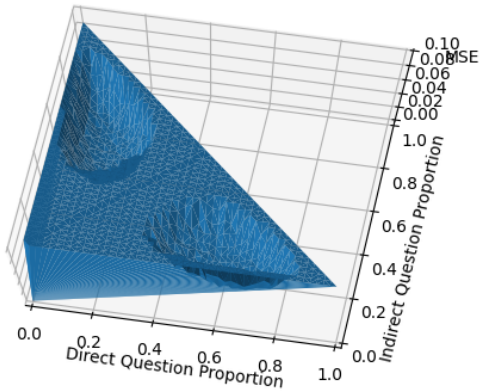


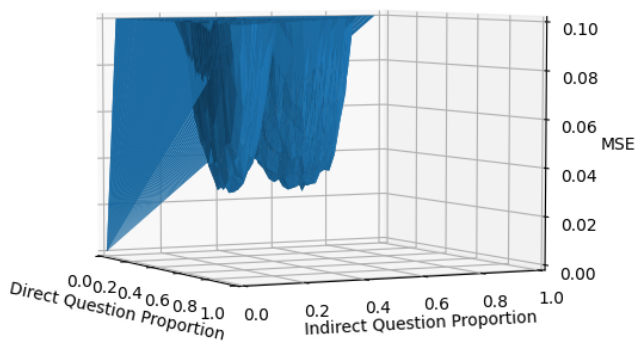


$$A = \frac{pq(1-p-q)}{(1/3)^3}$$

$$n = 100, \pi_x = .2, \pi_y = .7, p = 0.16, q = 0.62, MSE = 0.0156$$

With Incorporating Lies



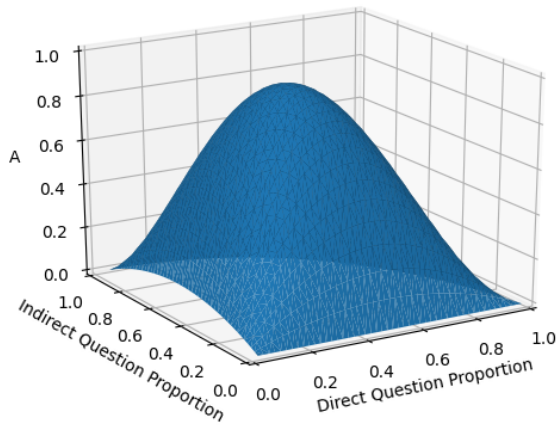


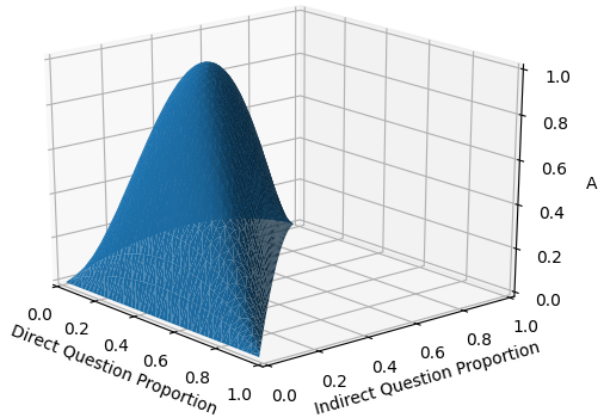
$$A = \frac{pq(1-p-q)}{(1/3)^3}$$

$$n = 100, \pi_x = .2, \pi_y = .7, p = 0.12, q = 0.64, MSE = .024$$

Graph of Lying Distribution

$$A = \frac{pq(1-p-q)}{(1/3)^3}$$





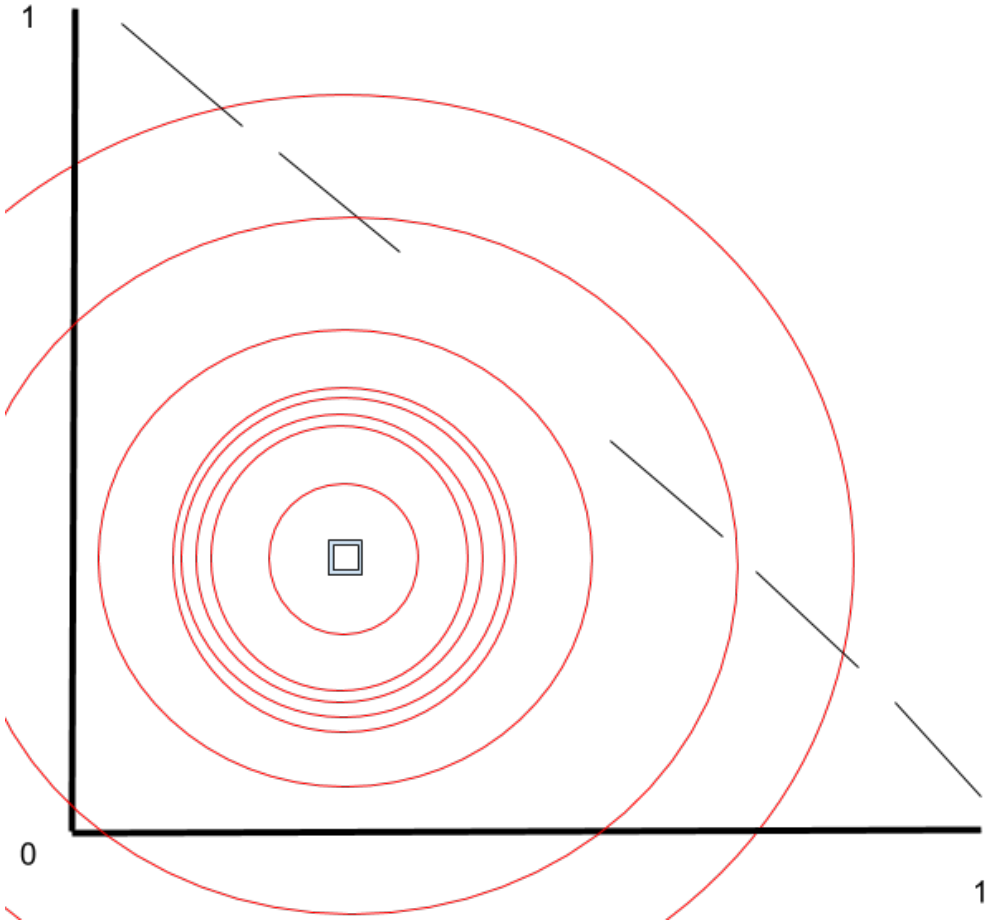
In order to avoid the assumption of how the distribution of lying based on the direct and indirect probabilities p and q we can brute force a peisewise distribution.

For Example

Task for the Interviewee: Place on the chart a circle in which inside the circle are values of p and q where you would be truthful for the model and the are outside is where you would be untruthful.

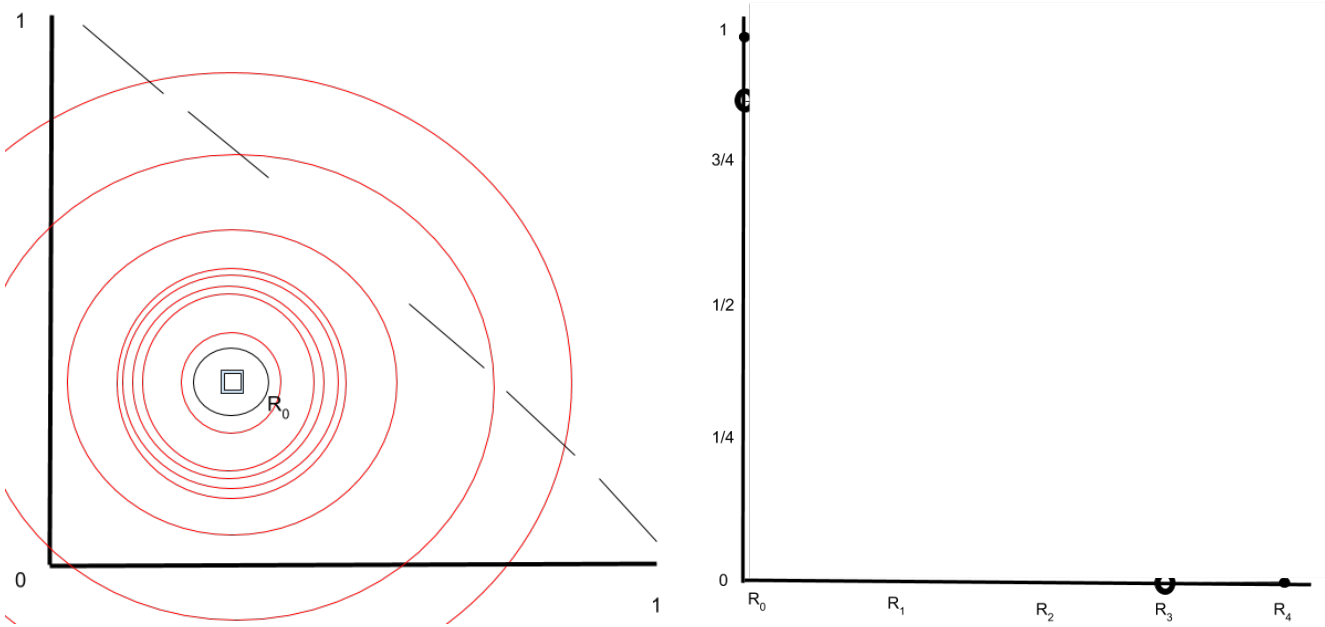
The x axis is the indirect question proportion, the y axis is the direct question proportion.

For Example



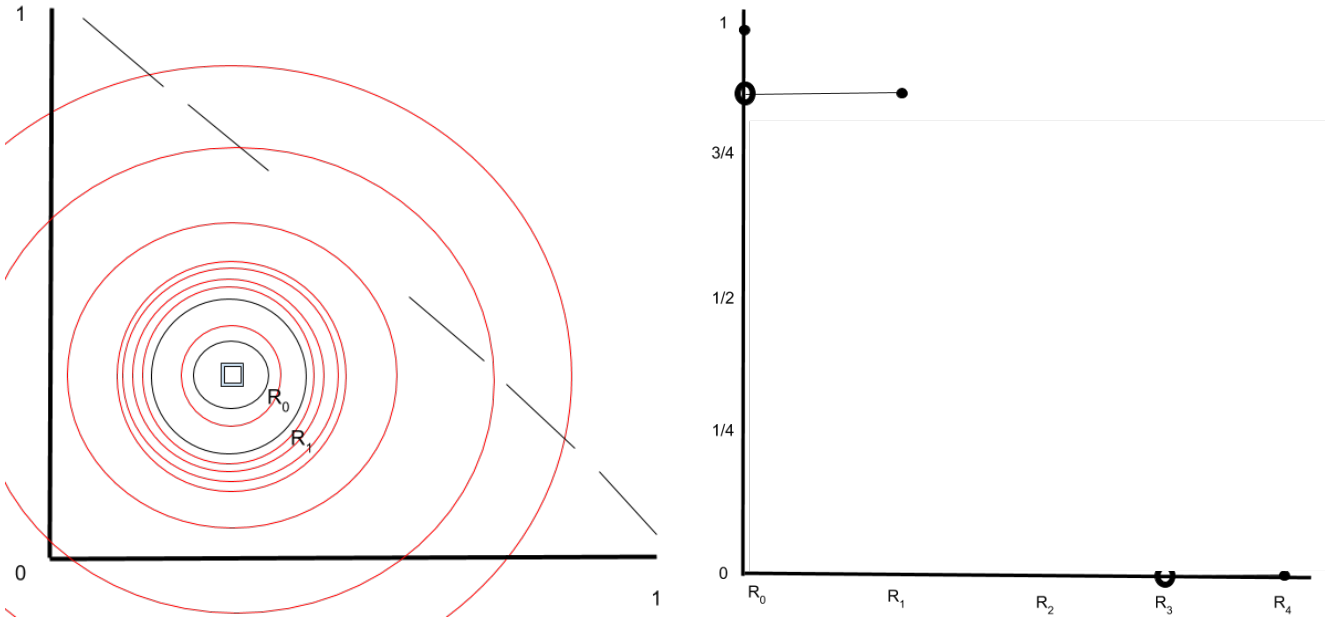
$$A_R = P(D_R > R)$$

For Example



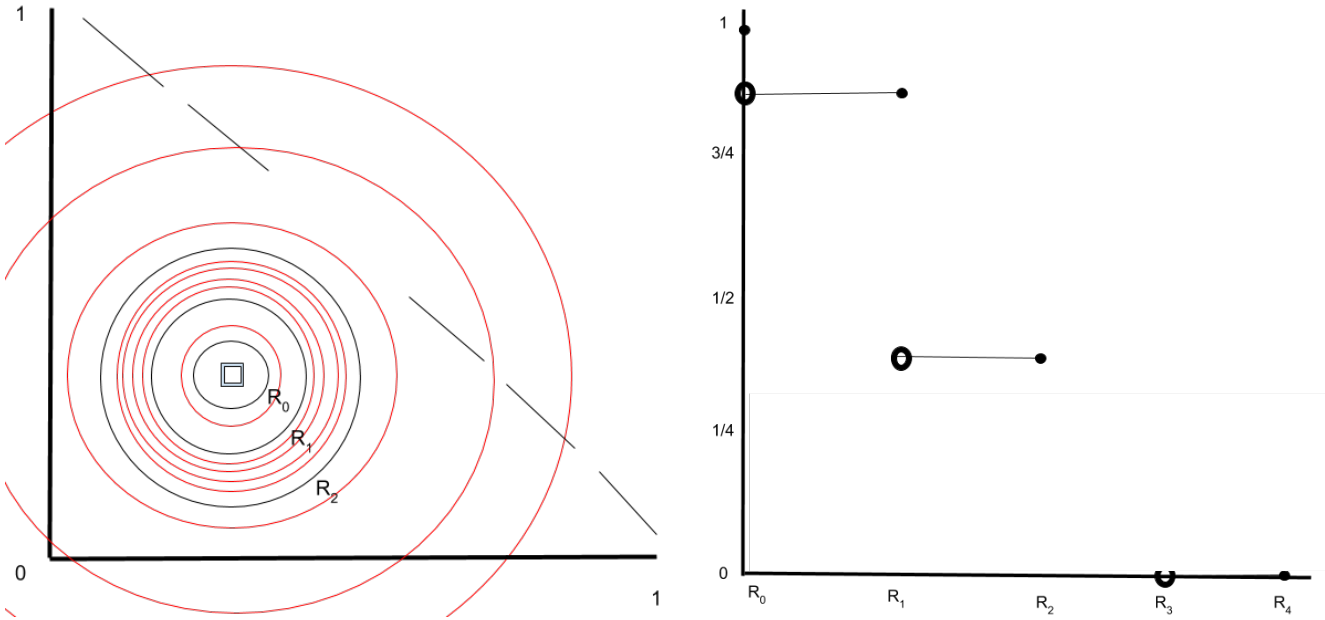
$$A_{R0} = 8/8 = 1$$

For Example



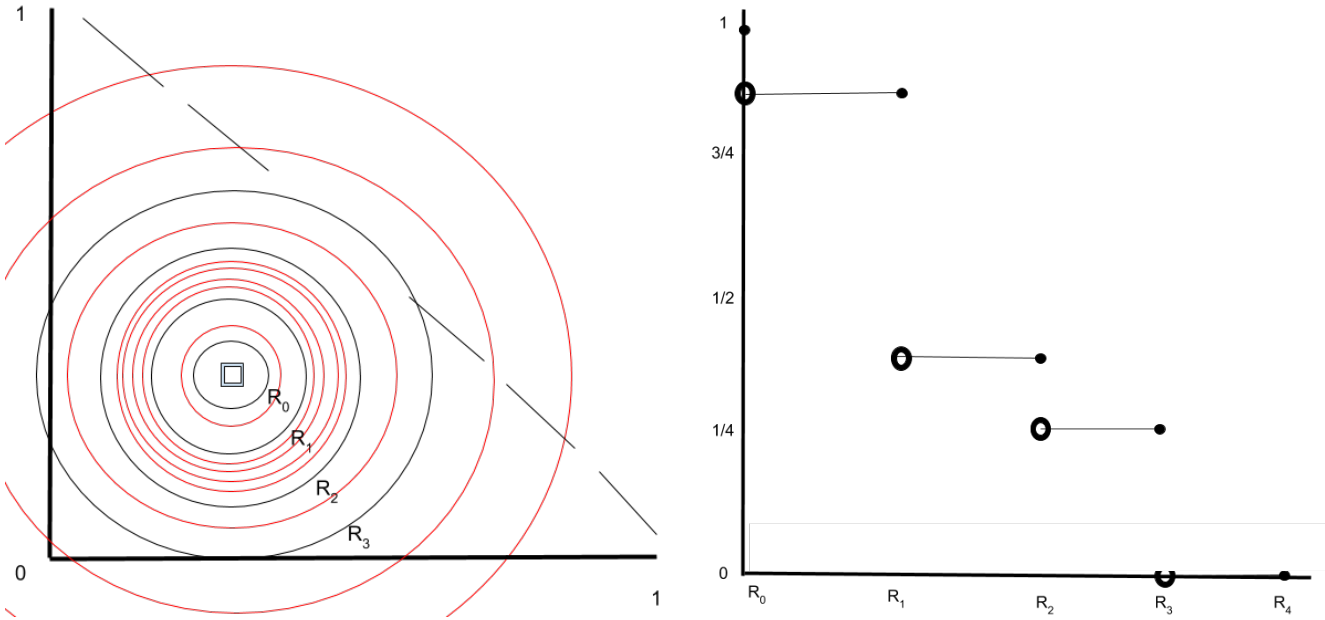
$$A_{R1} = 7/8 = ..875$$

For Example



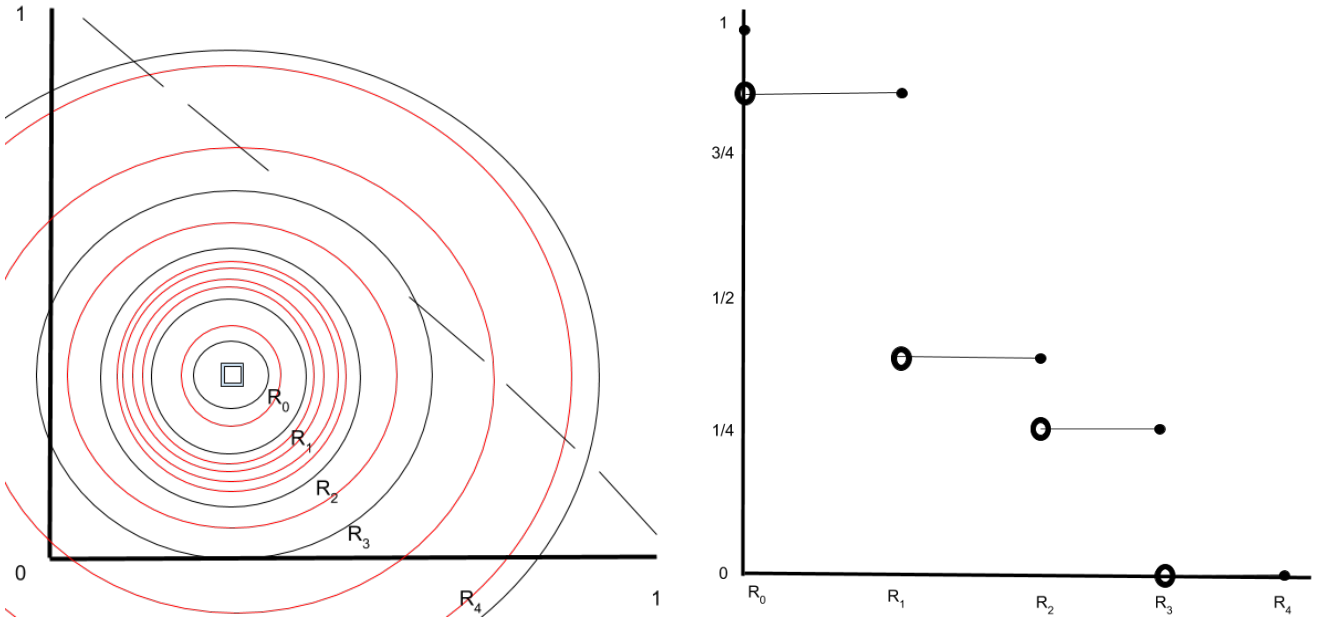
$$A_{R2} = 3/8 = .375$$

For Example



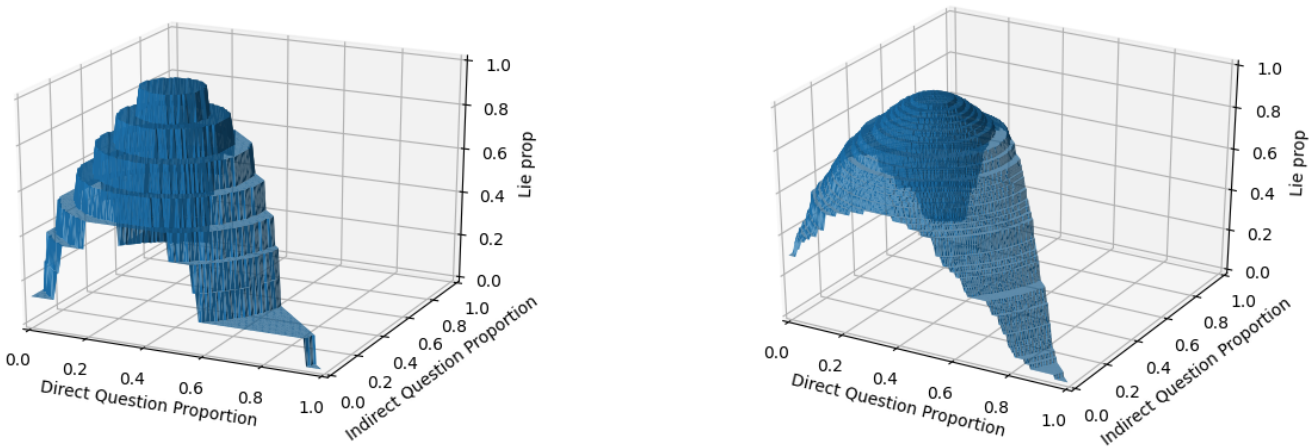
$$A_{R3} = 2/8 = .25$$

For Example



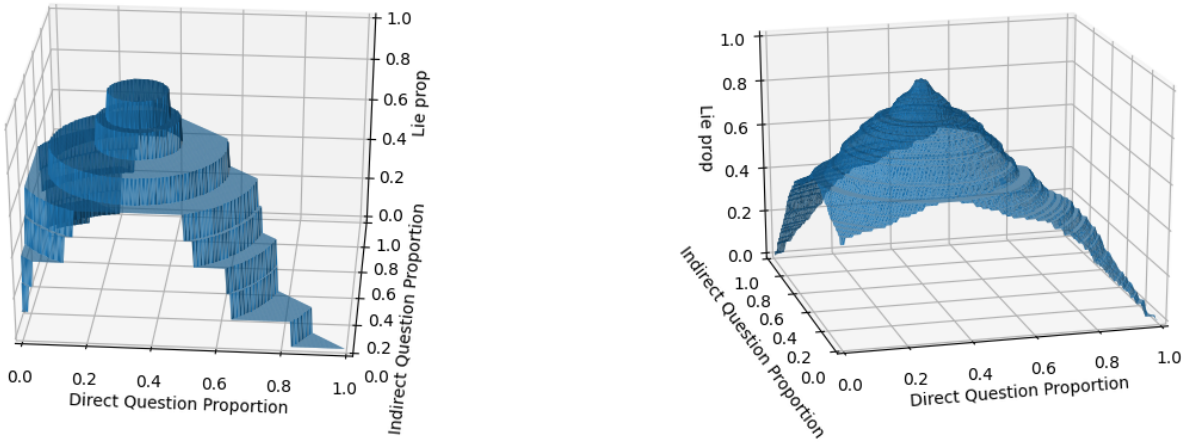
$$A_{R4} = 0/8 = 0$$

3D Case: Hill



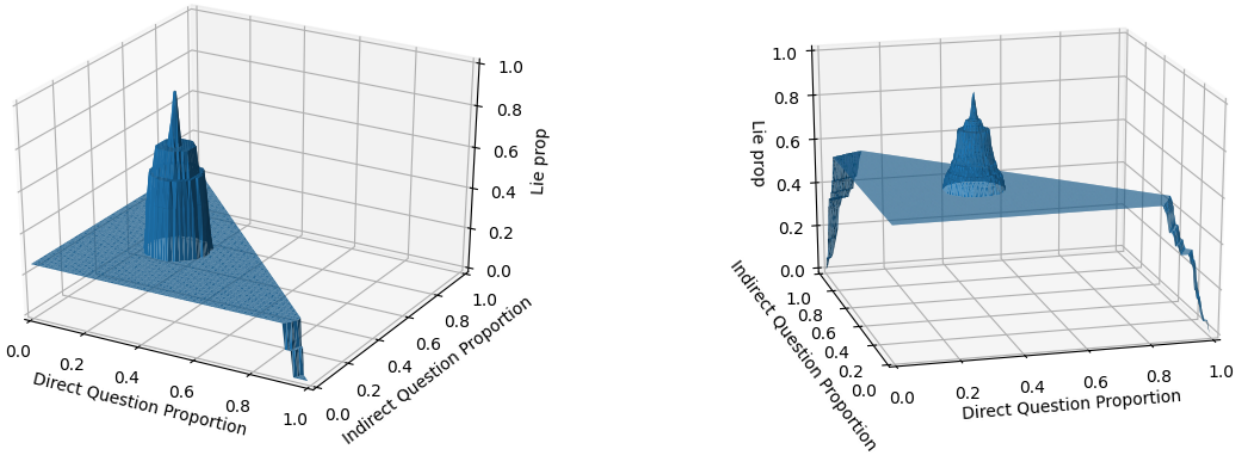
This is the hill shape, based off a random distribution of points or a normal distribution of radii to generate circles, with n = 8 and n = 50 respectively.

3D Case: Peak



This is the peak shape, based off a random distribution of radii to generate circles, with n = 8 and n = 100 respectively.

3D Case: Plateau



This is the plateau shape, based off a bi-model distribution of radii to generate circles, with n = 8 and n = 50 respectively. Specifically seen here is an example for 2 groups with vastly differing sensitivity

Incorporation of Quantitative RRT Model

Because this information is vital to our estimation of A it could as be categorized as sensitive information as A would be. However, the answers we get are not binary and are quantitative.

This means we are going to have to use a quantitative RRT to scramble the data but still attempt to get the mean and variance of our metric, we could assume normality and approximate a distribution.

Incorporation of Quantitative RRT Model

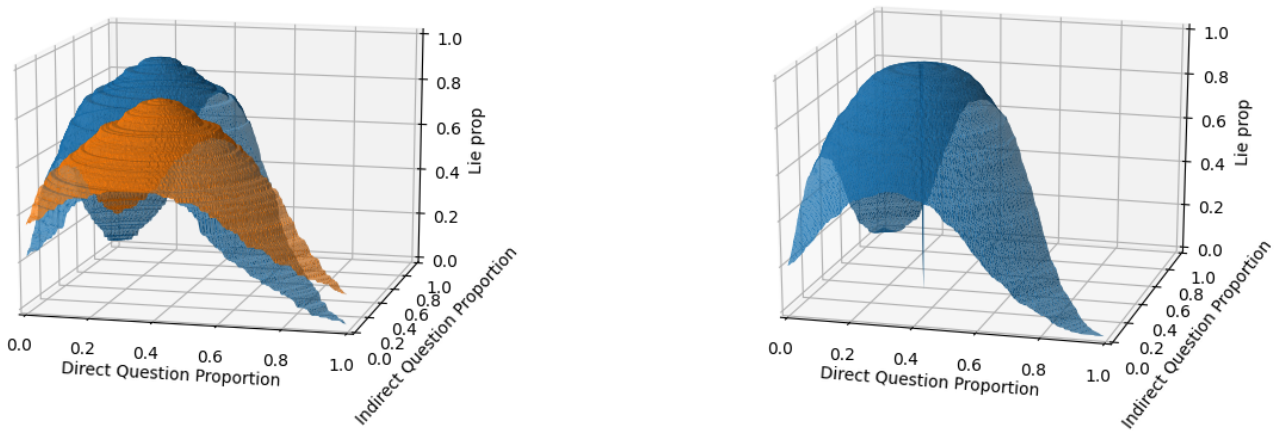
A basic additive model would be wonderful for this since, When $Z = Y + S$

$$E(Z) = E(Y) + E(S) \text{ and } Var(Z) = Var(Y) + Var(S) \implies$$

$$E(Z) - E(S) = E(Y) \text{ and } Var(Z) - Var(S) = Var(Y)$$

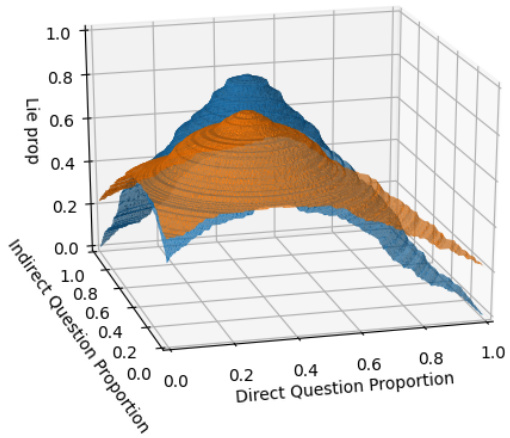
Dr. Sadia's quantitative model with untruthfulness would also work as it is unbiased and is even more robust.

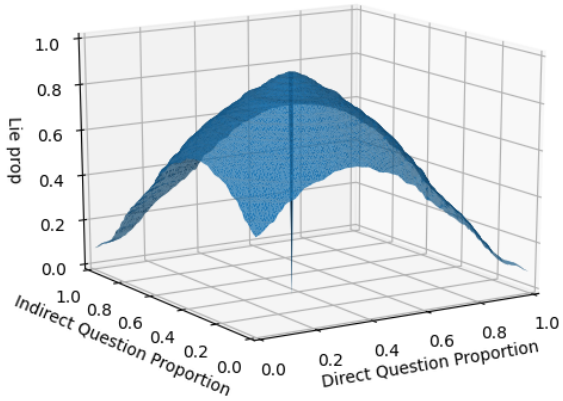
Approximation: Hill



On the left in blue is the original piesewise distribution (n=100) and orange is a scrambling distribution which is Gaussian (mean = 0, sd = .1). On the Right is the approximated distribution using a Gaussian distribution of radii (mean = $E^*(Y)$, sd = $\sqrt{Var^*(Y)}$)

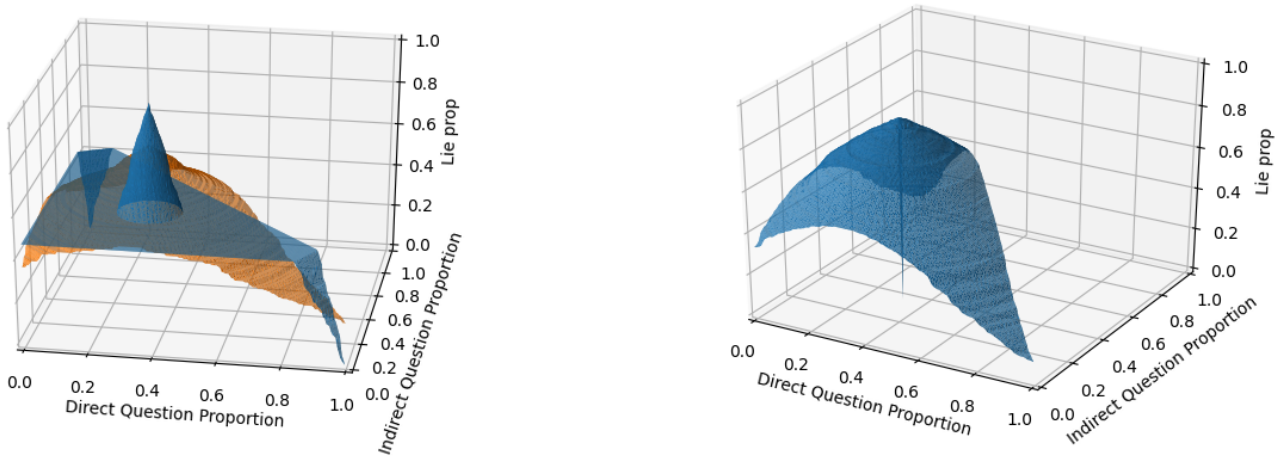
Approximation: Peak



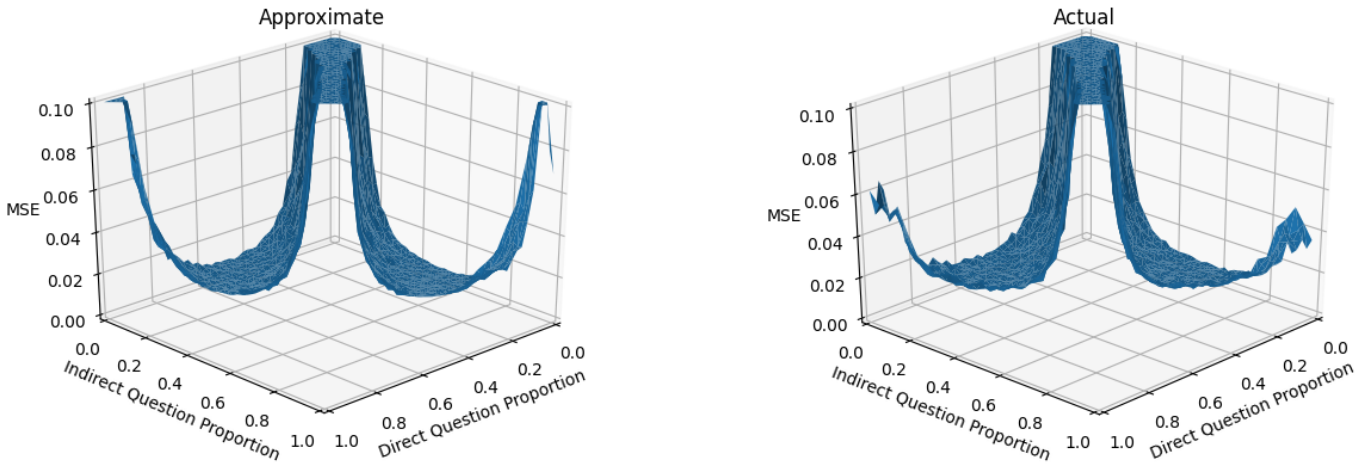


On the left in blue is the original piesewise (n=200) distribution and orange is a scrambling distribution which is Gaussian(mean $= 0$, sd $= .1$). On the Right is the approximated distribution using a Gaussian distribution of radii(mean $= E^*(Y)$, sd $= \sqrt{Var^*(Y)}$)

Approximation: Plateau



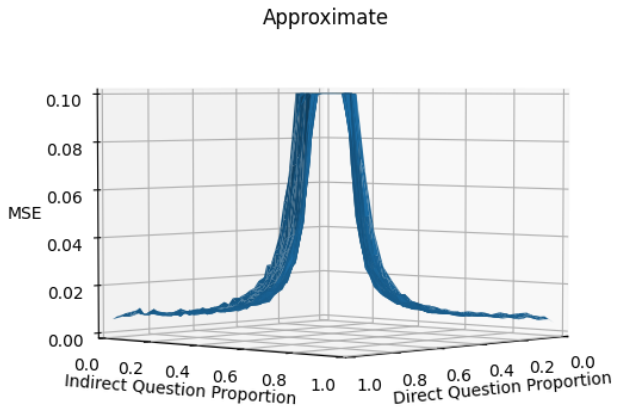
On the left in blue is the original piesewise (n=200) distribution and orange is a scrambling distribution which is Gaussian(mean = 0, sd = .1). On the Right is the approximated distribution using a Gaussian distribution of radii(mean = $E^*(Y)$, sd = $\sqrt{Var^*(Y)}$)

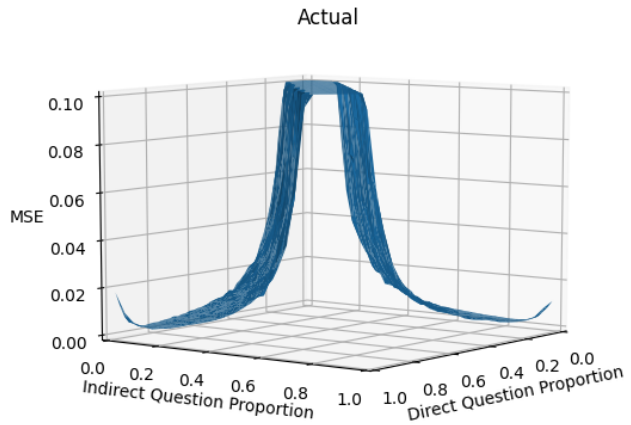


$$n = 500, \pi_x = .2, \pi_y = .7, A = \textit{Pure Hill}\ (\frac{1}{3}, \frac{1}{3})$$

$$\text{Gaussian Scrambler}, \mu = 0, \sigma^2 = .1$$

Approx Distribution: $p = 0.1,\ q = 0.52,\ \textit{MSE} = 0.0040956$
Actual Distribution: $p = 0.12,\ q = 0.54,\ \textit{MSE} = 0.0054223$





$$n = 500, \pi_x = .2, \pi_y = .7, A = \textit{Double Plateau}\ (\frac{1}{3}, \frac{1}{3})$$

$$\text{Gaussian Scrambler}, \mu = 0, \sigma^2 = .1$$

Approx Distribution: $p = 0,\ q = 0.98,\ \textit{MSE} = 0.00362$
Actual Distribution: $p = 0.02,\ q = 0.86,\ \textit{MSE} = 0.00132$

Where Can The Center Be?

Perfect Scrambler: Models where responses are scrambled so well the true responses can never be recovered. Most attractive to interviewee. Occurs at $p = q$ or $p = .5, q = 0$

Imperfect Scrambler: Not a perfect scrambler

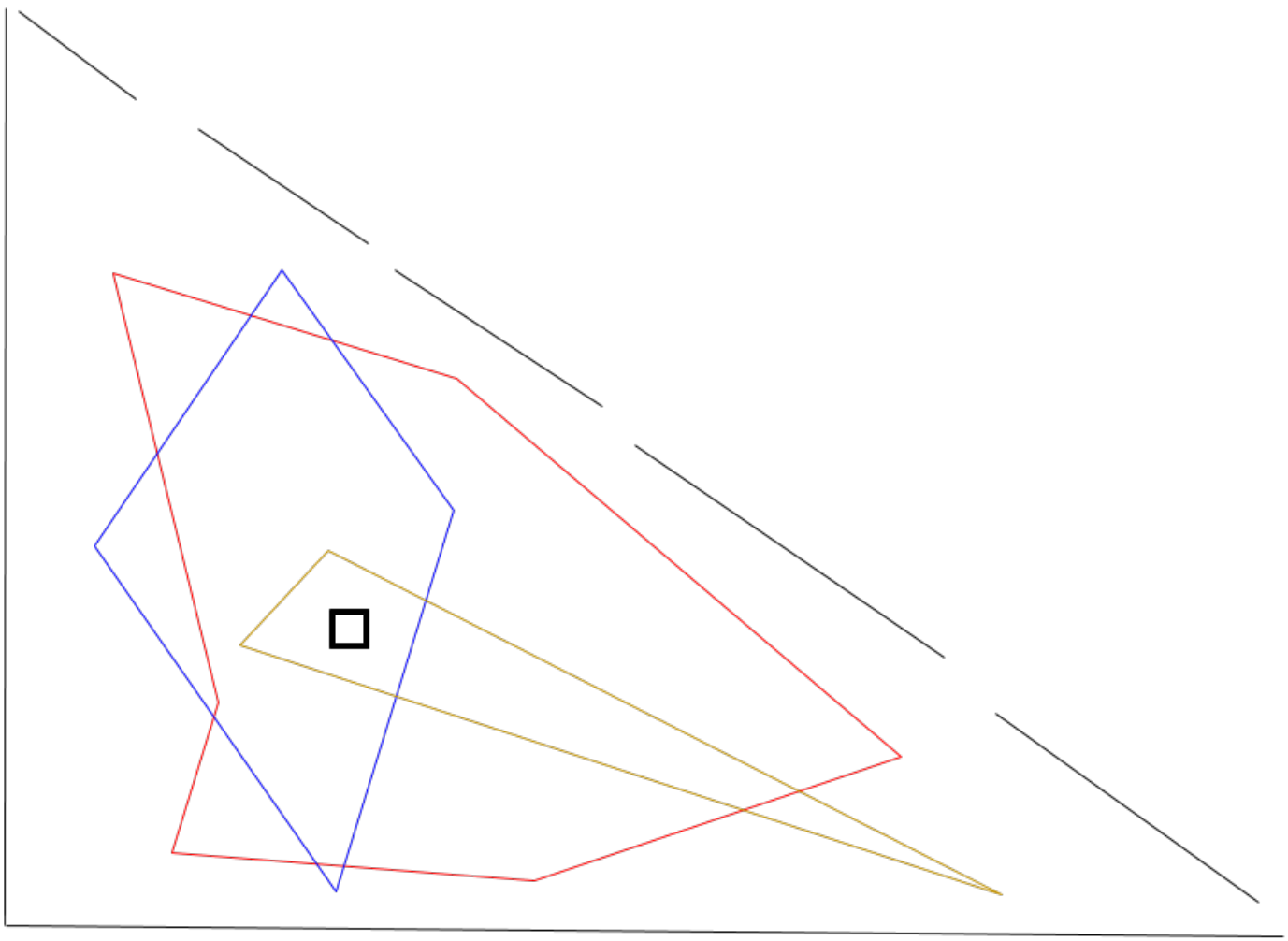
Perfect Scramblers are most reasonable centers for the circle method due to the assumption everyone should tell the truth when their responses are perfectly scrambled, $p = q = \frac{1}{3}$ was used here.

More Robust Method

Task for the Interviewee: Draw a shape which represents which areas of p and q which you would be truthful.

The x axis is the indirect question proportion, the y axis is the direct question proportion.

For Example



For Example