# Stepwise Data Fitting

This is an attempt to incrementally analyze the model as we complicate it slowly by adding new dynamics (except for identity which has been left out here).

## Models and Simulations

Every simulation run has 1000 agents that represent humans, and the richest model has 10 media agents as well as a variable media influence factor. For each of these steps, every parameter combination is run with 30 random initializations, except for the richest model (Step 4) which has been run with 5 random initializations.

The following models were tried:

1. Simple Hegselmann-Krause with full network. Uniformly distributed initial opinions.
2. Simple Hegselmann-Krause with scale-free network. Uniformly distributed initial opinions.
3. Hegselmann-Krause with normally distributed open-mindedness and scale-free network. Uniformly distributed initial opinions.
4. Hegselmann-Krause with normally distributed open-mindedness, normally distributed initial opinion, and scale-free network.
5. Hegselmann-Krause with normally distributed open-mindedness, normally distributed initial opinion, media agents (), variable media distribution parameters, variable media influence factor, and scale-free network.

**Notes:** 1. The media influence factor and is a weighting factor for media-sourced influence. So implies no influence from media sources, while implies that media-sourced influence is as strong as other people’s opinions. 2. The scale-free network parameter is set to 1. 3. For Step 4, initial opinion distribution parameters were fixed to values of and after pilot simulations showed significantly good fits for these values. 4. Media opinion positions are drawn from a deterministic set of positions approximating the normal pdf, and have the parameters and .

| Model | Parameters | Random Initializations per param combination |
| --- | --- | --- |
| Step 1 |  | 30 |
| Step 2 |  | 30 |
| Step 3 | , | 30 |
| Step 4 | , , | 5 |
| Step 5 | , , , , | 1 (more being run) |

## Survey Data

<>

### Parlemeter Data - Initial fitting patterns

<>

### Dynamic Data -

<>

## Fitting Strategy

We have used Jensen-Shannon Divergence (JSD) as the metric here.

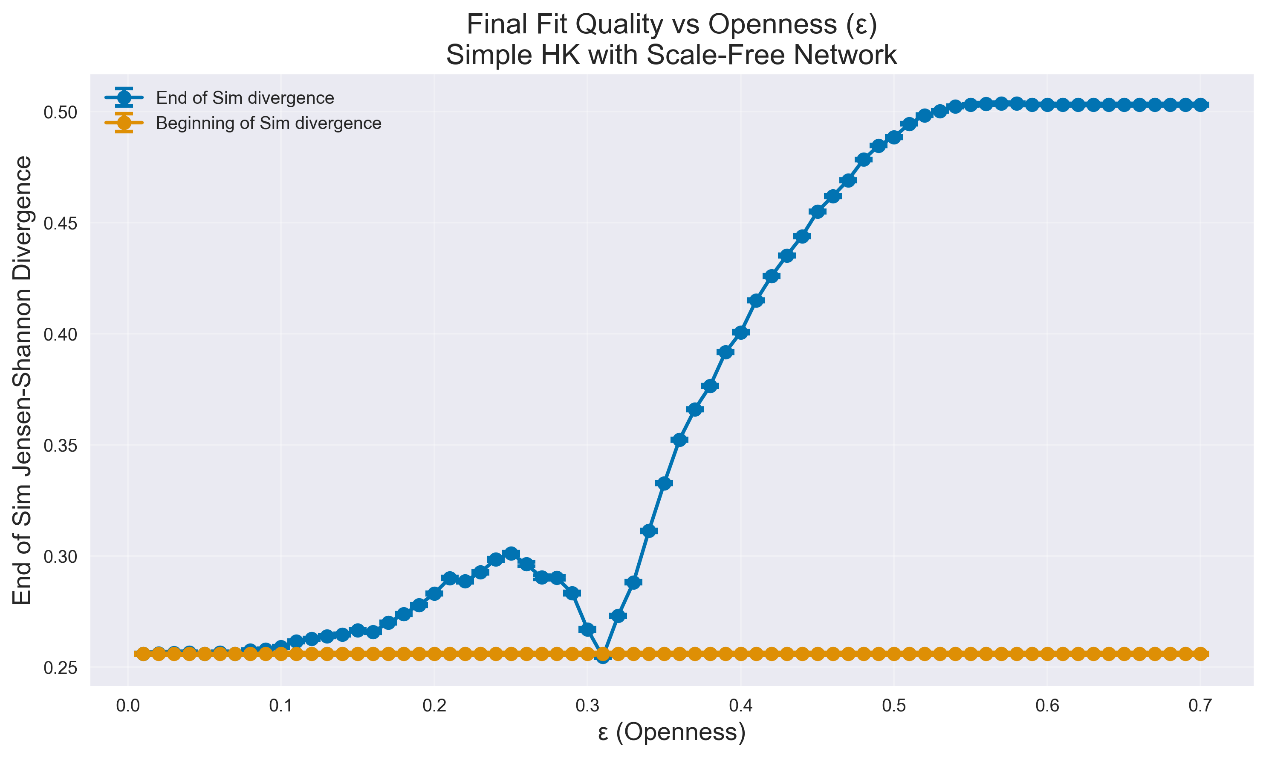
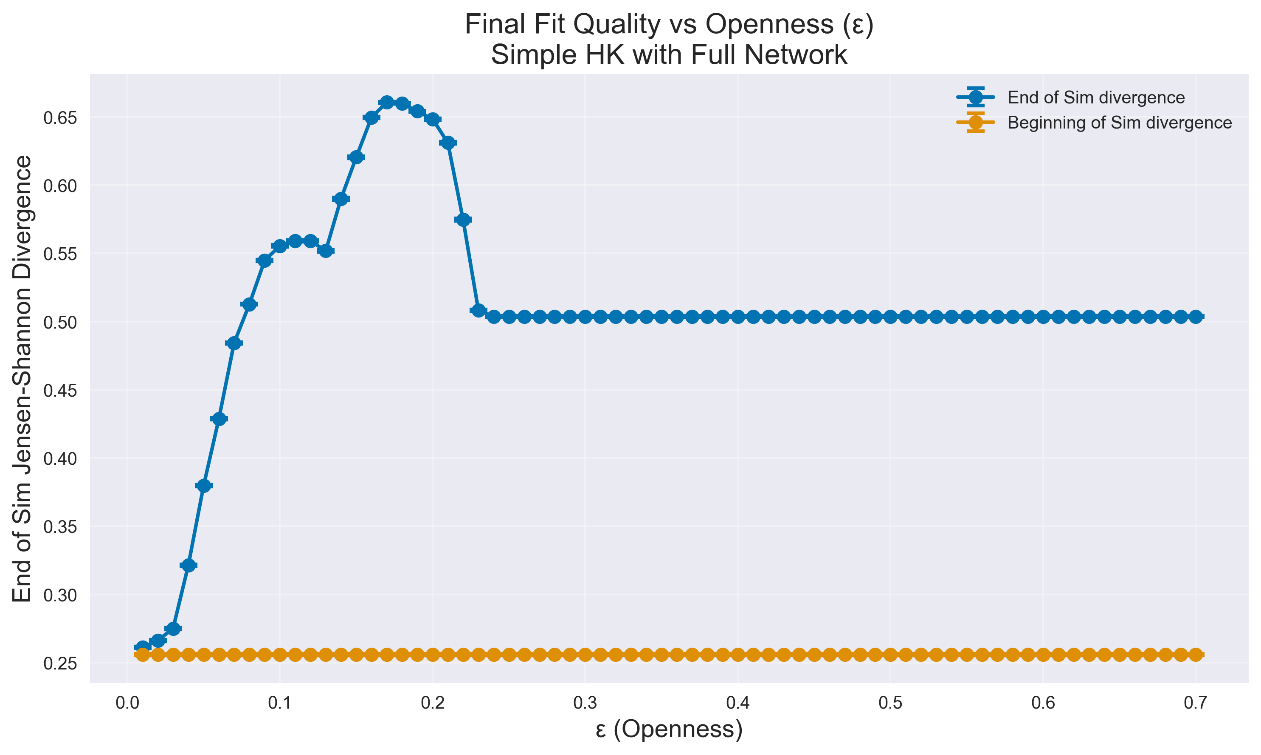
Data from every simulation were binned to match the 10-point scale, and the middle two bins were collapsed (the same was done for the survey data) to create a true middle bin.

Every simulation was compared to every survey point (one for every country and year from the Parlemeter) to derive an initial and final JSD with respect to the survey data (considering opinions from the simulation before v/s after the simulation).

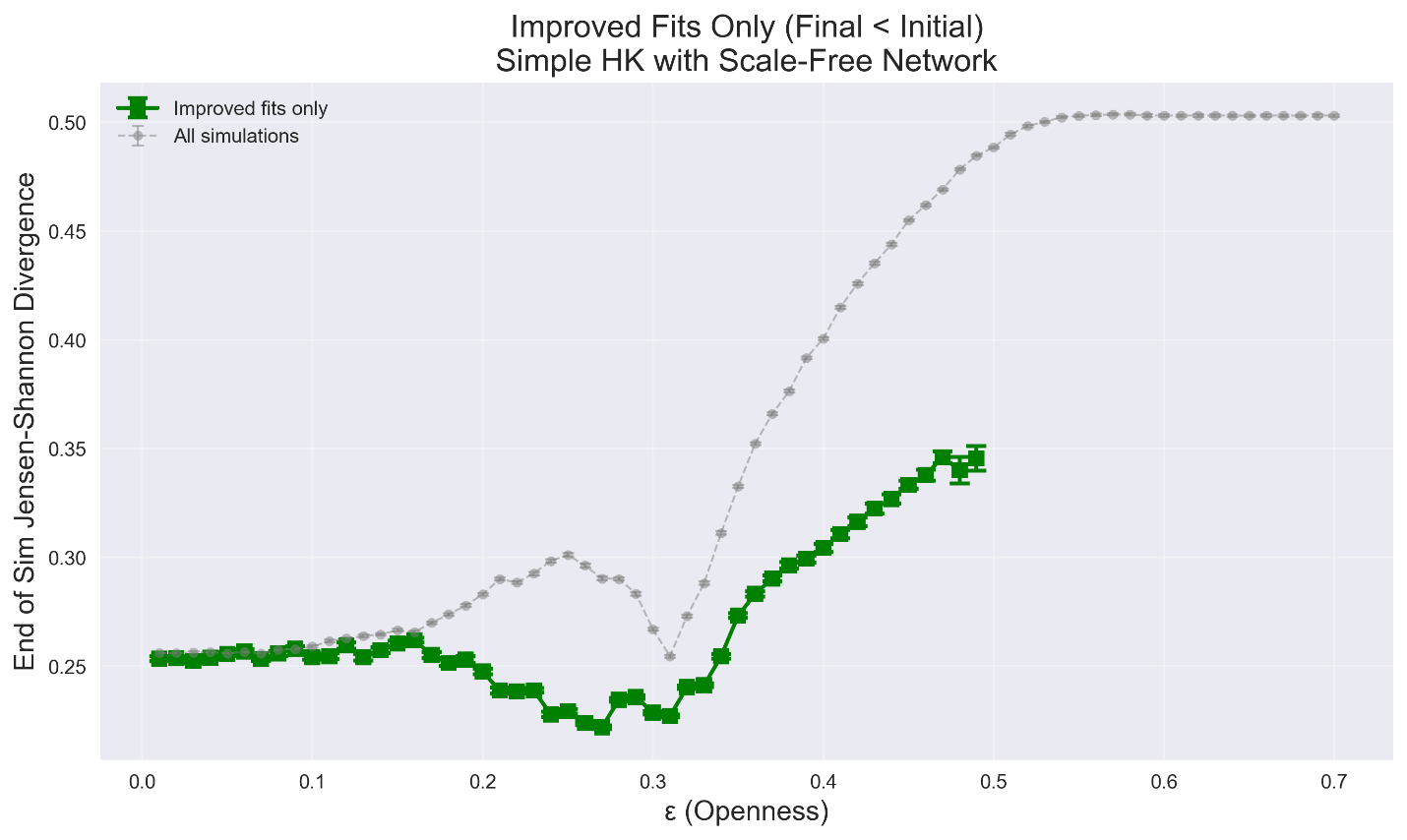
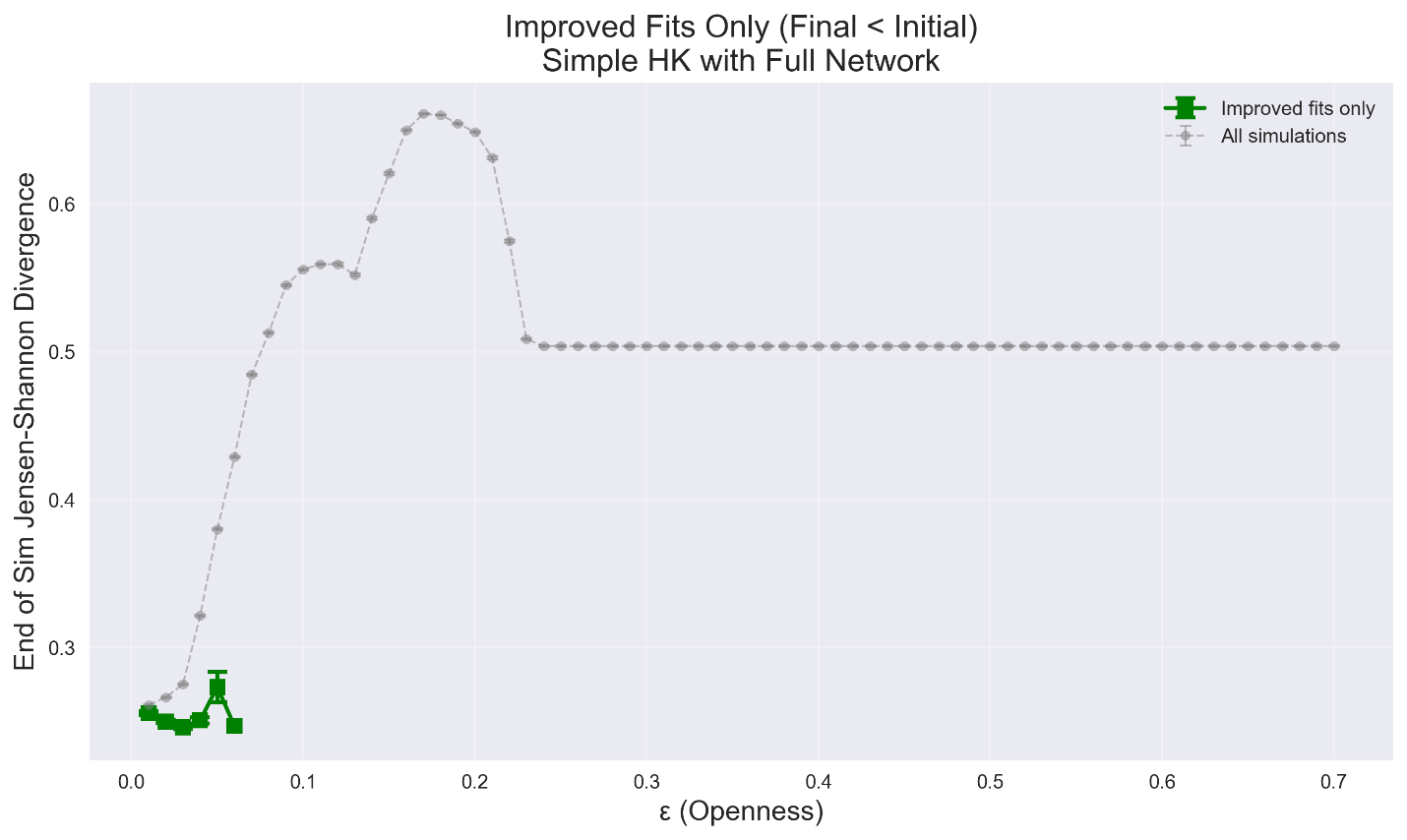
## Data

For the line plots below, unless specified otherwise data from different runs (and different comparisons to survey data) for a given parameter or parameter vector were averaged. For the colorplots (unless specified otherwise) the best fit (among all initializations and all comparisons to country\*year data) is plotted instead of the average.

### Steps 1 and 2

These two steps differ only in the type of network used. Neither shows very good fits – there was no simulation in either step for which the final JSD was below 0.05. 

*Figure 1: Mean JSD plotted against Openness for Step 1 (top) and 2 (bottom). Each data point is an average of several comparisons as each simulation is compared to all combinations of year and country in the Parlemeter.*



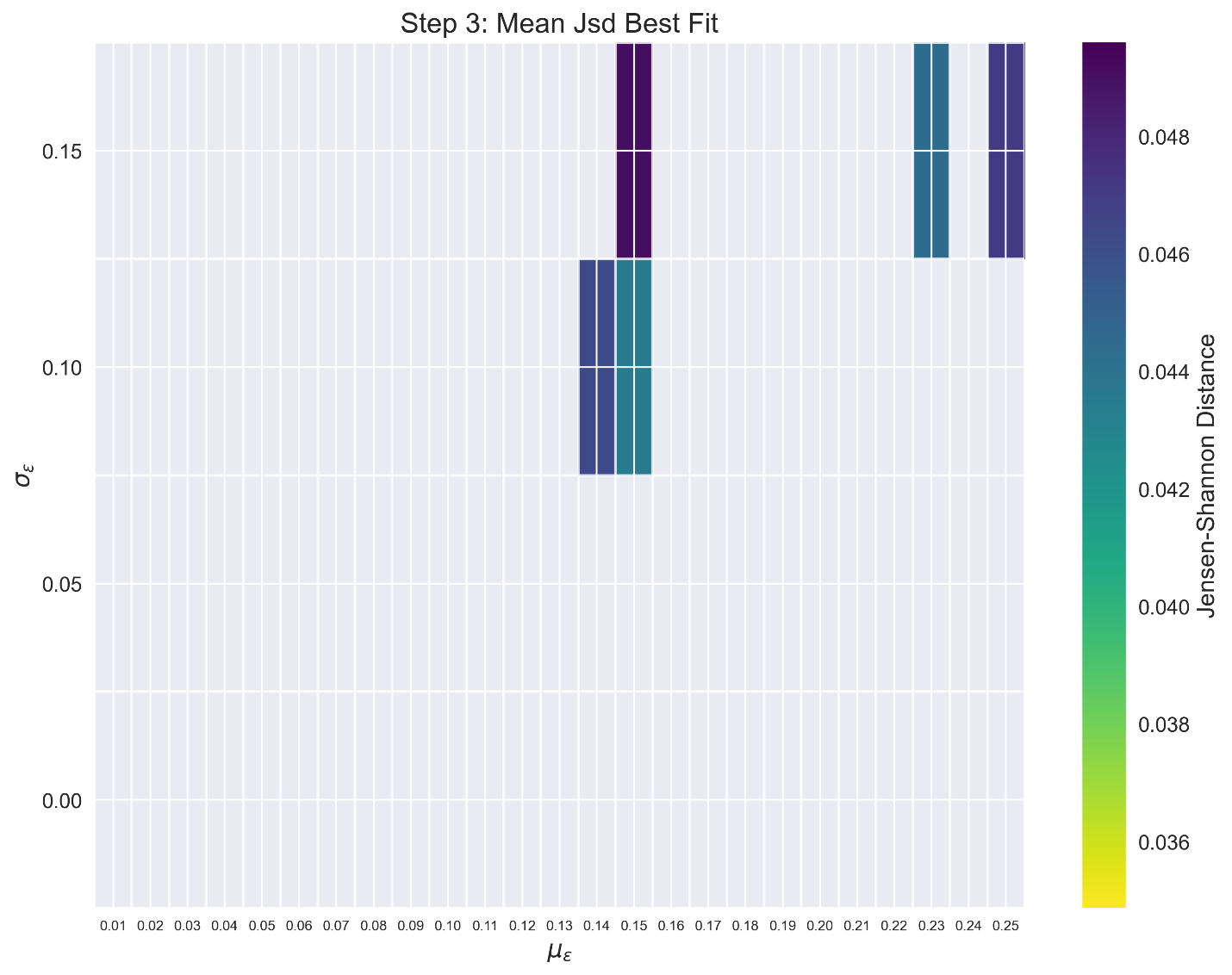
*Figure 2: Mean JSD plotted against Openness for Step 1 (top) and 2 (bottom) for all sims (grey) in comparison with only those that showed an improvement at the end of the simulation compared to the beginning (green). Simulations were run until Epsilon = 0.7, so the lack of datapoints for epsilon above 0.07 for step 1 and 0.49 for step 2 indicates that none of the simulations for high Epsilon values improved the fit compared to uniformly distributed opinions at the the beginning of the simulation.*

### Step 3

This step has variable openness (normally distributed with mean and variance of openness as parameters).

### 

*Figure 3: Best fit (of all initializations and comparisons) JSD for each parameter combination in Step 3 (HK with heterogenous openness and uniformly distributed initial opinions).*

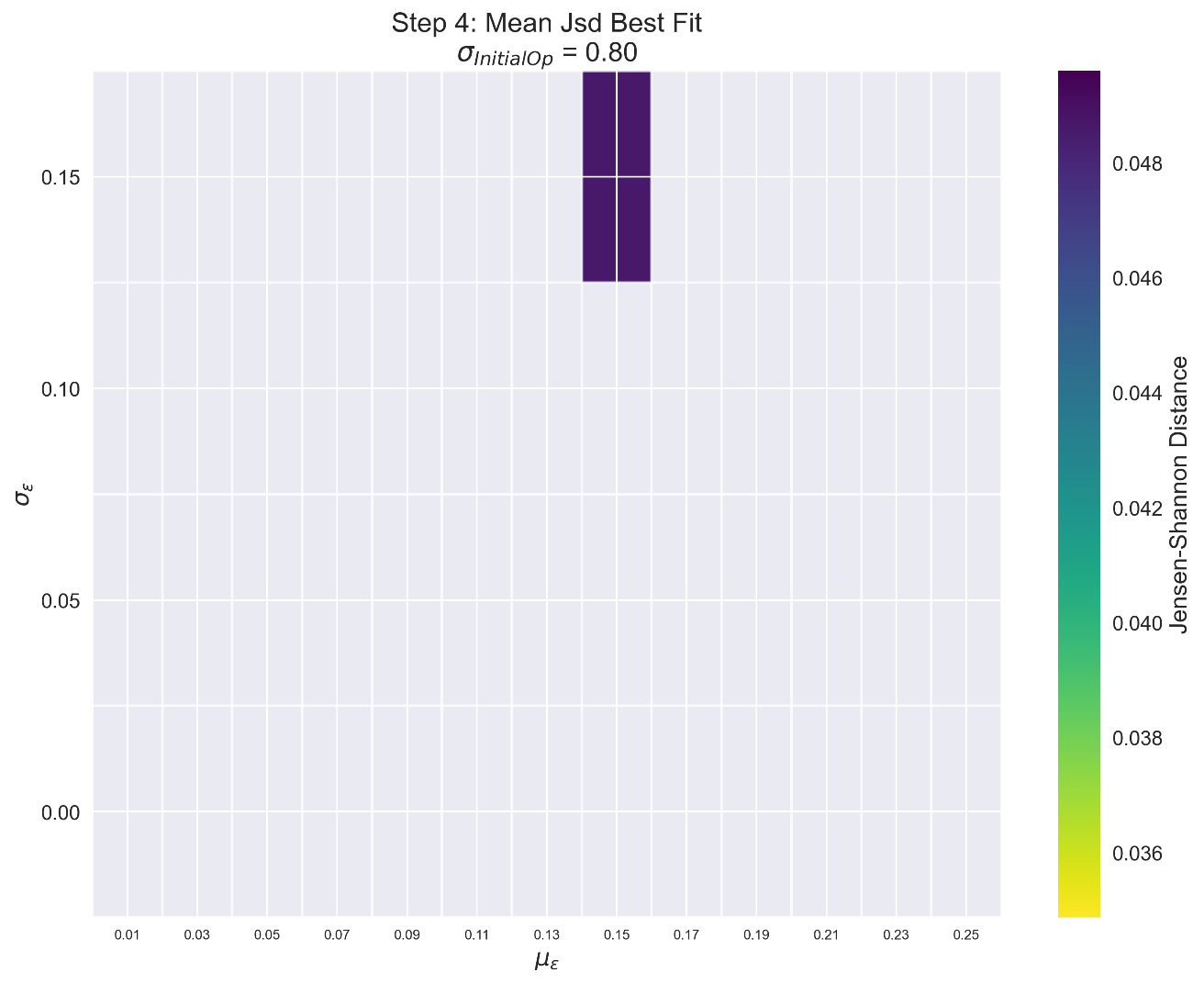


*Figure 4: Mean JSD for the Step 3 dataset* ***filtered such that maximum JSD < 0.05****. Only some parameter combinations are represented as others do not have any simulations that fulfilled the criteria for any comparison.*

### Step 4

### 

*Figure 5: Best fit (of all initializations and comparisons) JSD for each parameter combination in Step 4 (HK with heterogenous openness and normally distributed initial opinions).*

**

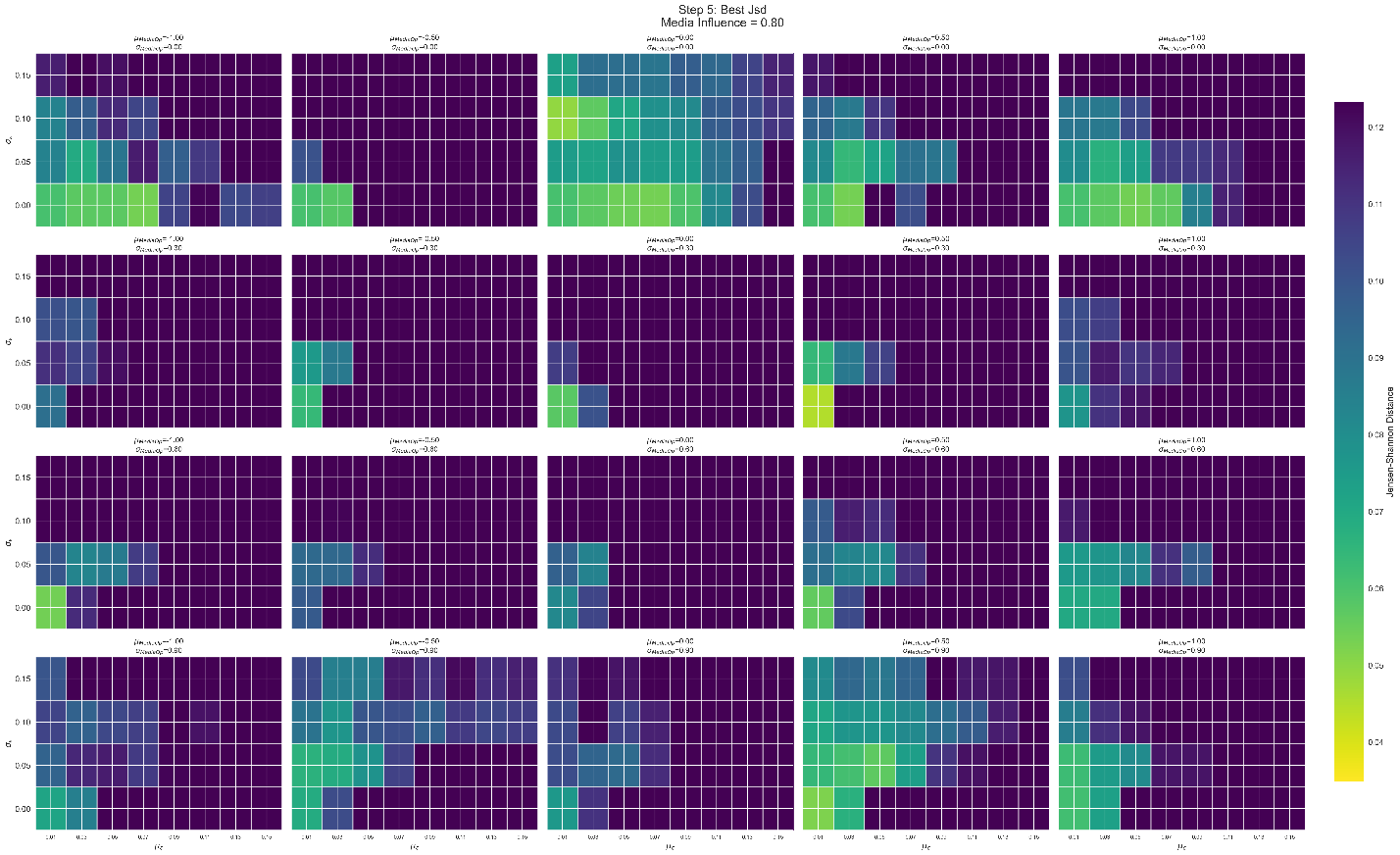
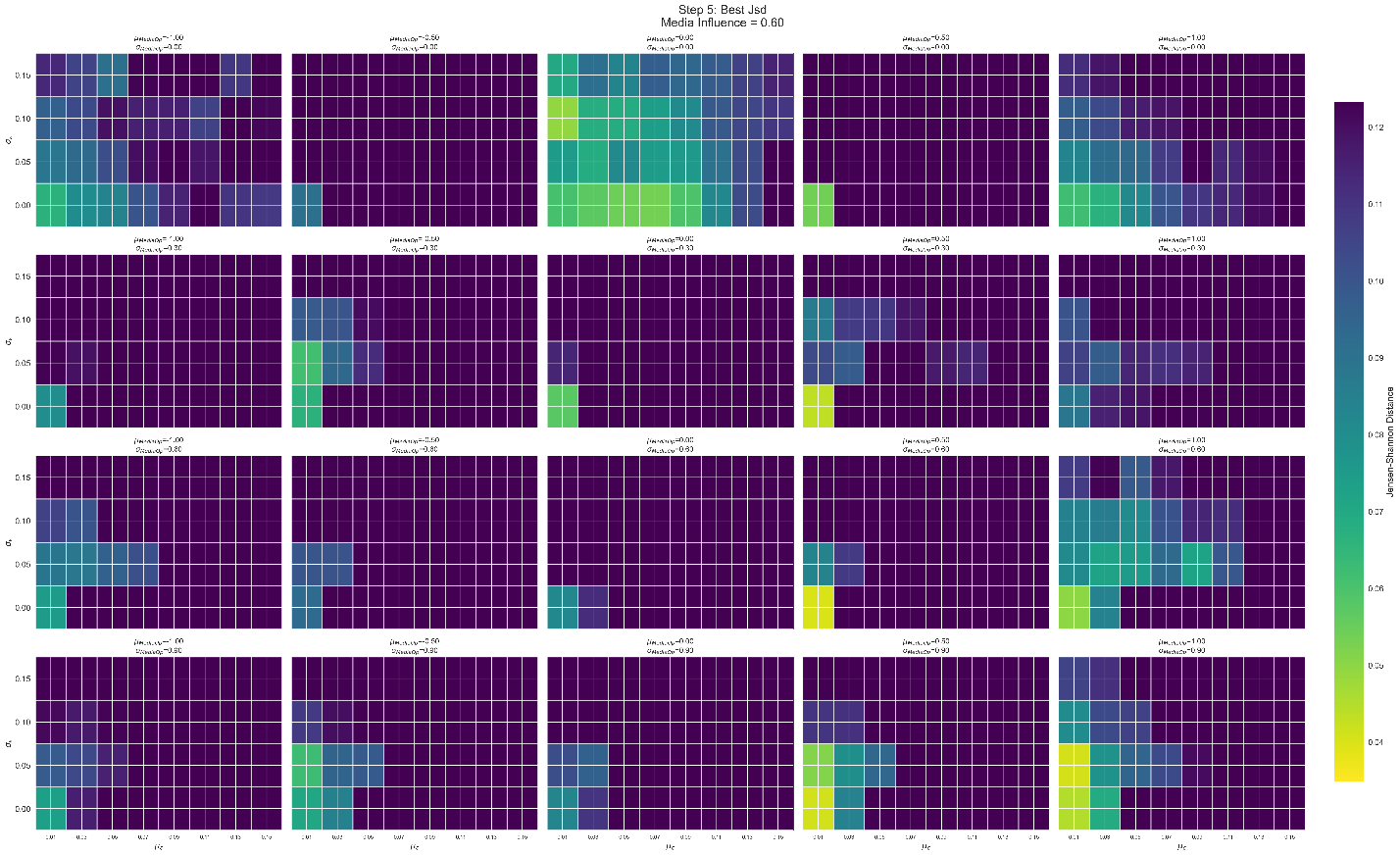
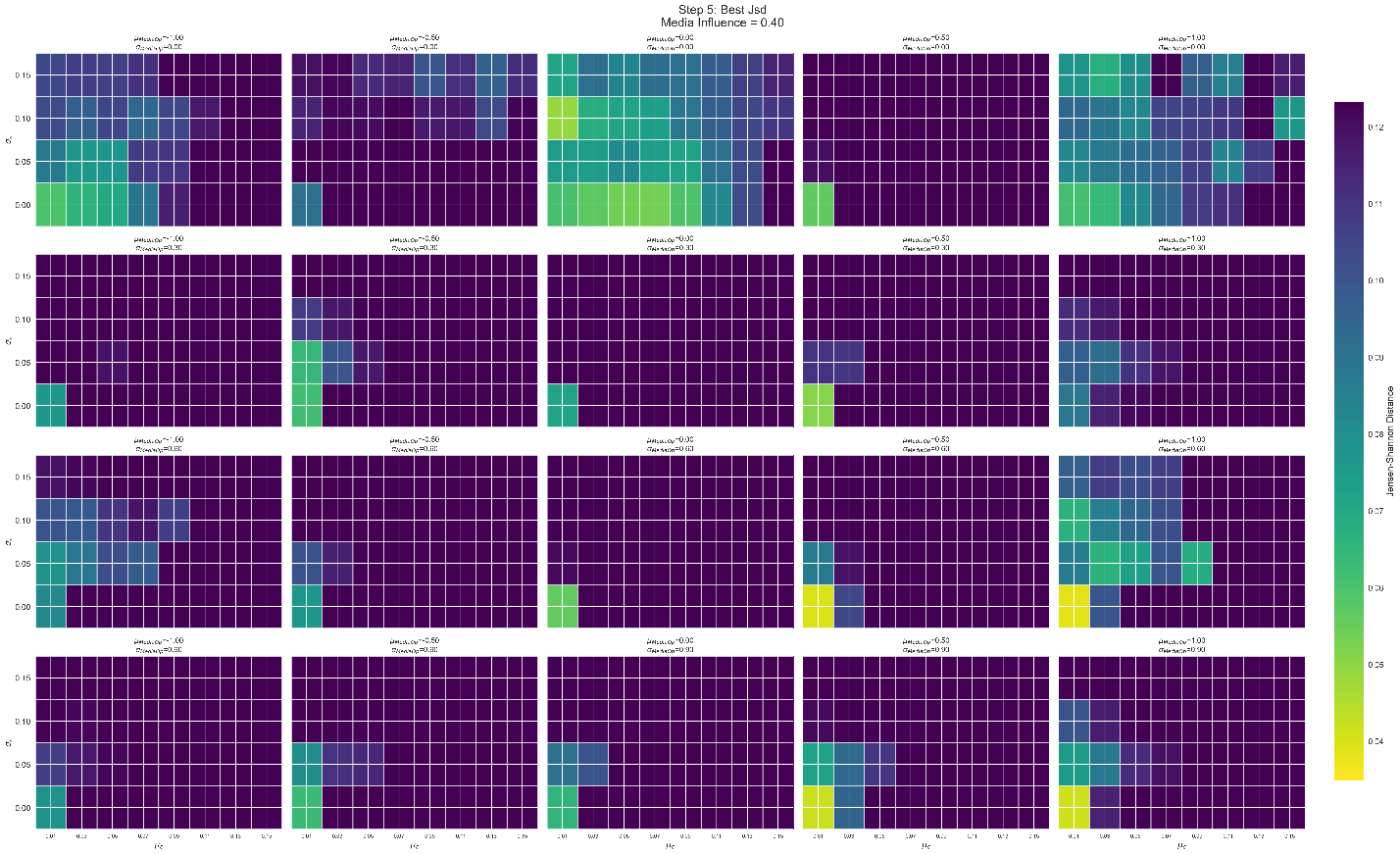
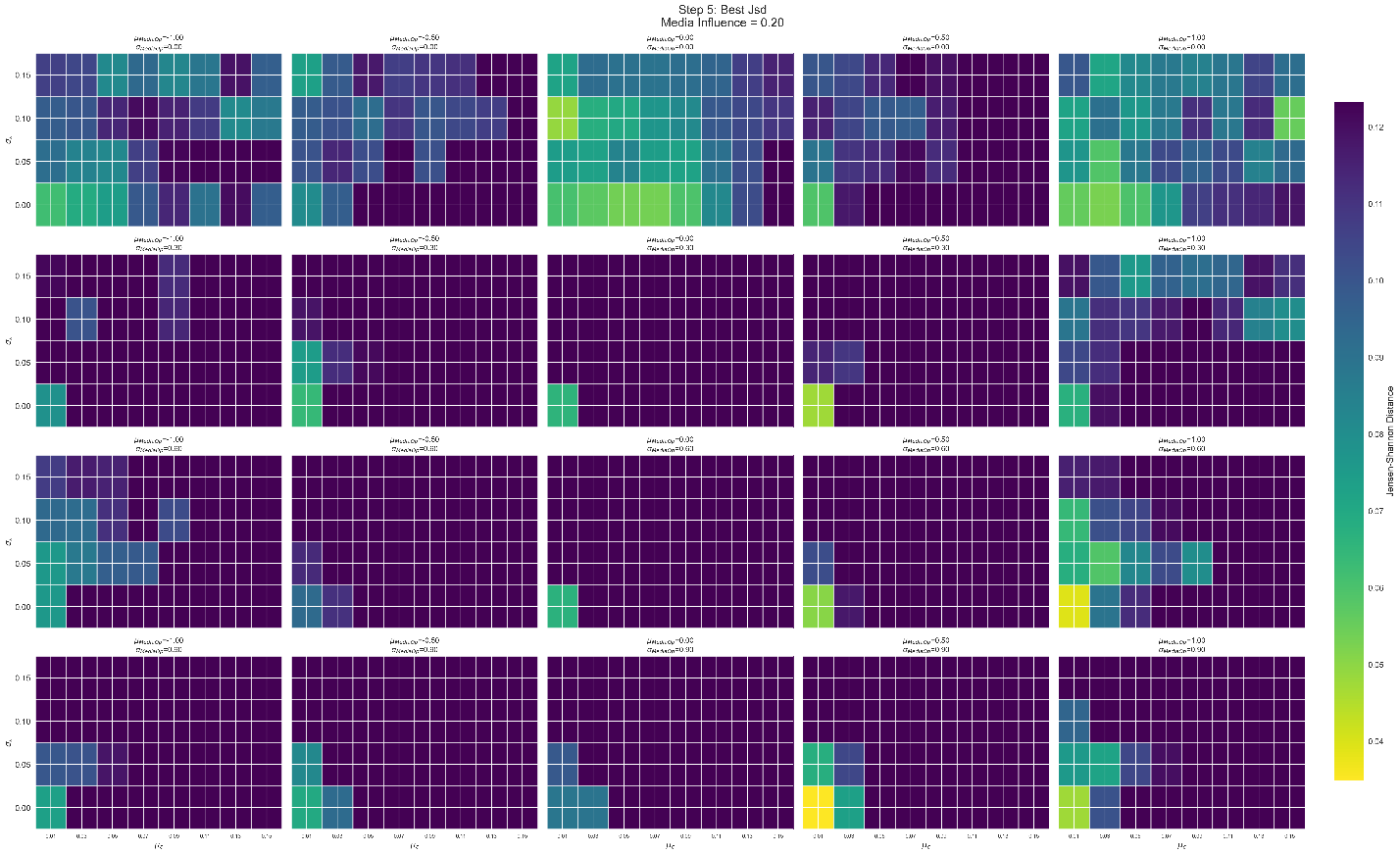
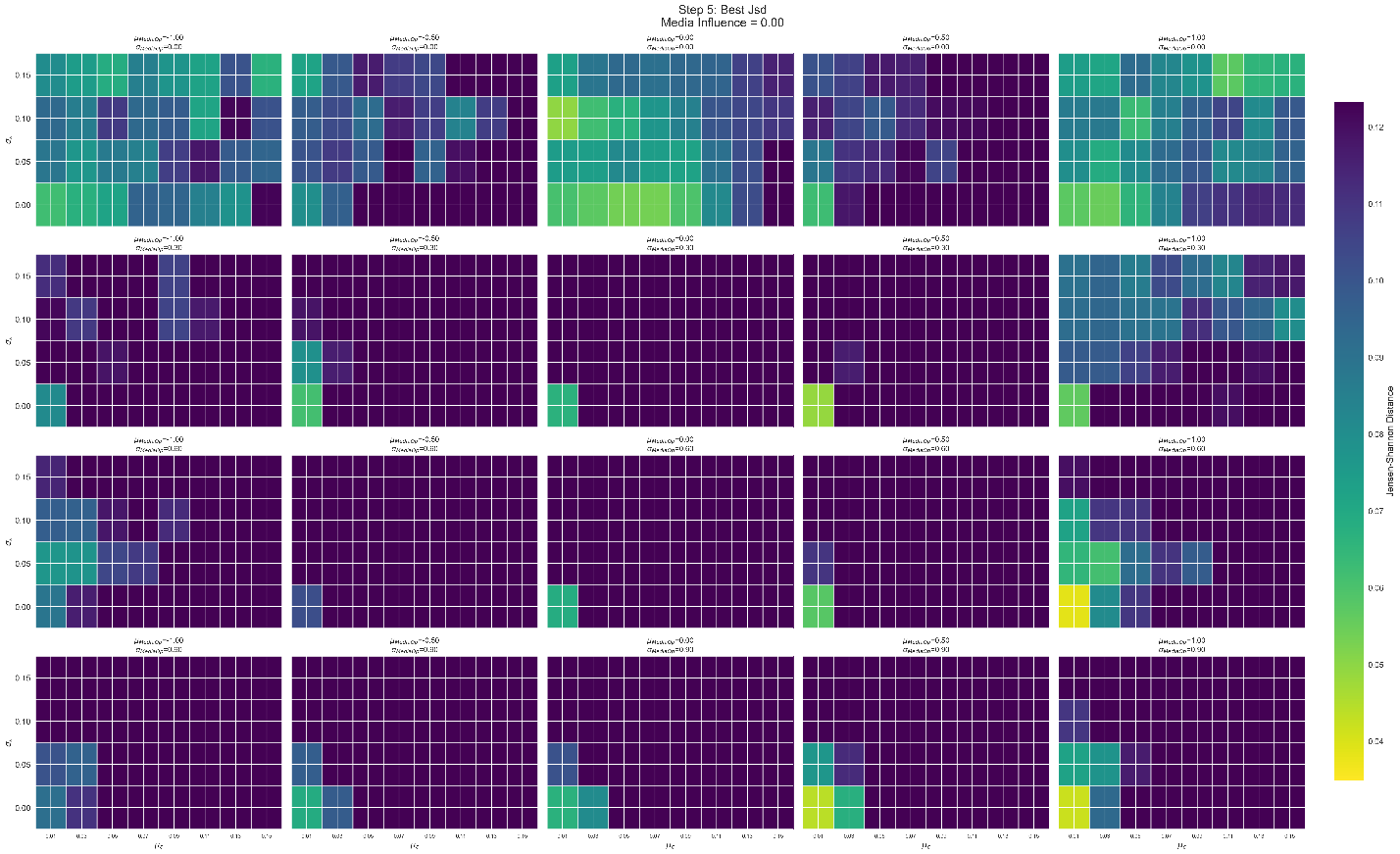
*Figure 6: Mean JSD for the Step 4 dataset* ***filtered such that maximum JSD < 0.05****. Only some parameter combinations are represented as others do not have any simulations that fulfilled the criteria for any comparison.*

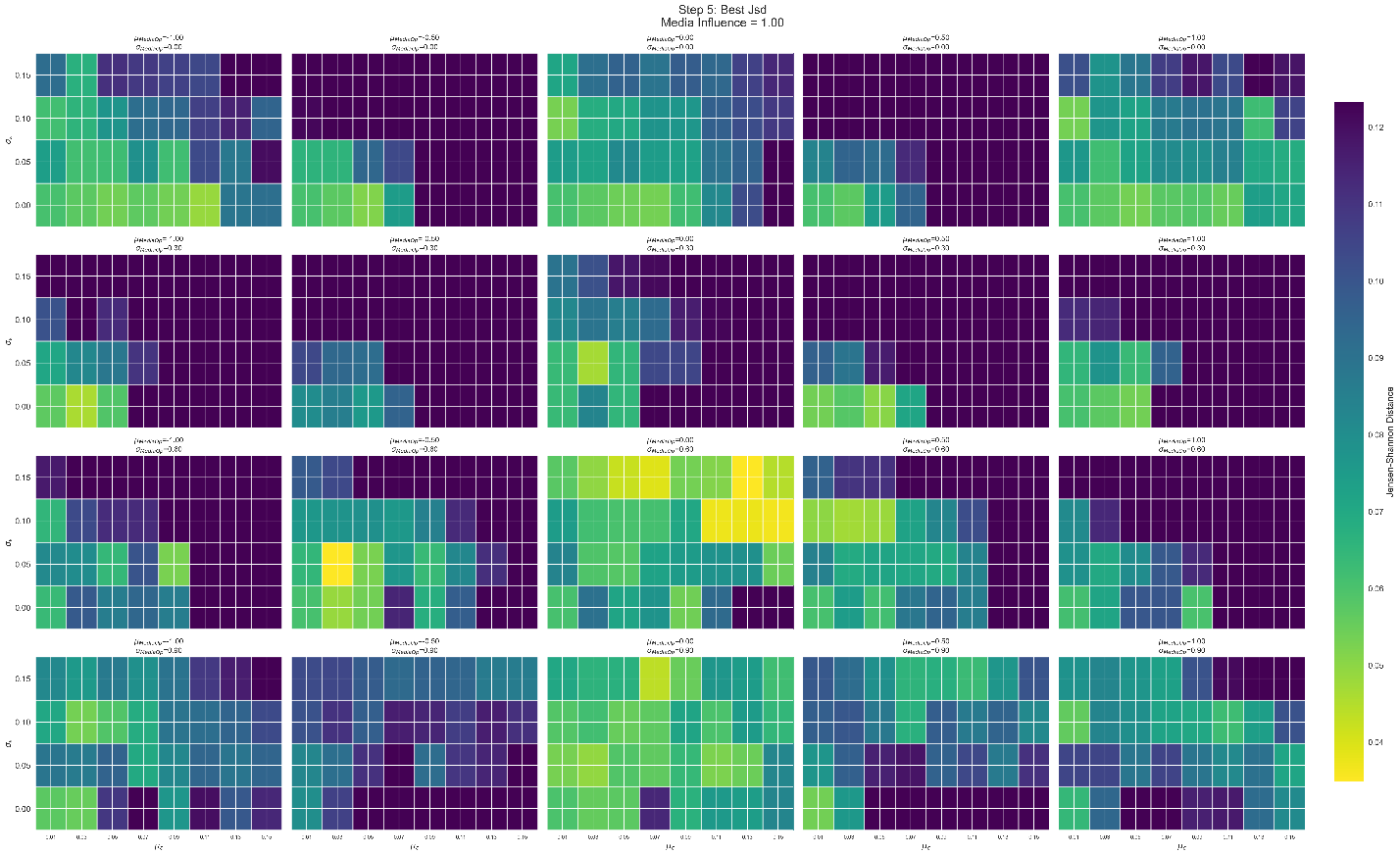
### Step 5 – Best fit plots

The following plots all show the best fitting comparison within a given parameter combination (across all initializations and comparisons).

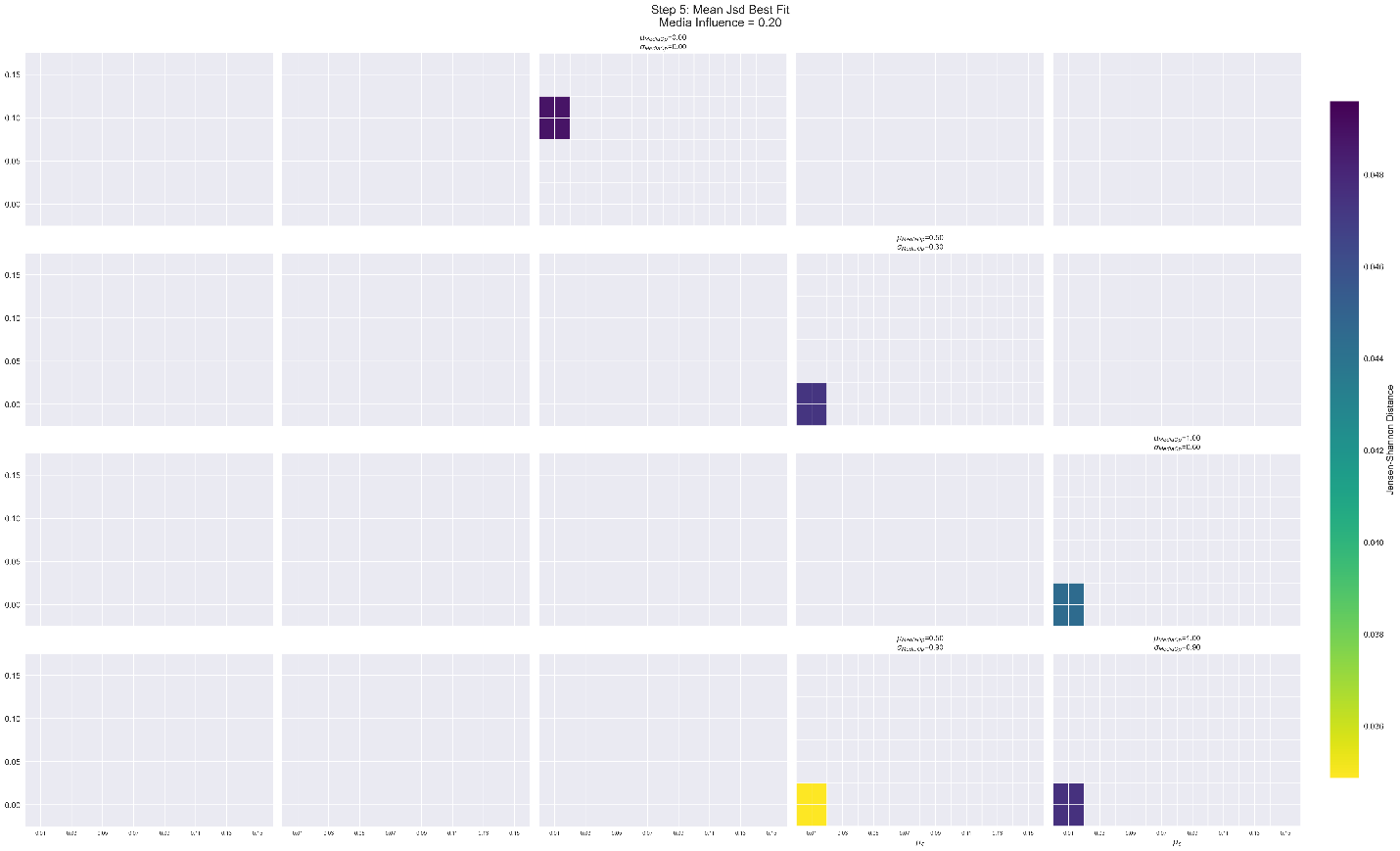
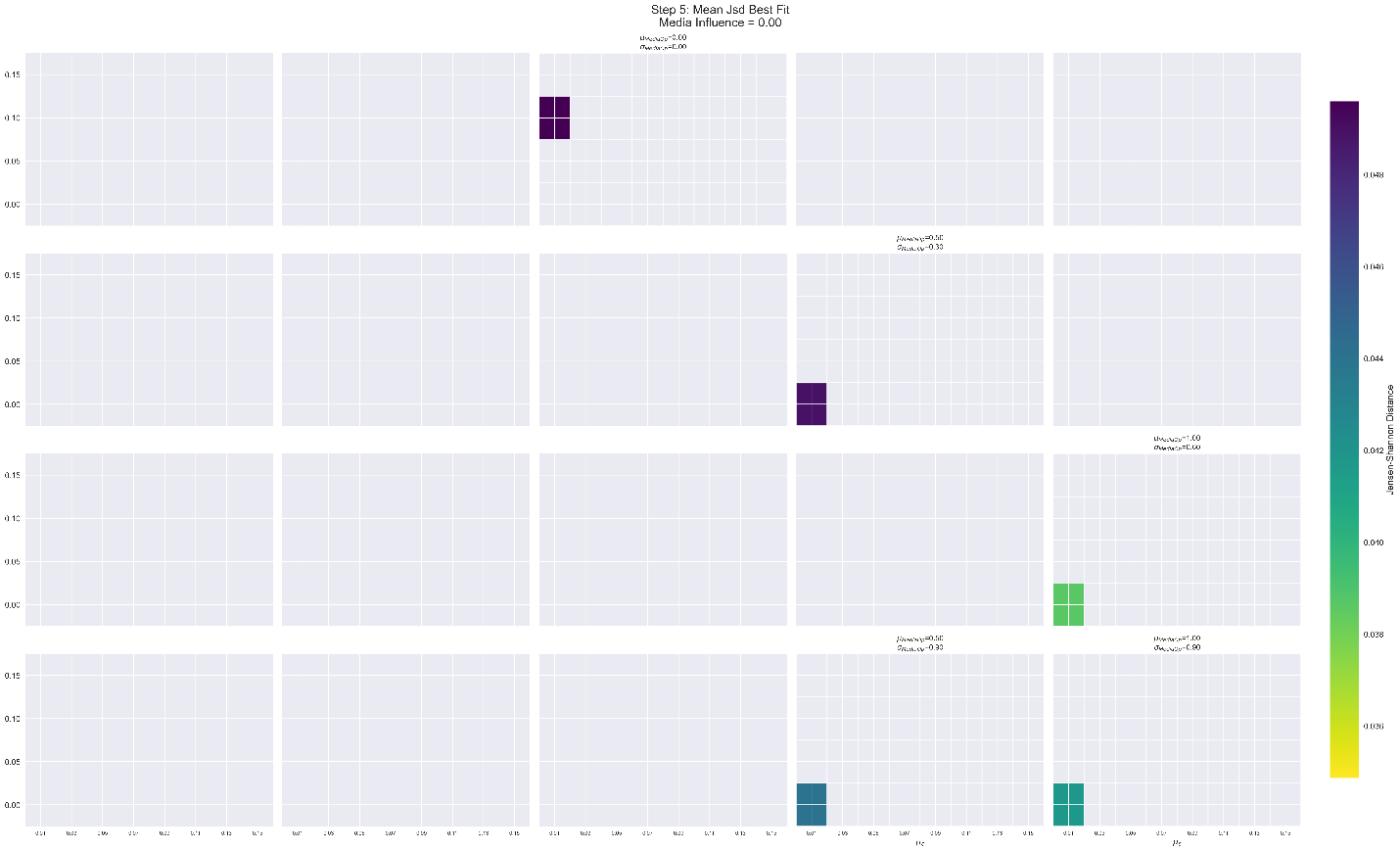
There are five plots, each containing 20 subplots. The different (larger) color plots all represent different values of the **Media Influence Factor** (fractional weightage assigned to opinions from media agents, 0 = no influence, 1 = full influence). This value increases as you go top to bottom from one image to another in this document. Within each image is 20 subplots – as you go left to right along the subplots the **mean media position** increases from -1 to +1 in intervals of 0.5 (while comparing to survey data -1 is anchored to the ‘left’ and +1 to the ‘right’ ends of the 10-point survey). As you go from top to bottom within a plot, each subplot has increasing **media standard deviation** from 0 to 0.9 in intervals of 0.3. Note however that within each panel/subplot, the y-axis (openness Standard deviation) increases from bottom to top like in the other color plots (this may be confusing so I should change it in next iterations).

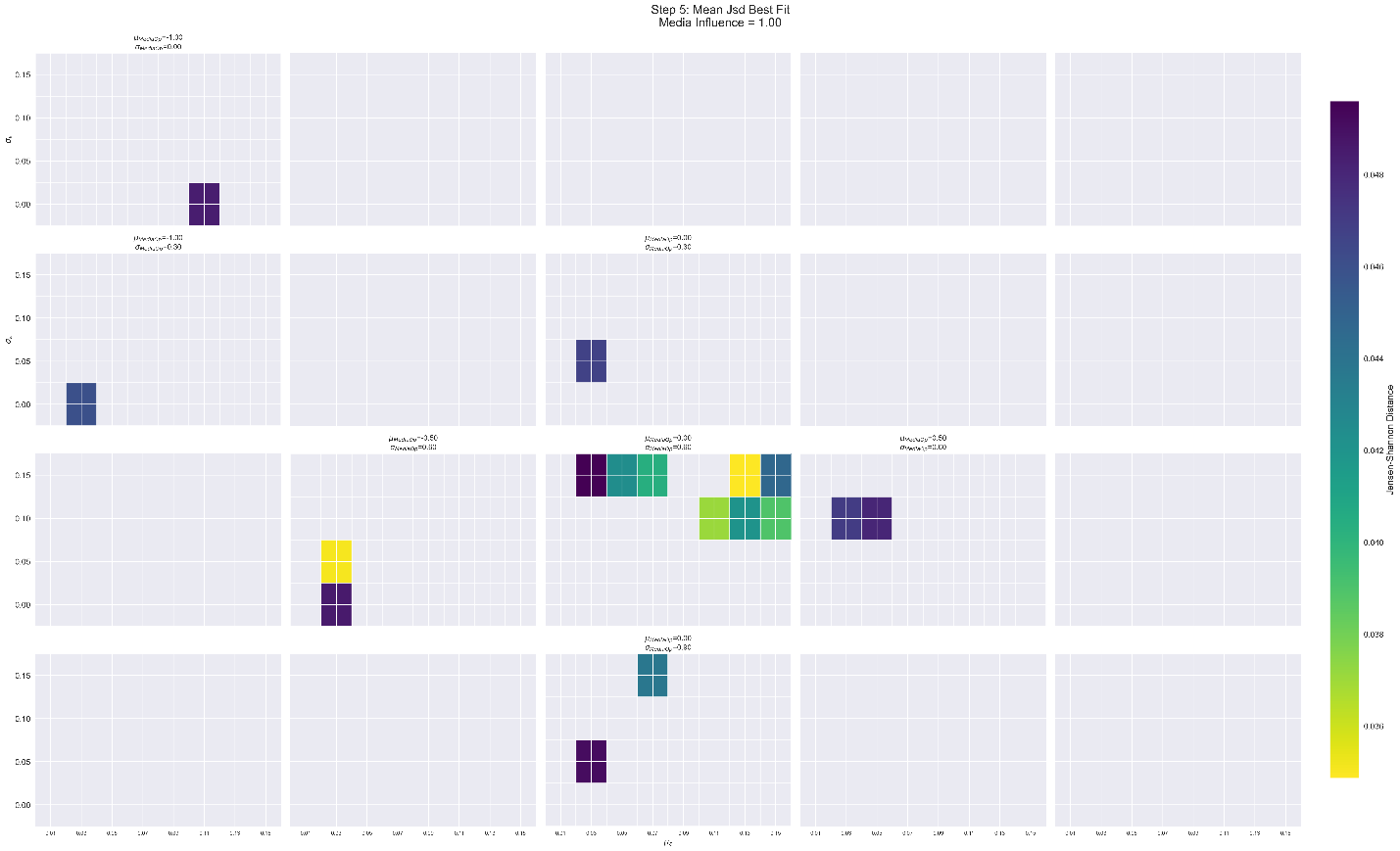
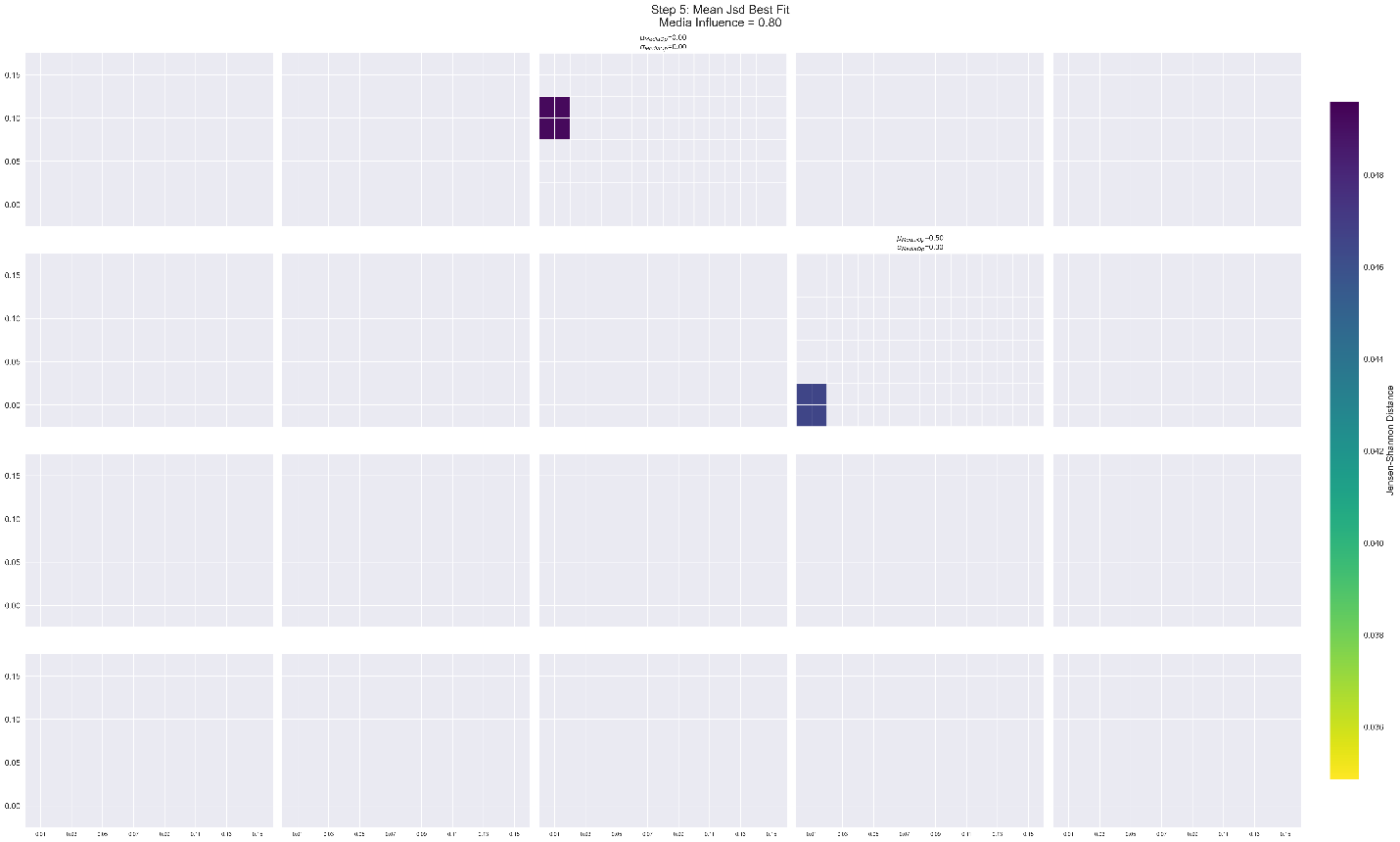
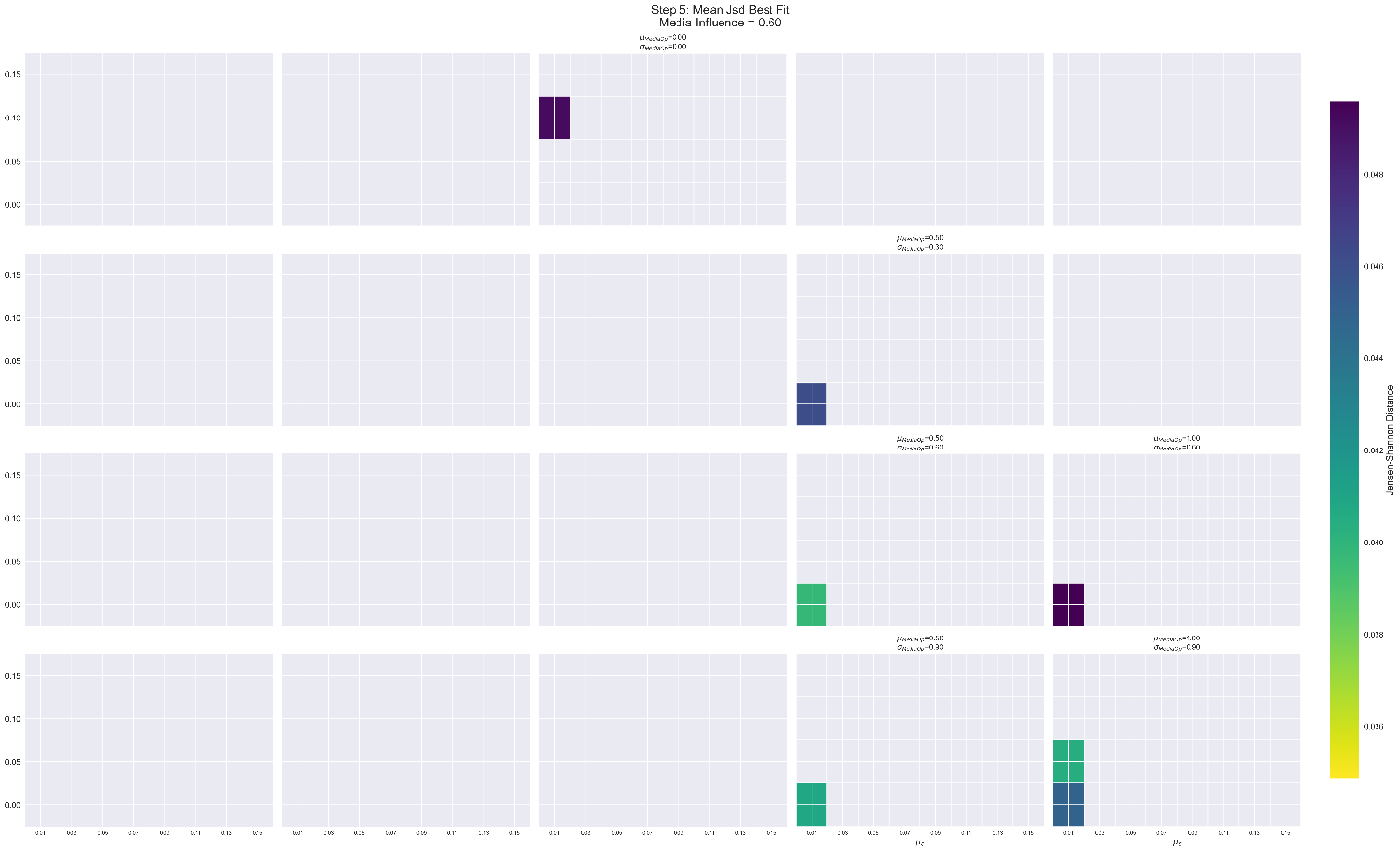
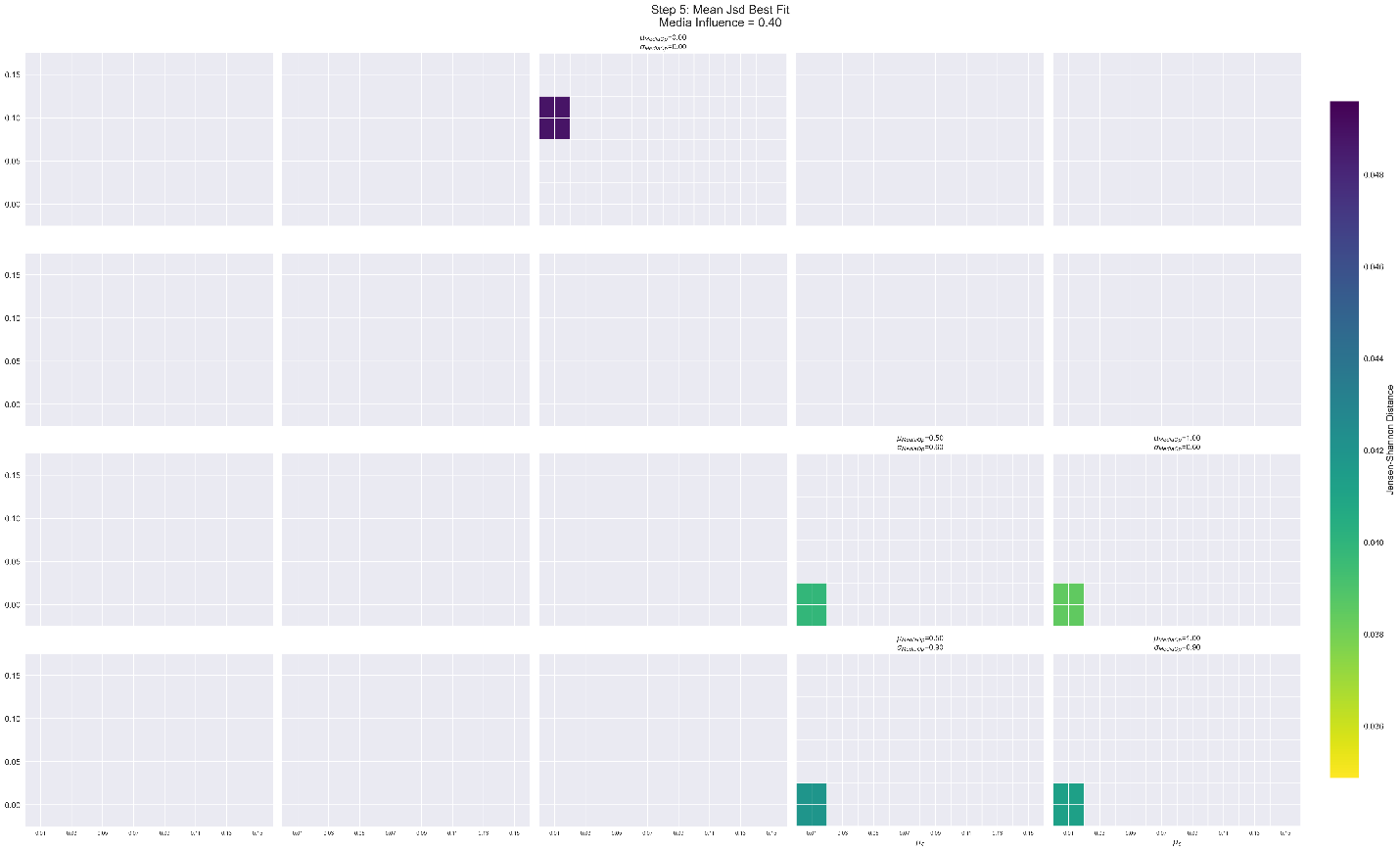
PS: The media distributions aren’t actually random, they are drawn from a deterministic approximation of a normal pdf.





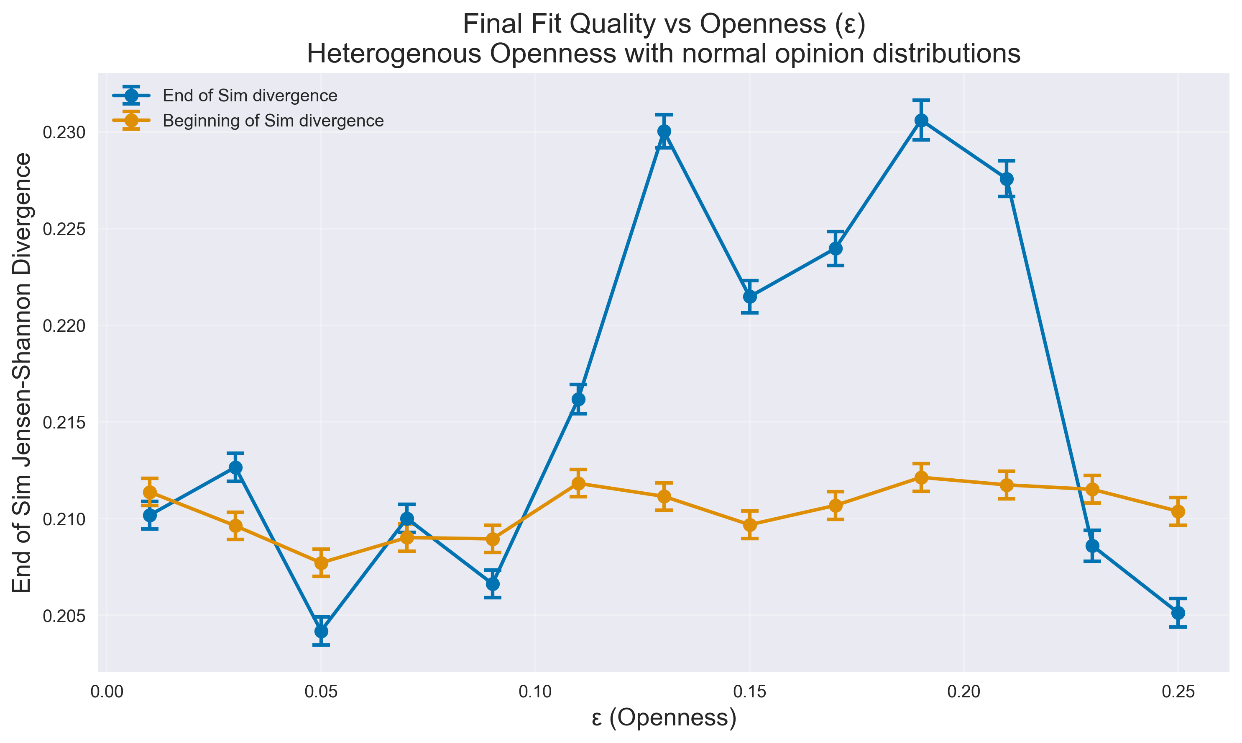
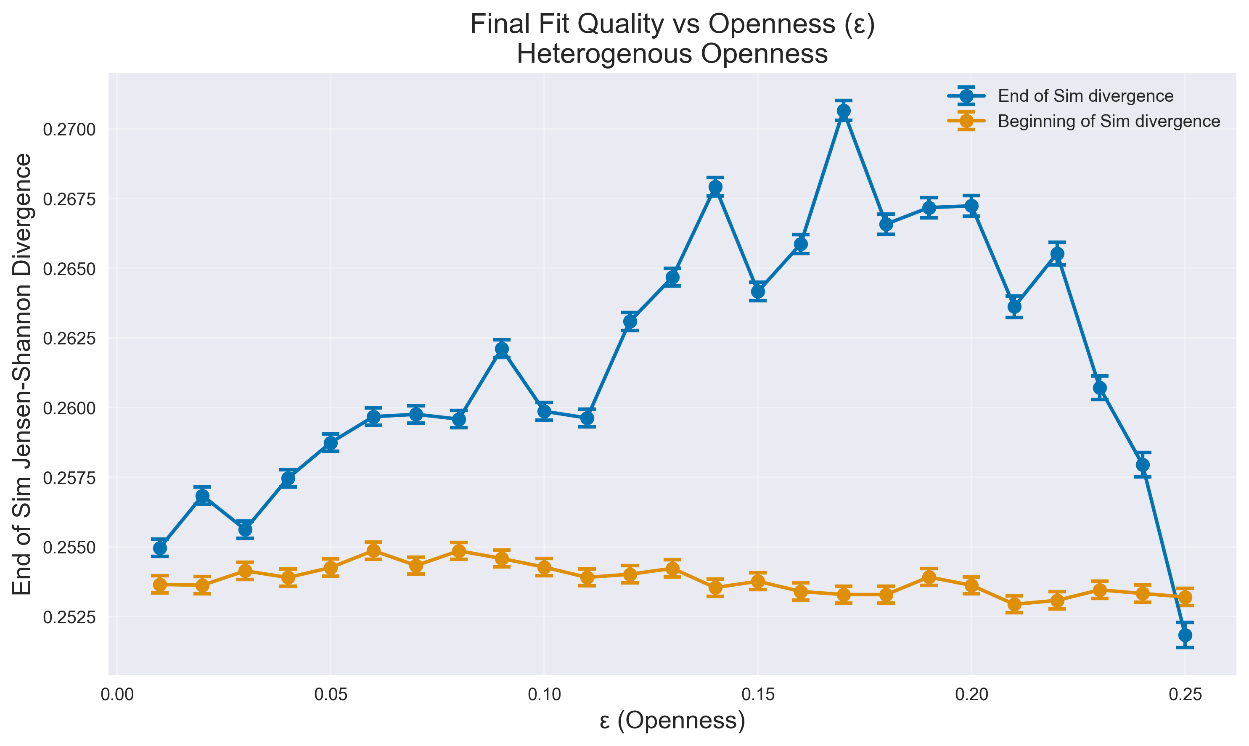
### Step 5 – Filtered by fit (JSD < 0.05 after simulation)



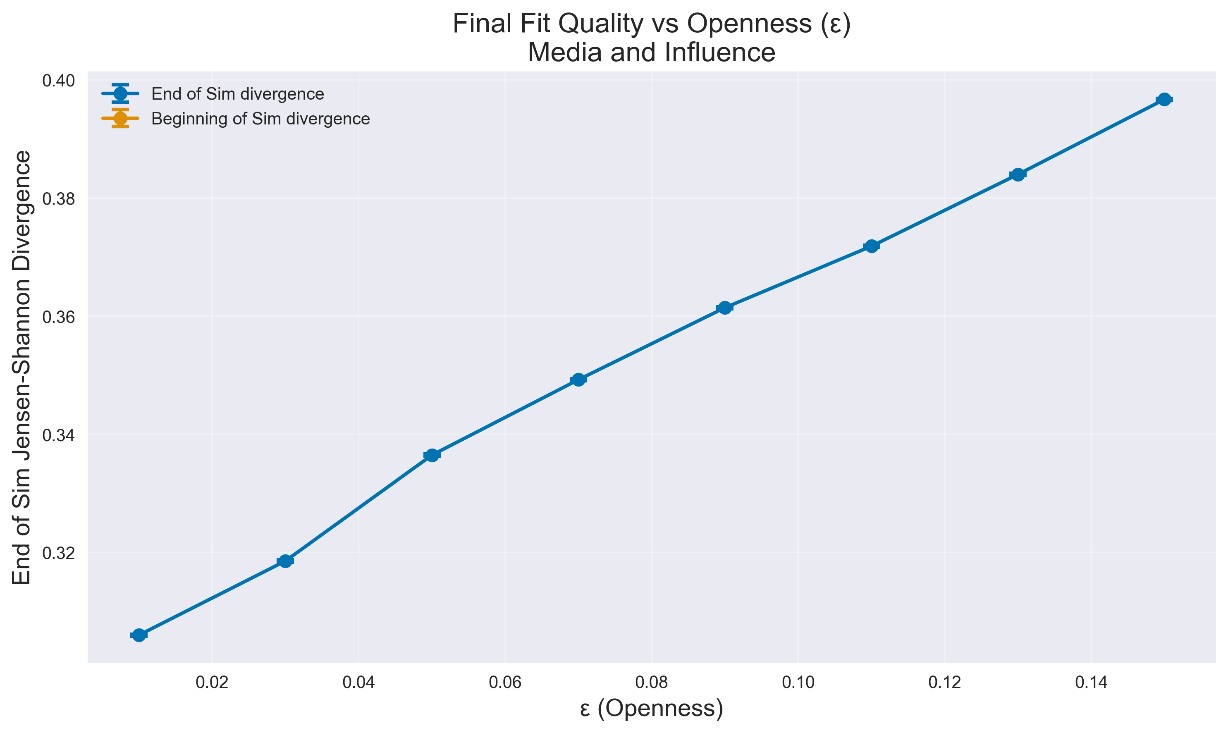


### JSD vs Epsilon for steps 3, 4, and 5

These are less informative plots as there are more parameters than just epsilon in these models. A small number of simulations for steps 3 and 4 show final JSD to be less than 0.05 (0.2% of all step 2 sims and 0.3% of all step 3 sims).



*Figure A: Mean initial and final JSD plotted against Openness for Step 3 (top) and 4 (bottom) for all sims. Some subranges show high fit improvement on average.*



*Figure X: Mean initial and final JSD plotted against Openness for Step 5 for all sims. Initial opinions were unavailable for this plot, but this analysis will be re-run with new data.*