Model Flexibility Analysis

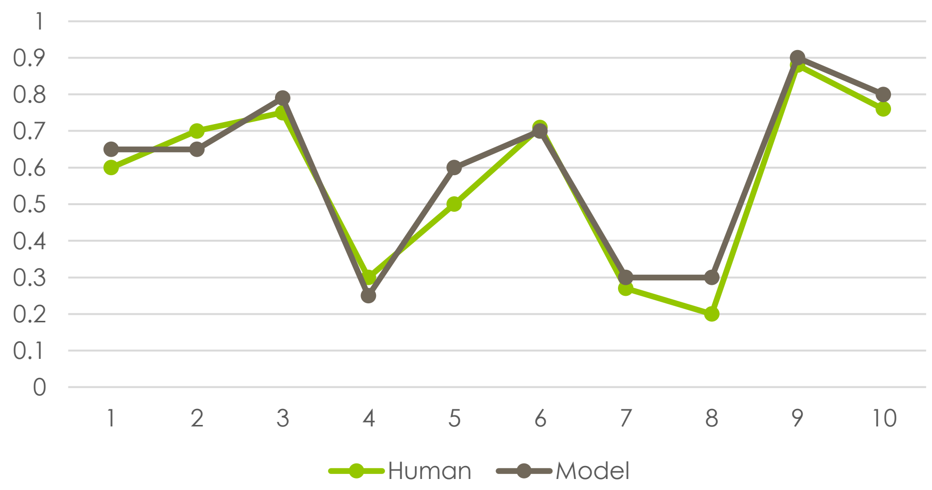
Vladislav "Dan" Veksler
Christopher W. Myers
Kevin A. Gluck

Model evaluation

- How well does the model fit the empirical results?
- Could the model have also fit any other data (i.e., to what degree is the model falsifiable)?

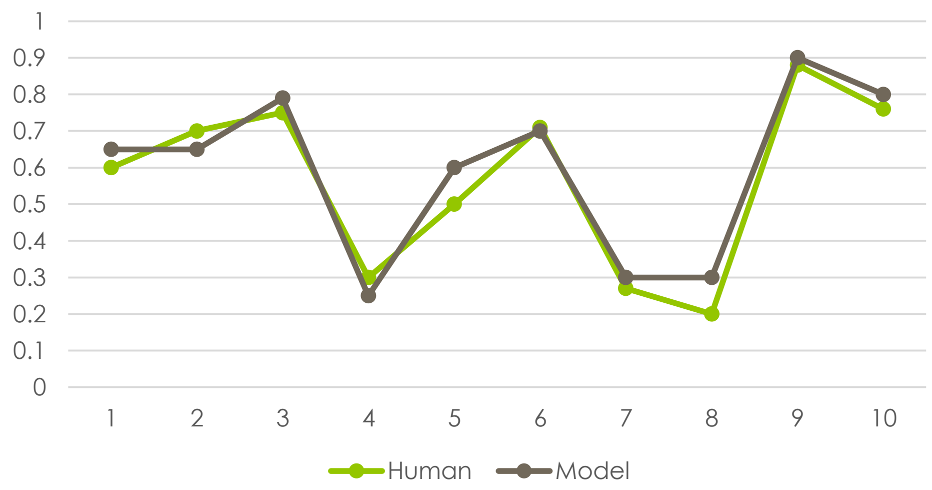
Does the model fit the data?

- Is this a good fit?



Does the model fit the data?

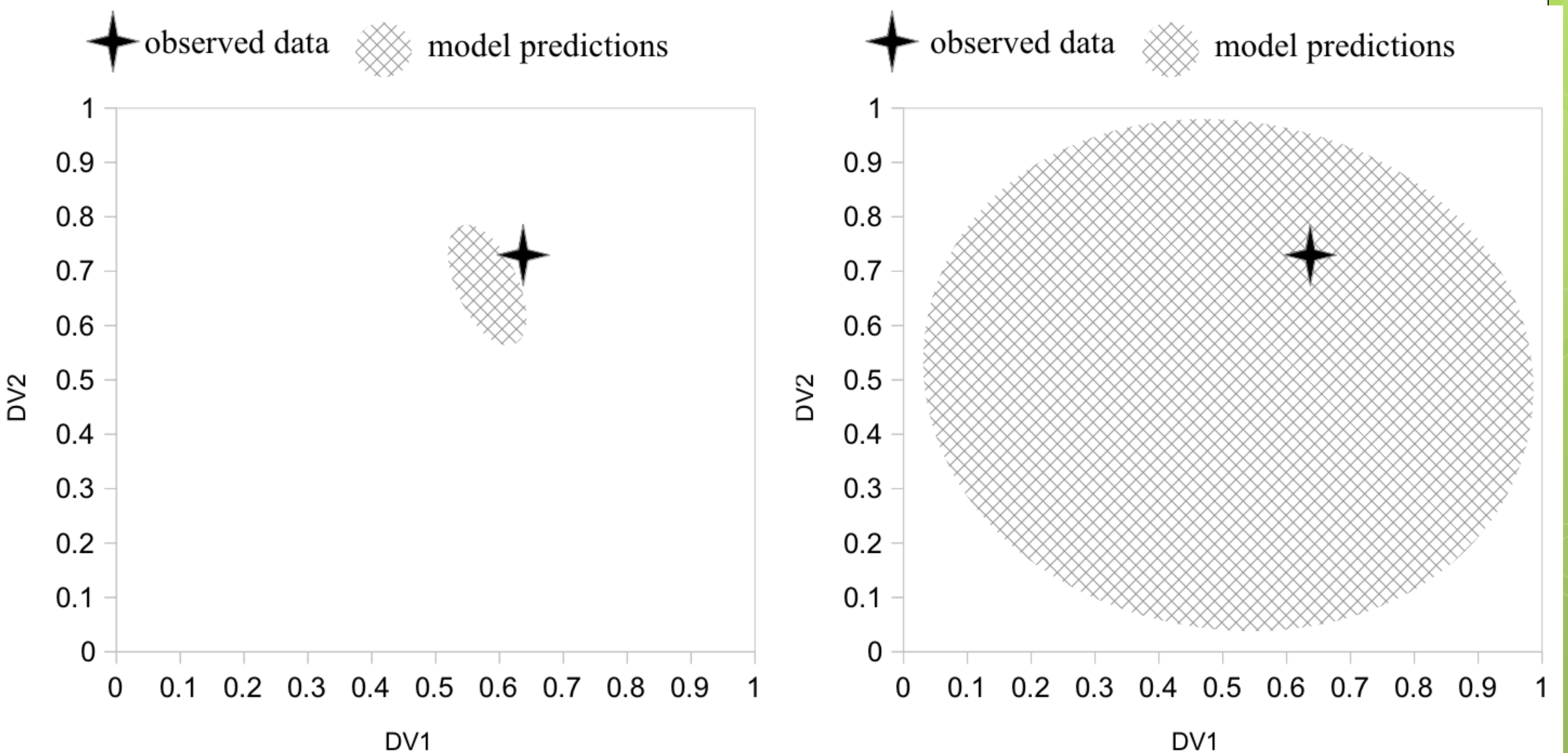
- What if it was outside of the confidence intervals?



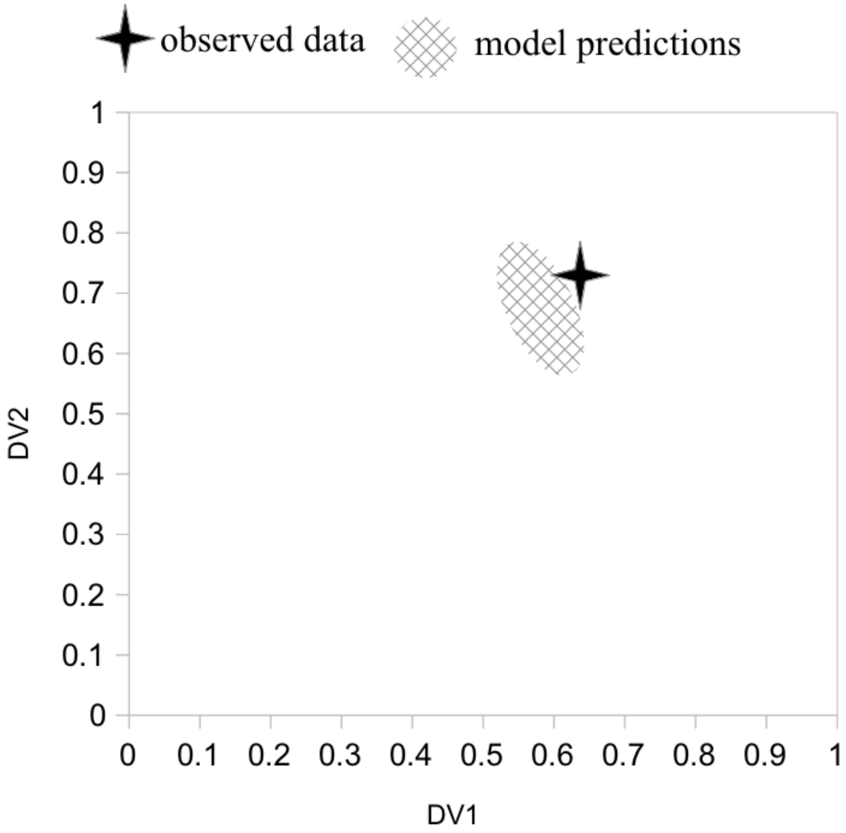
Best method for evaluating model fit

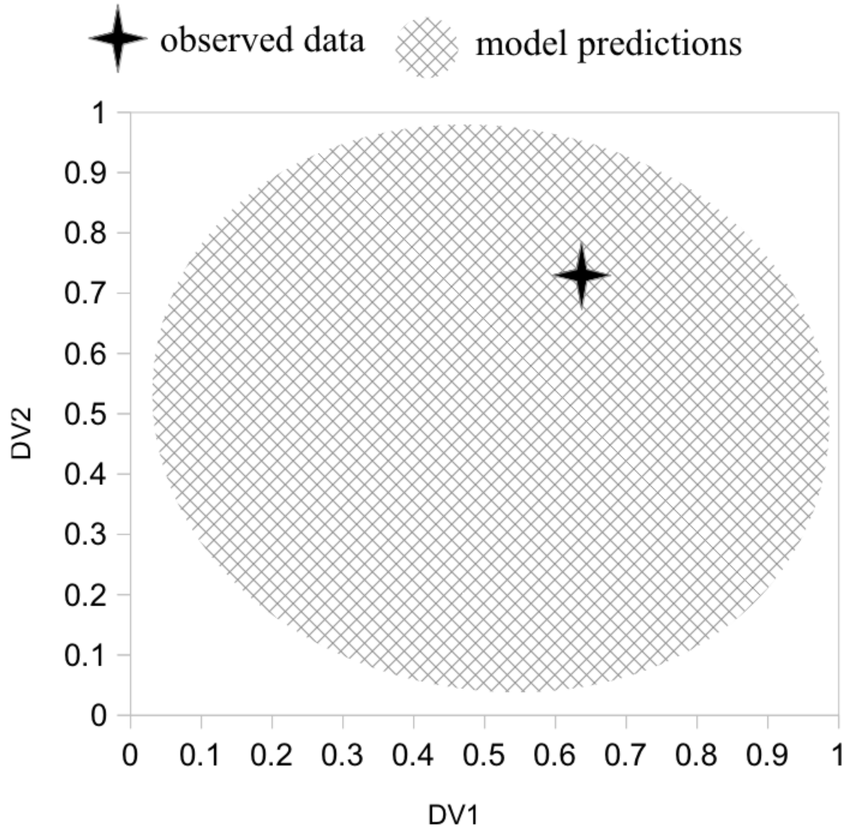
- RMSE
  - mean behavioral tendencies
- r-squared
  - trend of mean tendencies
- likelihood
  - behavior distribution

Which model provides a better fit to empirical results?



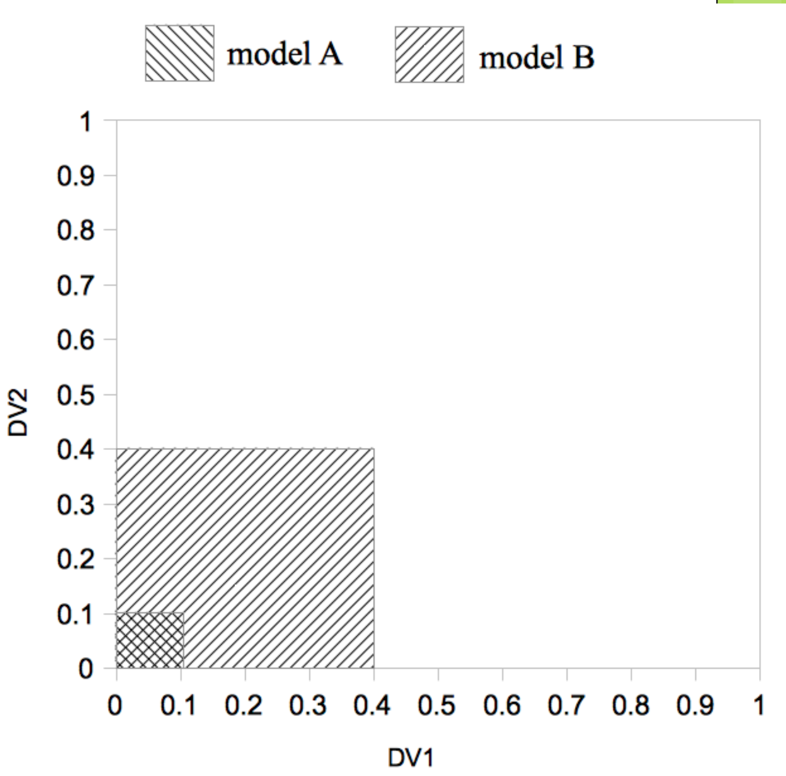
Which model is more convincing?





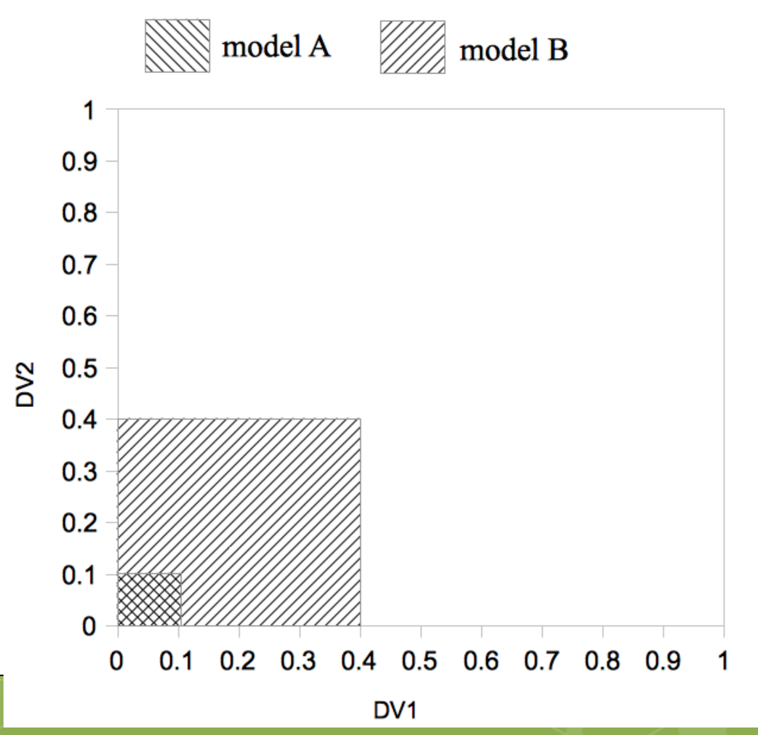
Model flexibility

- Model flexibility (often referred to as model complexity)
  - how likely the model is to fit any one dataset
  - i.e. what proportion of hypothetical dataspace do model predictions cover



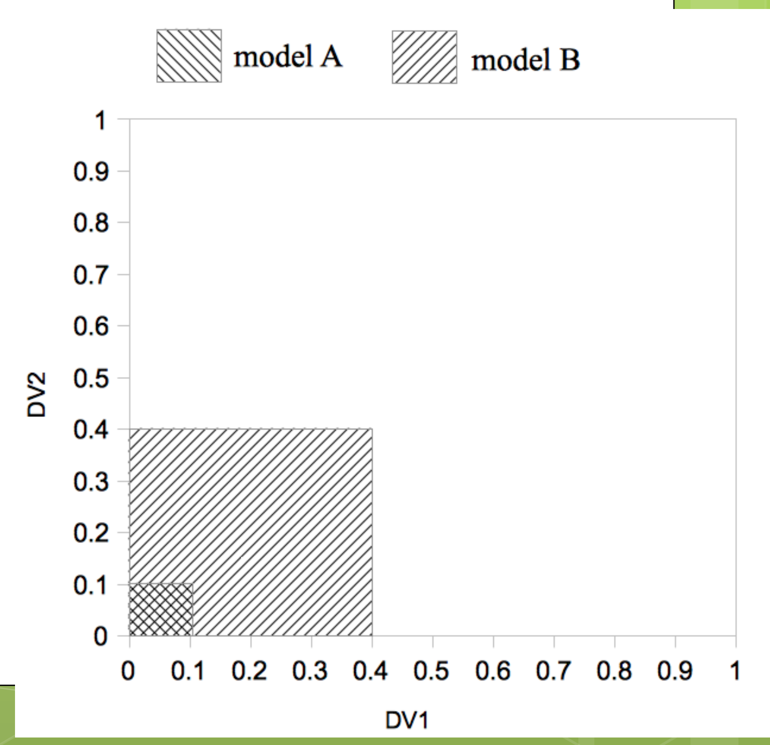
Number of Free Parameters

- AIC, BIC, adjusted r-squared
- Poor estimate of model flexibility
  - assume x, y, and z are all between 0 and 1
  - A(x,y,z) = (.1x, .1y + .001z)
  - B(x,y) = (.4x, .4y)
  - C(x) = (x, .5x + .5sin(10000x))



Why use latent estimates?

- Why not directly calculate the volume of model predictions?

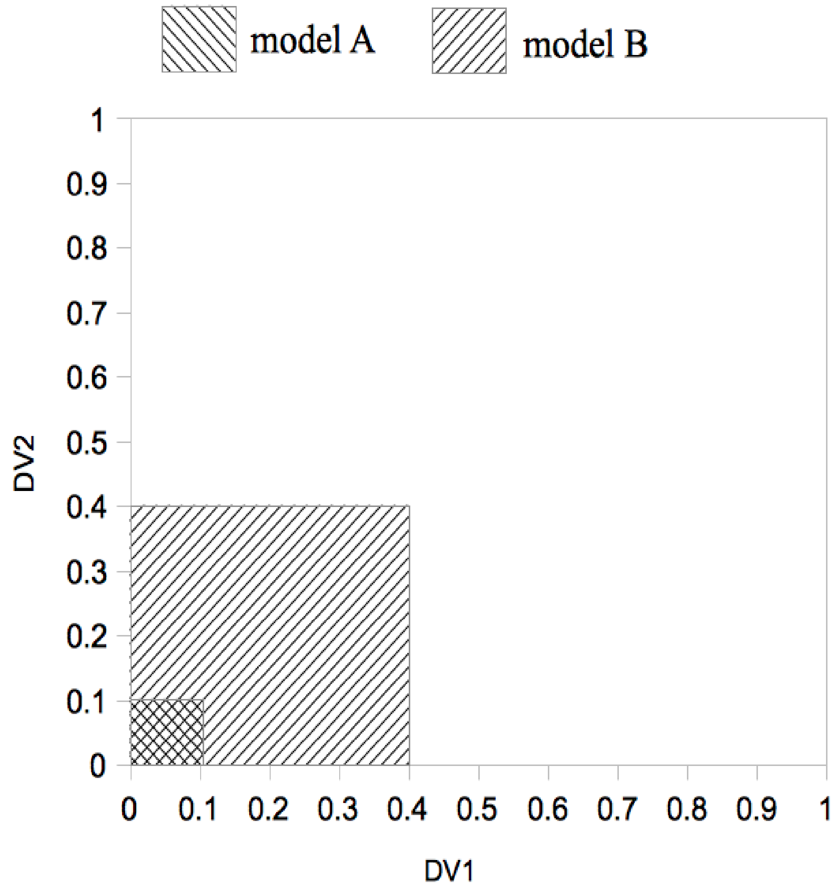


Model Flexibility Analysis

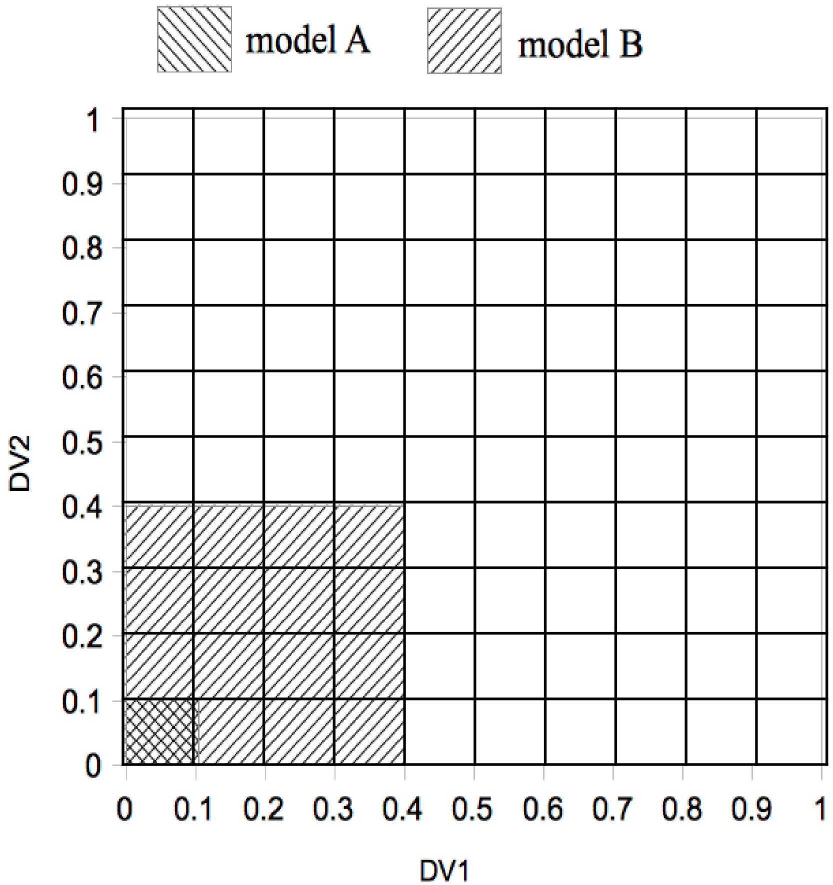
- Veksler V. D., Myers C. W., & Gluck K. A. (2015) Model Flexibility Analysis. *Psychological Review*, 122(4) 755-769.

  - lay a grid over the space of potential behavior
  - estimate model flexibility as the proportion of grid cells that include model predictions

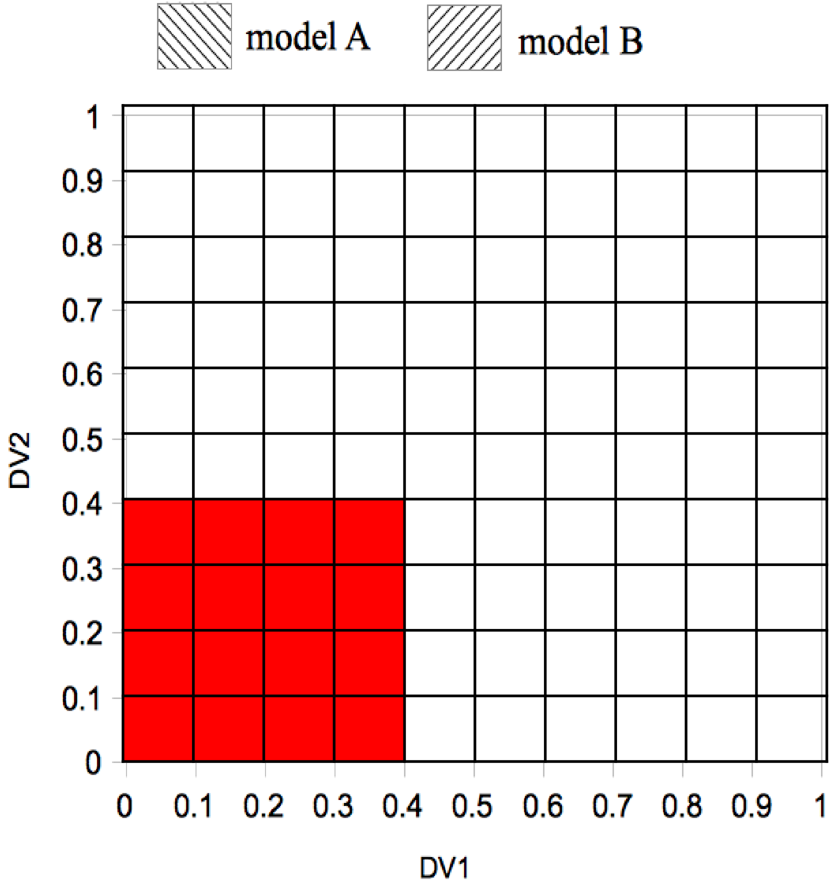
Model Flexibility Analysis



Model Flexibility Analysis



Model Flexibility Analysis



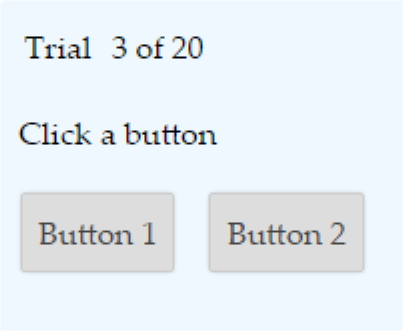
Model Flexibility Analysis

R code for computing MFA. Potential range of each behavioral measure is assumed to be between 0 and 1. If the potential range is different, data should be scaled for use with this function.

```
mfa = function( modelPredictions ){
        numberOfMeasures = ncol( modelPredictions )
        numberOfPredictions = nrow( modelPredictions )
        totalCells = numberOfPredictions
        granularity = totalCells ^ ( 1 / numberOfMeasures )
        cellPredictions = floor( modelPredictions[,] * ( granularity - .0000001 ) )
        return ( nrow( unique( cellPredictions ) ) / totalCells )
}
```

Two-arm bandit

- Reward probabilities: .7, .3



- Reinforcement learner with two free parameters:
  - alpha (learning rate)
  - egs (expected gain noise)

MindModeling

Powered by μBatch

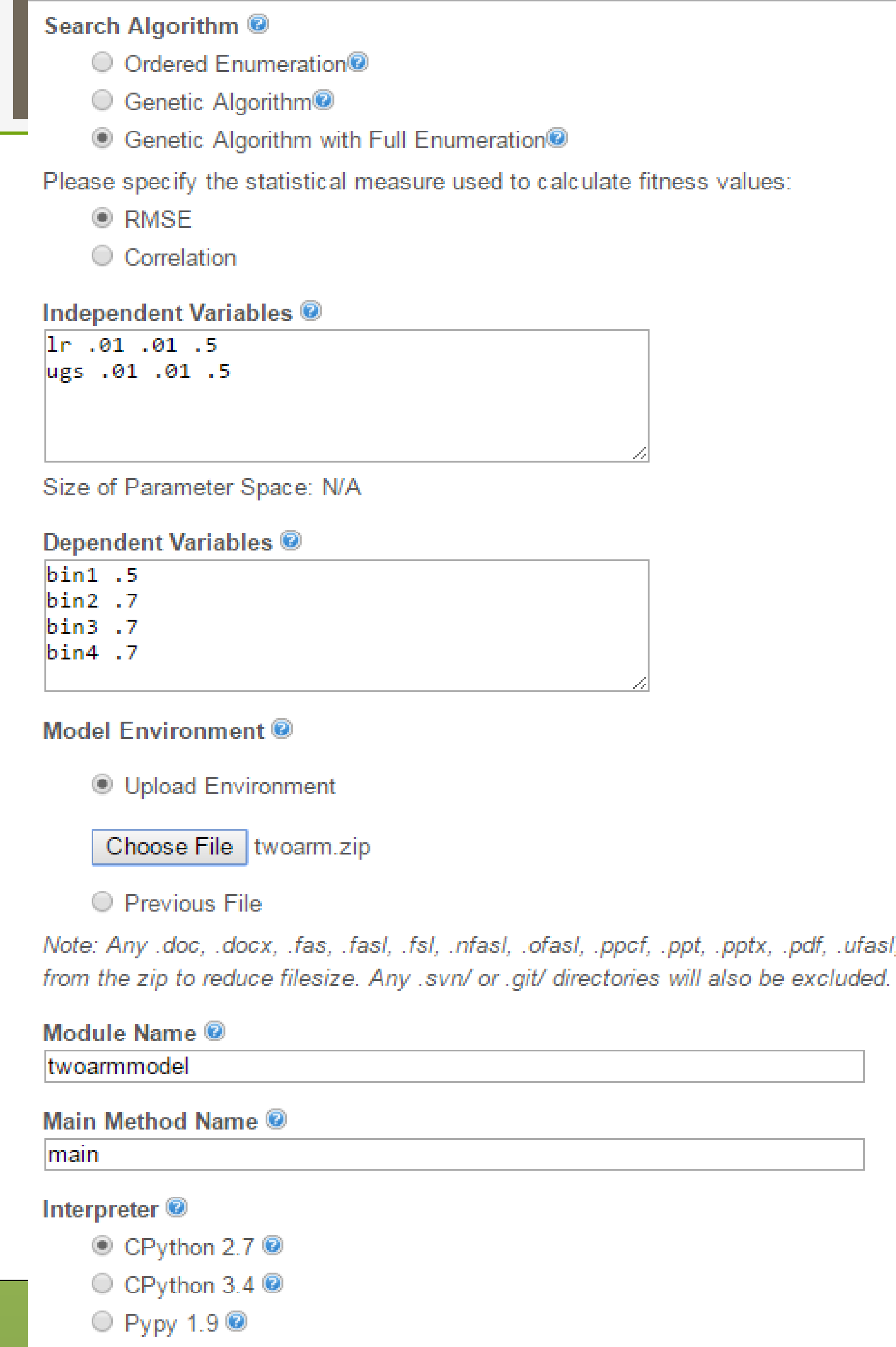
Jobs | Projects | **Submit** | Statistics | Downloads | Docs | Contact | Publications | Account | Logout |

Please select the type of model you would like to submit:

- ACT-R
- Command-line
- Java
- Julia
- Matlab
- Python
- R

Submit job

- specify starting, step, end values for each simulation parameter
- genetic algorithm option requires target values
- upload zip file
- specify dependencies



lr	ugs	bin1	bin2	bin3	bin4	MindModelingRunTime
0.03	0.01	0.54742	0.58348	0.58384	0.5871	6.14
0.01	0.01	0.53686	0.59424	0.6121	0.61864	6.14
0.02	0.01	0.5487	0.5905	0.5973	0.5926	6.11
0.38	0.14	0.54944	0.6113	0.6379	0.65602	2.359375
0.38	0.15	0.55156	0.61186	0.63924	0.65022	2.296875
0.38	0.16	0.55408	0.61132	0.64212	0.65108	2.34375
0.38	0.17	0.5479	0.6136	0.63782	0.6527	2.328125
0.38	0.18	0.55114	0.60798	0.63504	0.65086	2.3125
0.38	0.19	0.55186	0.61244	0.63604	0.6499	2.3125
0.38	0.2	0.54516	0.61248	0.63724	0.65016	2.28125
0.38	0.21	0.54788	0.60616	0.63434	0.6452	2.296875
0.38	0.22	0.54738	0.60854	0.63512	0.64796	2.25
0.38	0.23	0.5457	0.6111	0.6318	0.64524	2.234375
0.38	0.24	0.5504	0.61014	0.63134	0.6461	2.21875
0.38	0.25	0.5435	0.60652	0.62926	0.6416	2.3125

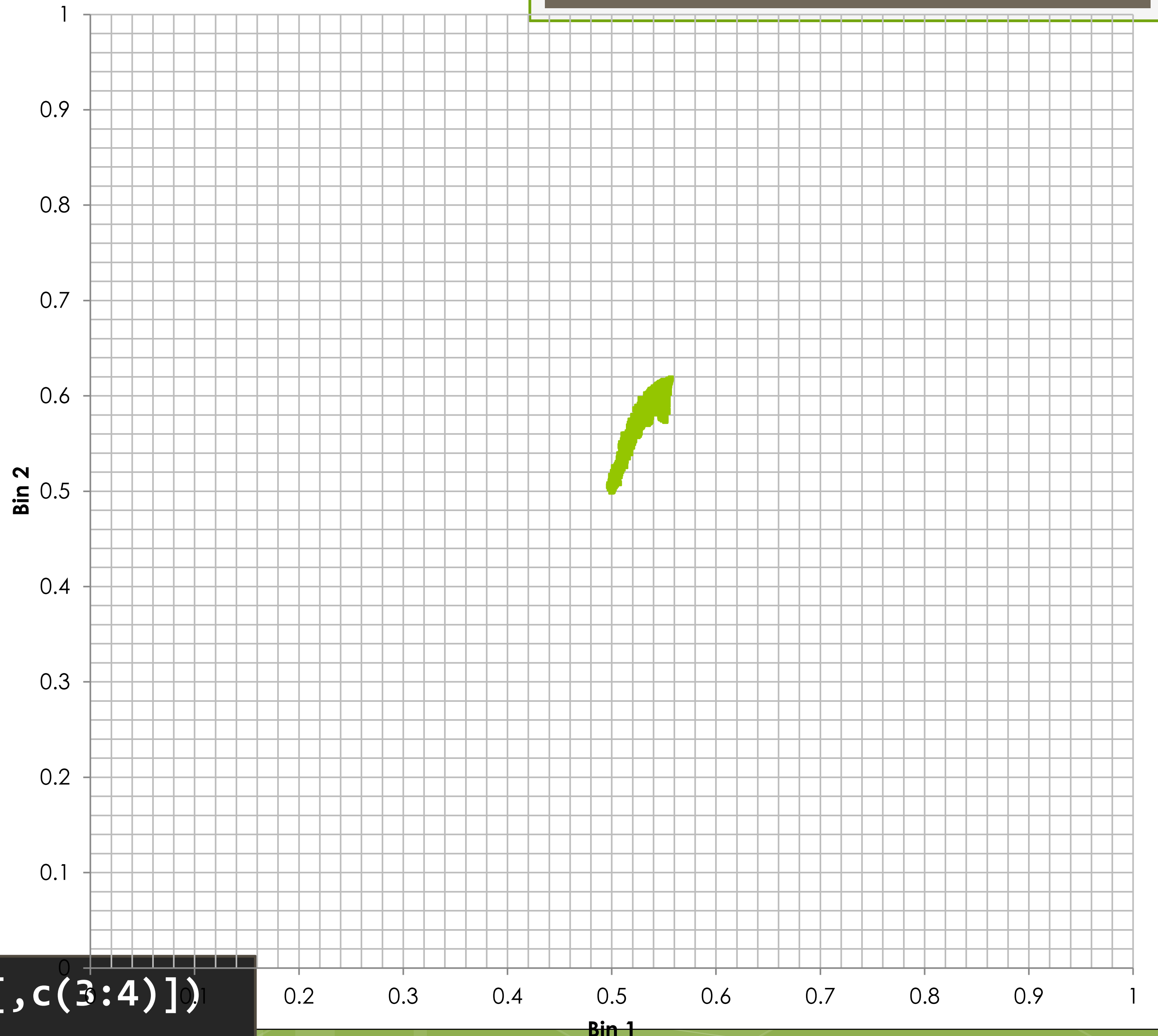
```
> mfa(data[,c(3:6)])
[1] 0.0016
```

lr	ugs	bin1	bin2	bin3	bin4	MindModelingRunTime
0.03	0.01	0.54742	0.58348	0.58384	0.5871	6.14
0.01	0.01	0.53686	0.59424	0.6121	0.61864	6.14
0.02	0.01	0.5487	0.5905	0.5973	0.5926	6.11
0.38	0.14	0.54944	0.6113	0.6379	0.65602	2.359375
0.38	0.15	0.55156	0.61186	0.63924	0.65022	2.296875
0.38	0.16	0.55408	0.61132	0.64212	0.65108	2.34375
0.38	0.17	0.5479	0.6136	0.63782	0.6527	2.328125
0.38	0.18	0.55114	0.60798	0.63504	0.65086	2.3125
0.38	0.19	0.55186	0.61244	0.63604	0.6499	2.3125
0.38	0.2	0.54516	0.61248	0.63724	0.65016	2.28125
0.38	0.21	0.54788	0.60616	0.63434	0.6452	2.296875
0.38	0.22	0.54738	0.60854	0.63512	0.64796	2.25
0.38	0.23	0.5457	0.6111	0.6318	0.64524	2.234375
0.38	0.24	0.5504	0.61014	0.63134	0.6461	2.21875
0.38	0.25	0.5435	0.60652	0.62926	0.6416	2.3125

```
> mfa(data[,c(3:5)])
[1] 0.0028
```

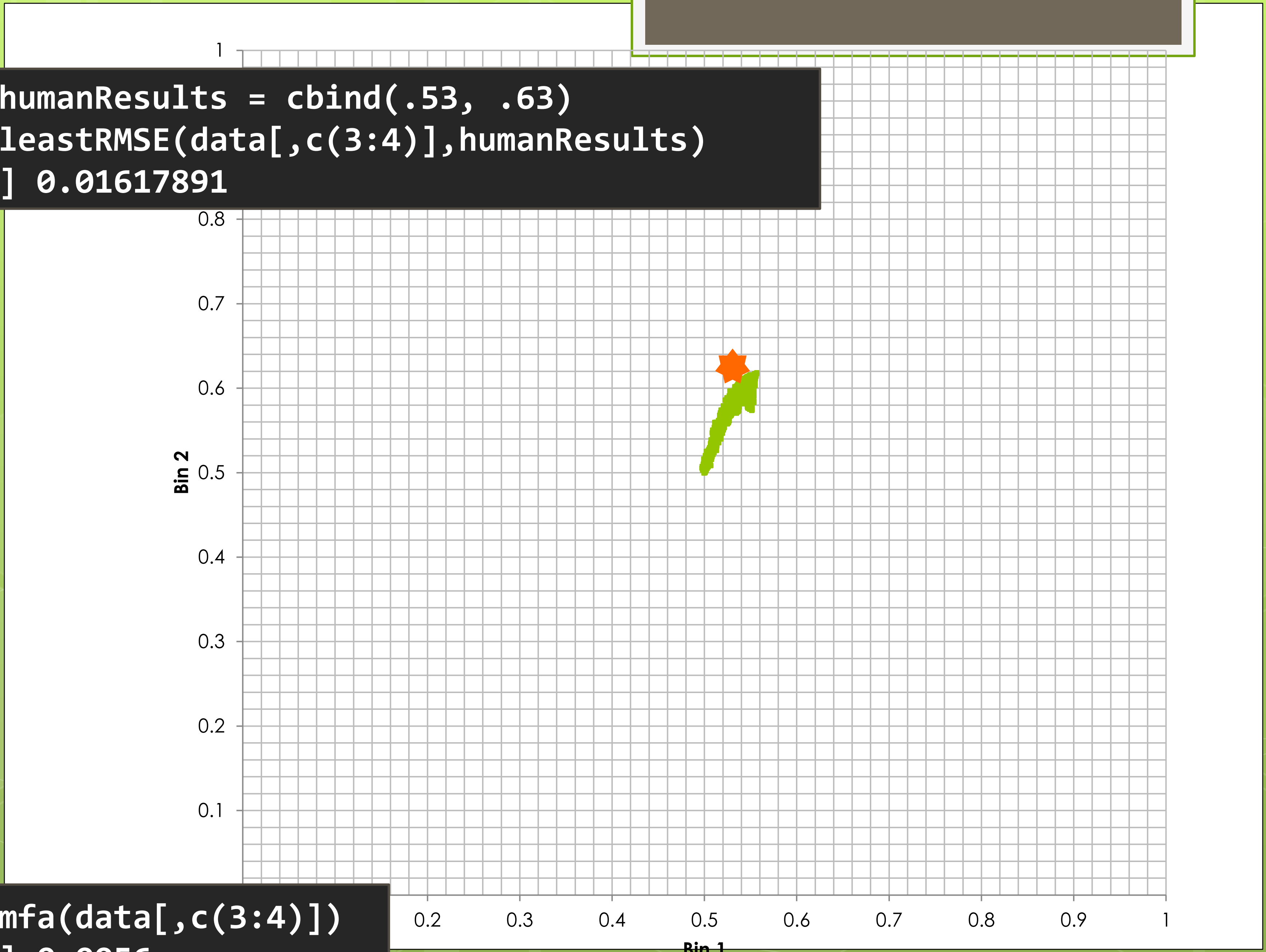
lr	ugs	bin1	bin2	bin3	bin4	MindModelingRunTime
0.03	0.01	0.54742	0.58348	0.58384	0.5871	6.14
0.01	0.01	0.53686	0.59424	0.6121	0.61864	6.14
0.02	0.01	0.5487	0.5905	0.5973	0.5926	6.11
0.38	0.14	0.54944	0.6113	0.6379	0.65602	2.359375
0.38	0.15	0.55156	0.61186	0.63924	0.65022	2.296875
0.38	0.16	0.55408	0.61132	0.64212	0.65108	2.34375
0.38	0.17	0.5479	0.6136	0.63782	0.6527	2.328125
0.38	0.18	0.55114	0.60798	0.63504	0.65086	2.3125
0.38	0.19	0.55186	0.61244	0.63604	0.6499	2.3125
0.38	0.2	0.54516	0.61248	0.63724	0.65016	2.28125
0.38	0.21	0.54788	0.60616	0.63434	0.6452	2.296875
0.38	0.22	0.54738	0.60854	0.63512	0.64796	2.25
0.38	0.23	0.5457	0.6111	0.6318	0.64524	2.234375
0.38	0.24	0.5504	0.61014	0.63134	0.6461	2.21875
0.38	0.25	0.5435	0.60652	0.62926	0.6416	2.3125

```
> mfa(data[,c(3:4)])
[1] 0.0056
```

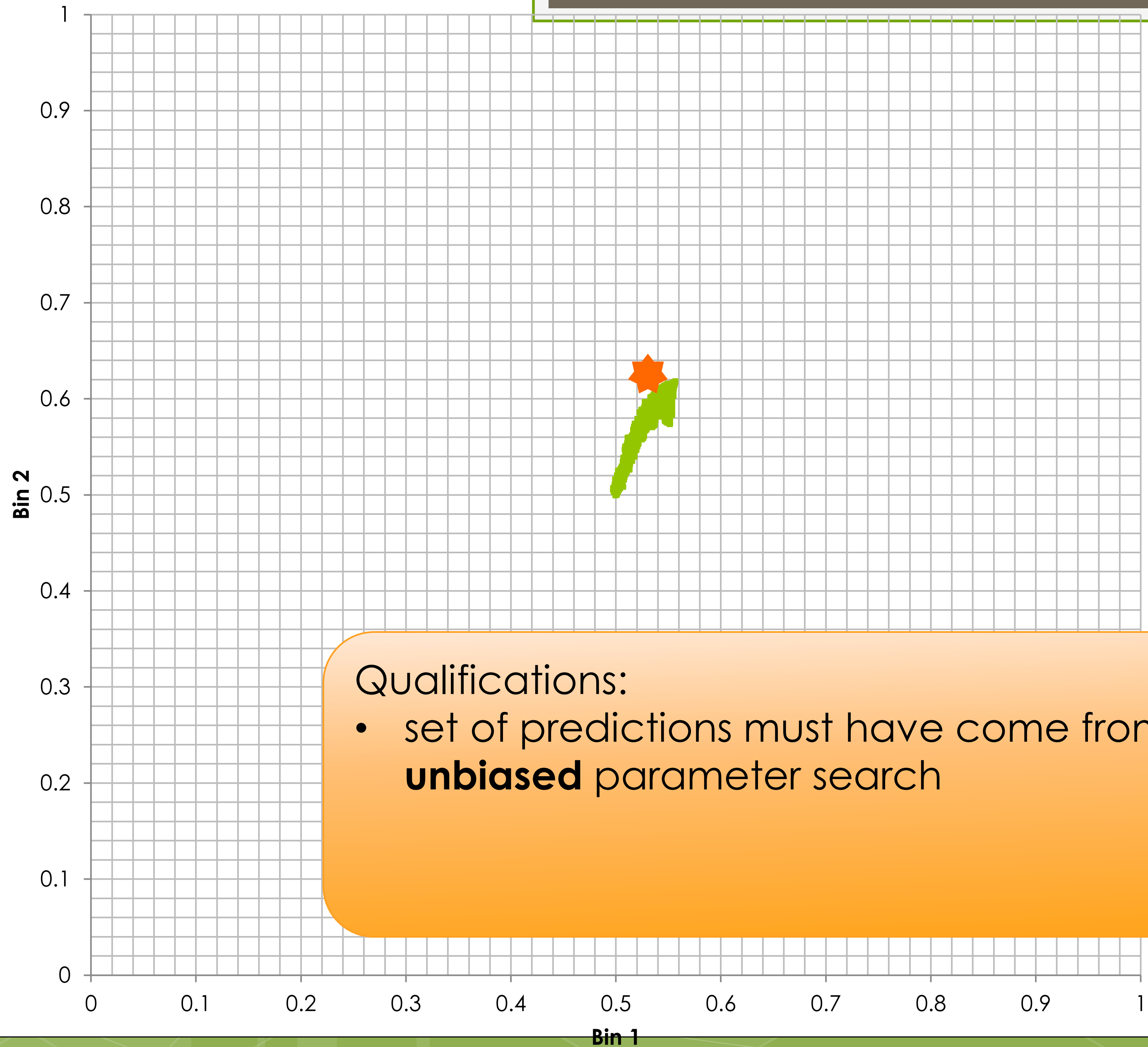



```
> mfa(data[,c(3:4)])
[1] 0.0056
```

```
> humanResults = cbind(.53, .63)
> leastRMSE(data[,c(3:4)],humanResults)
[1] 0.01617891
```

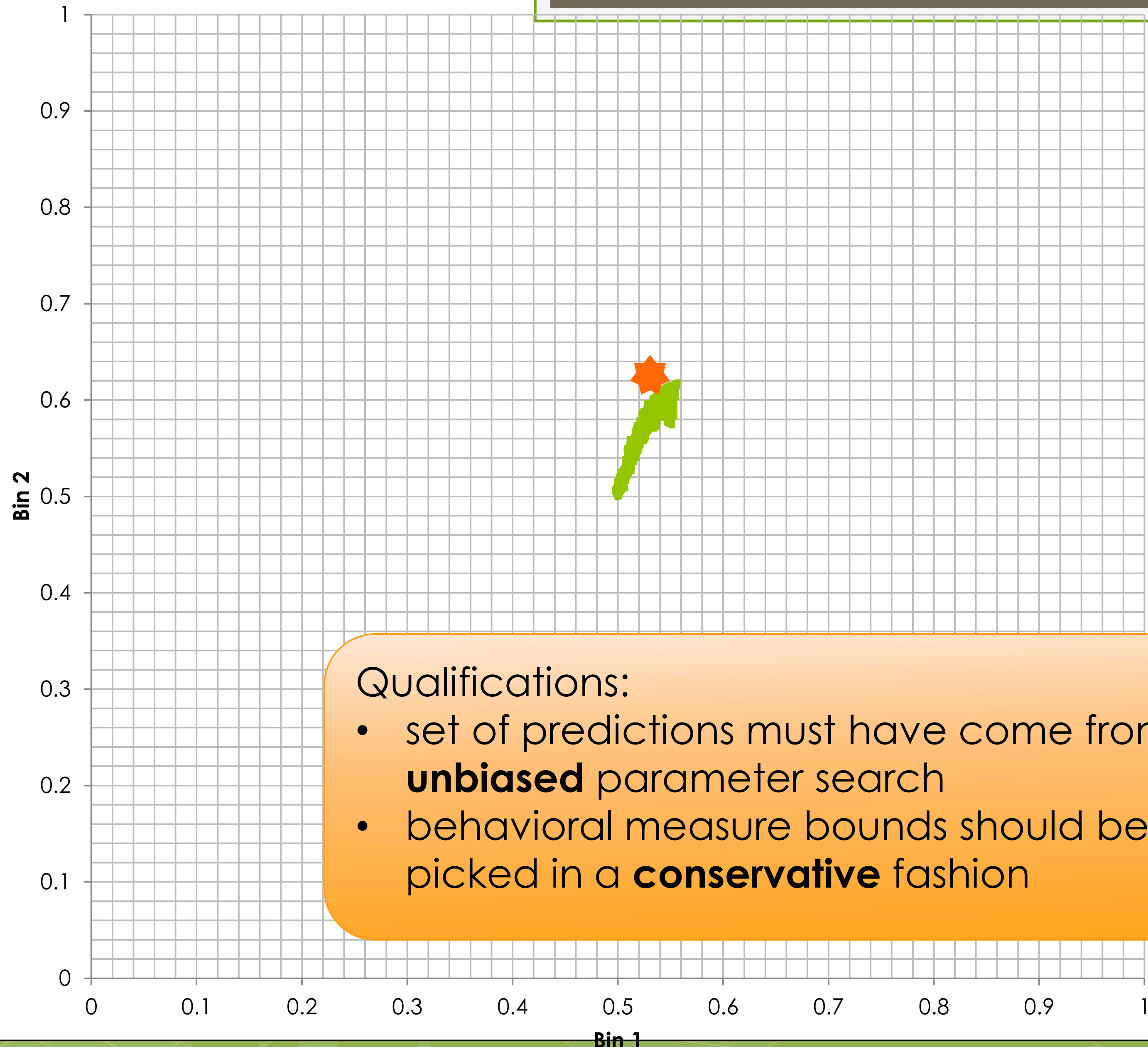



```
> mfa(data[,c(3:4)])
[1] 0.0056
```



Qualifications:

- set of predictions must have come from an **unbiased** parameter search



Qualifications:

- set of predictions must have come from an **unbiased** parameter search
- behavioral measure bounds should be picked in a **conservative** fashion

Summary

- Status quo
  - you find a good fit of model to data
  - you report best-found fit
  - you discount the fit by the number of model parameters
- Additionally we should
  - provide context for interpreting the impressiveness of the fit, e.g.
  - "model predicts .56% of the potential behavioral outcomes"

Flexibility Assessment	Appropriate for All Model Types	Appropriate for Quantitative Analyses	Interpretable Flexibility Metric for each Model
Number of free parameters (e.g., AIC, BIC, RMSEA, adjusted r^2)	No[1]	Yes	Yes
Parameter Space Partitioning (PSP; Pitt et al. 2006)	Yes	No[2]	Yes
Model Mimicry (Navarro et al. 2004; Wagenmakers et al. 2004)	Yes	Yes	No[3]
Normalized Maximum Likelihood (NML; Grünwald, 2005, 2007)	No[4]	Yes	Yes[5]
Bayesian Model Selection (BMS; e.g., Myung & Pitt, 1997)	No[4]	Yes	No[3]
Prior Predictive Evaluation (Vanpaemel, 2009)	No[6]	Yes	Yes
Model Flexibility Analysis	Yes	Yes	Yes

1. appropriate for linear mathematical models, where with n free quantitative parameters one can fit any n-dimensional dataset
2. appropriate for qualitative analyses of relationships between behavioral measures
3. provides rank-order of model flexibility
4. appropriate for probability models where goodness of fit is measured via a log-likelihood
5. provides non-strict p-values, which may be difficult to interpret in some cases
6. appropriate where model predictions are independent across behavioral measures