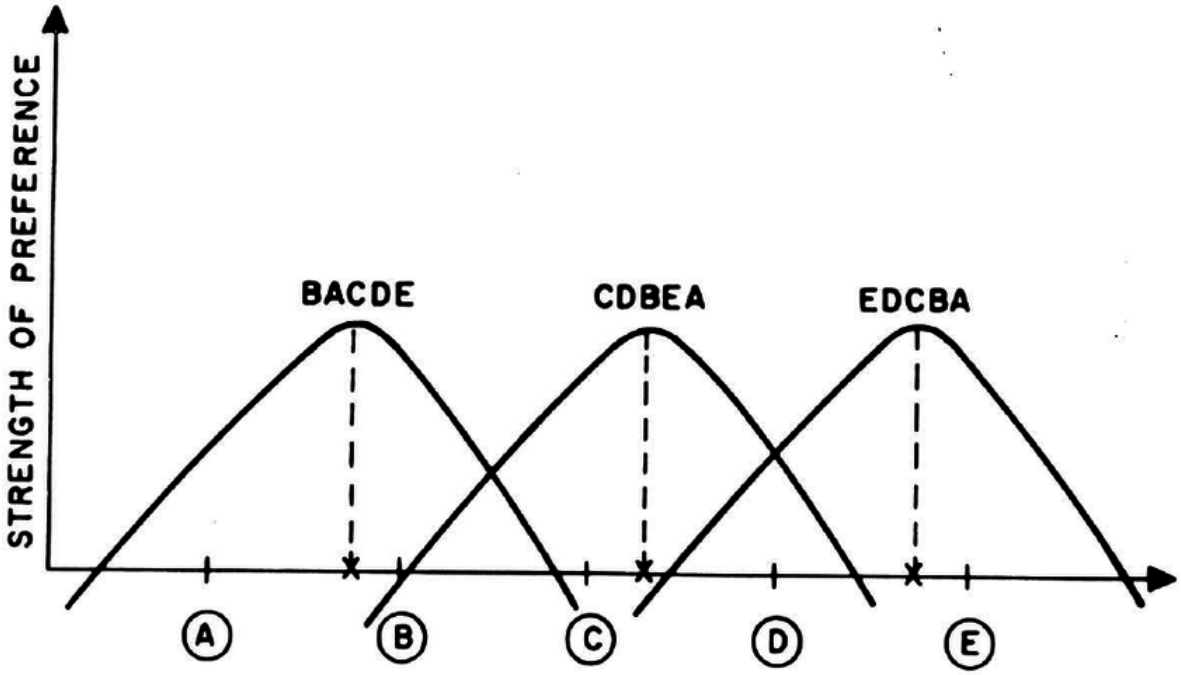
(Unidimensional) Unfolding

Analyzing Survey Research Data
GSERM, Summer 2019
Prof. Adam M. Enders

What is (Unidimensional) Unfolding?

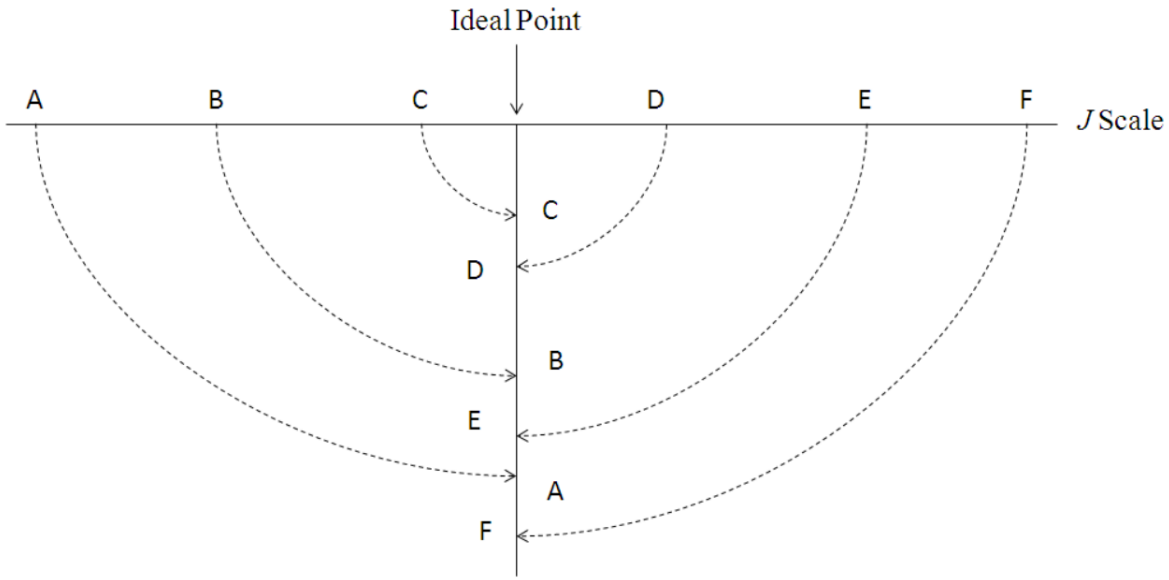
- Scaling model that assumes "single-peaked" item response functions, rather than monotonic ones, like with the cumulative scaling model
- Represents **proximities** between the n rows and k columns of a **rectangular** data matrix as distances between points along a single continuum
- Objectives:
  - Represent row and column objects along a single latent continuum
  - Proximity between row and each column object should represent, to the best extent possible, the preferences/(dis)similarities from the original data matrix

Subject "Utility Functions"



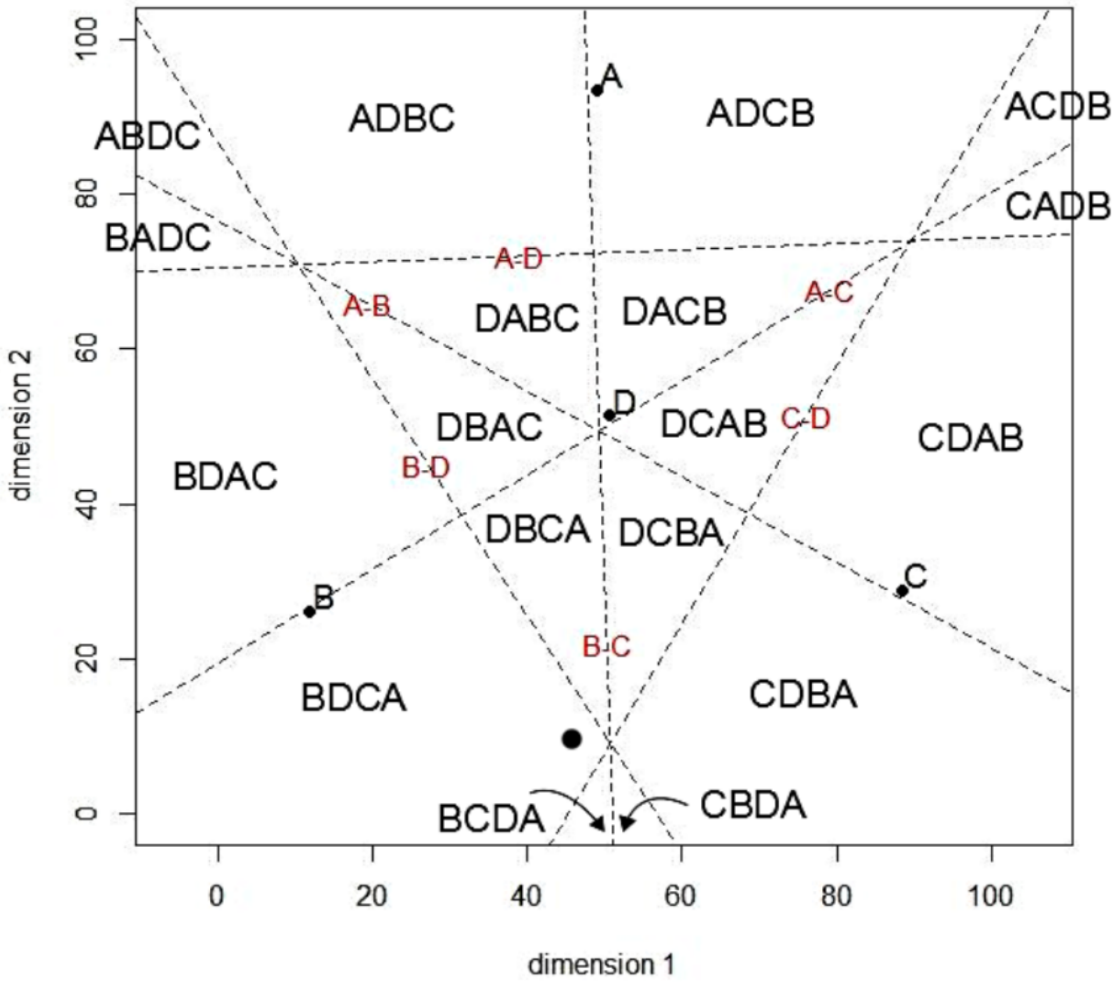
Preferences Can Be "Unfolded"

Just imagine a piece of string!



Generalization to Multiple Dimensions

Now imagine a napkin!



The Quintessential Example

- Spatial theories of voting, like the one popularized by Downs (1957)
  - The model itself (the unfolding model), devoid of substantive content, was first proposed Hotelling (1929) and completely developed by Coombs (1950)
- Commonality between most spatial theories of voting:
  1. Each voter can be represented by a point in some hypothetical space such that the point reflects the person's ideal set of policies
  2. The policy position of each candidate can be represented by a point in the same space
  3. A voter chooses the candidate whose policy position is closest to his or her own
- Note that "spatial" more or less implies "proximity," hence why proximity data is most appropriate for the unfolding model

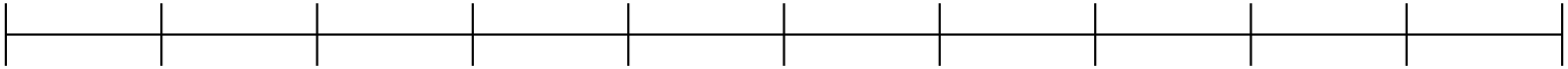
Basic Issue Space

Basic Issue Space

Abortion

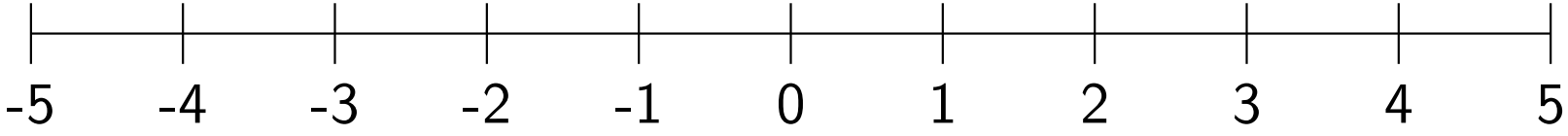
Basic Issue Space

Abortion

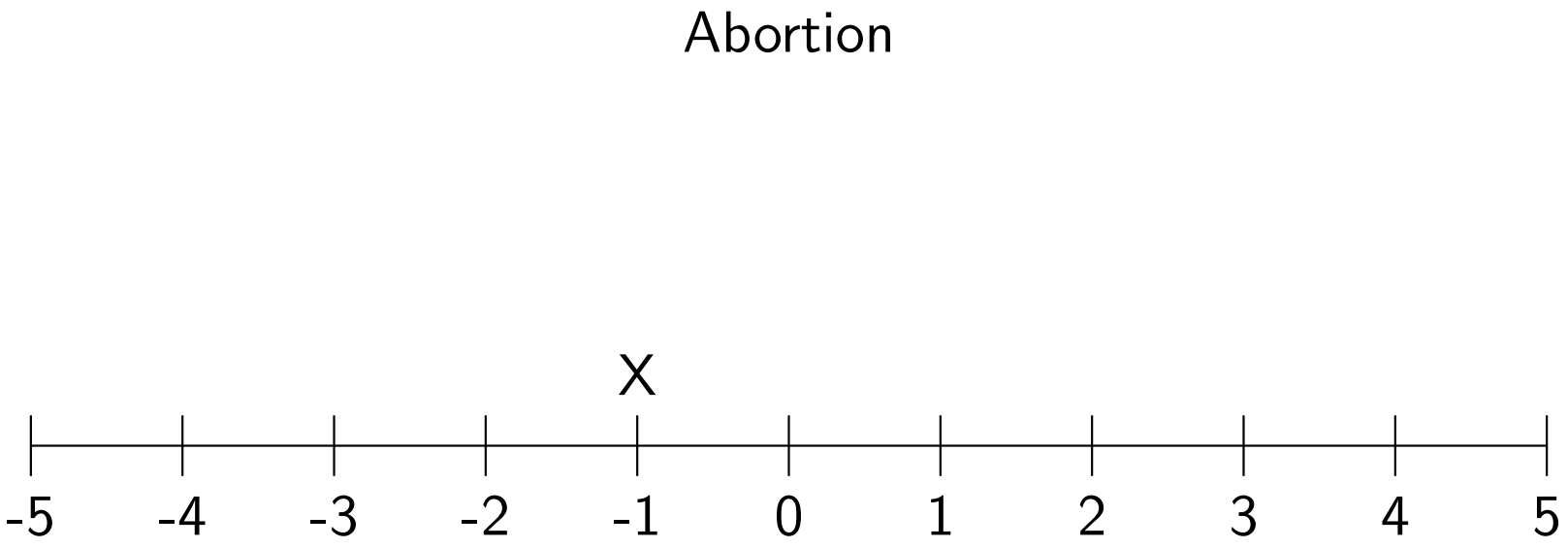


Basic Issue Space

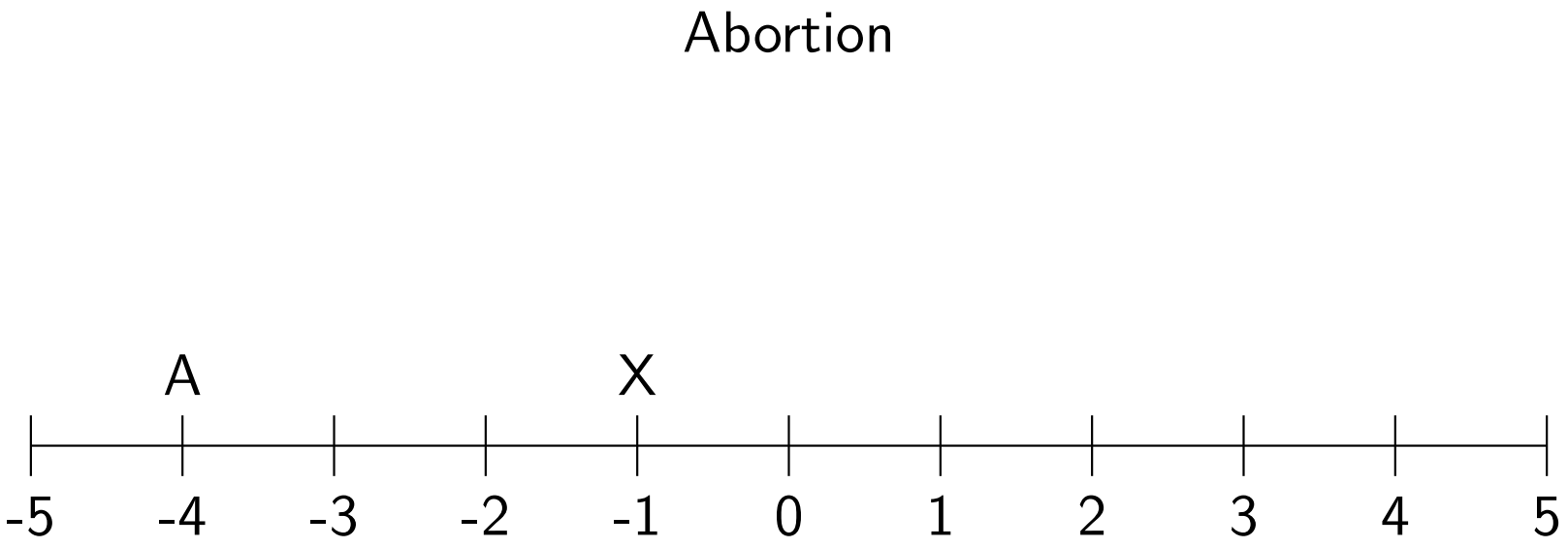
Abortion



Basic Issue Space

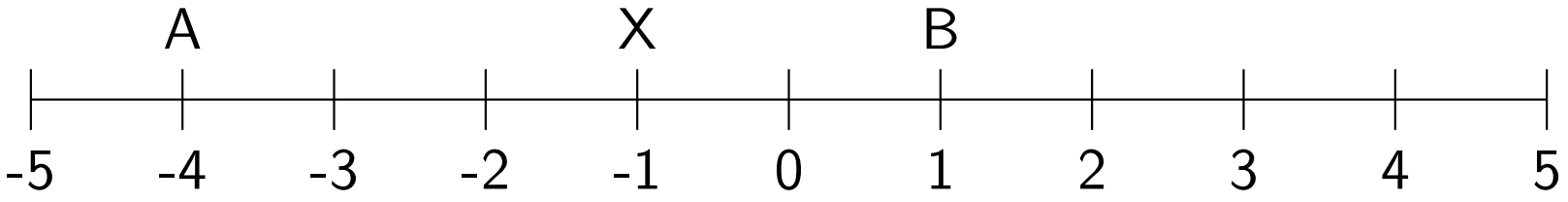


Basic Issue Space

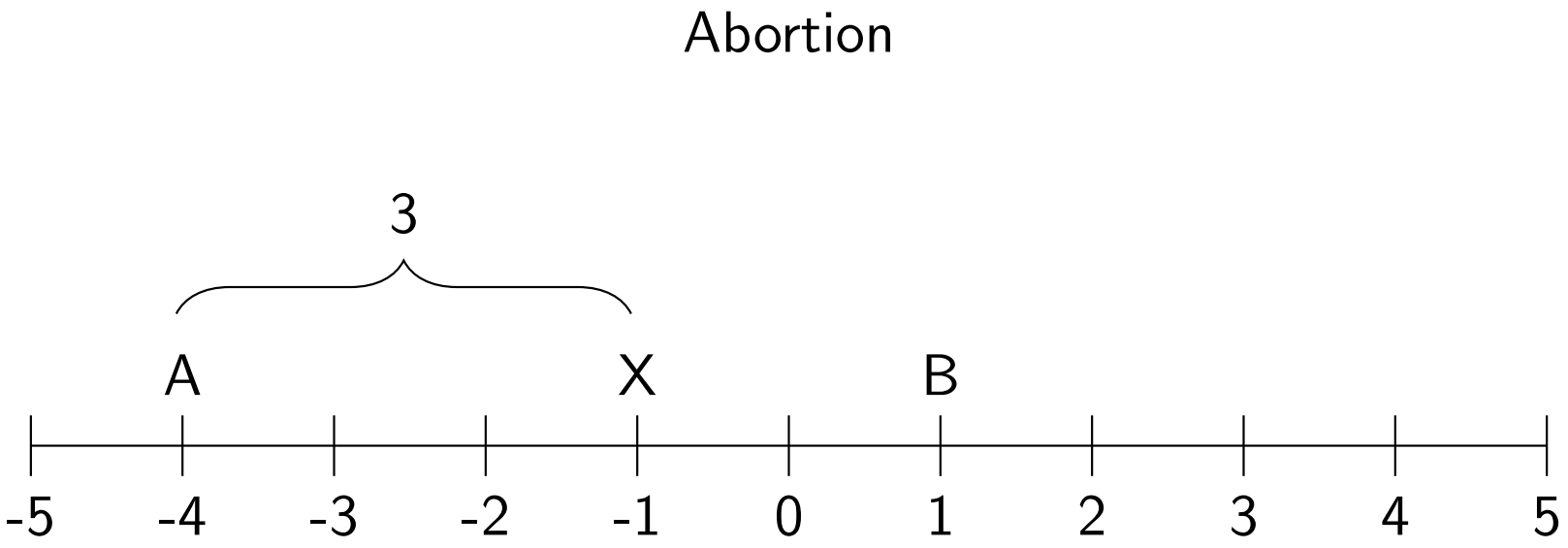


Basic Issue Space

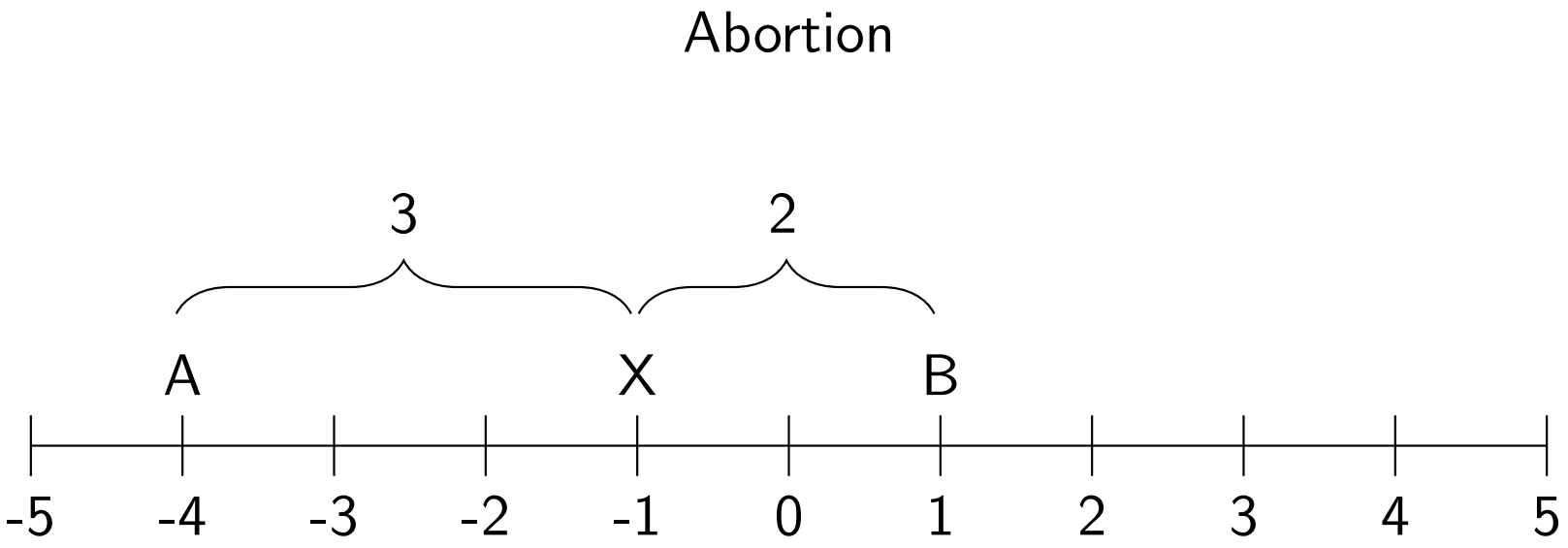
Abortion



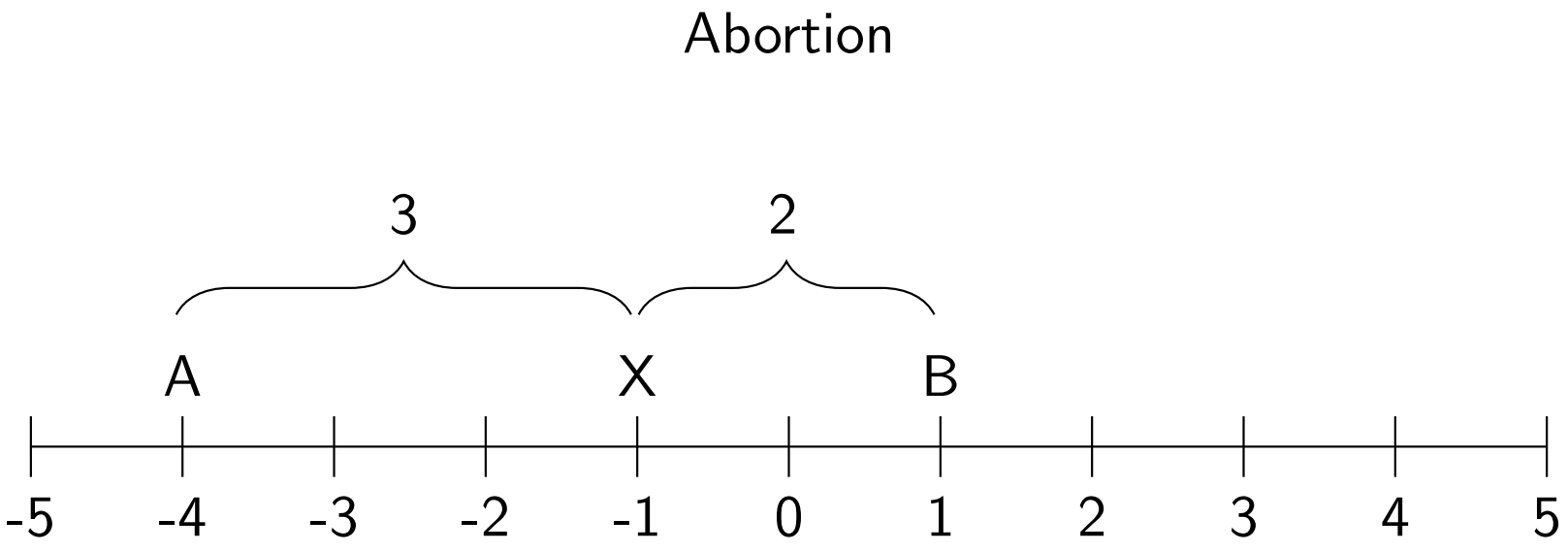
Proximity Model



Proximity Model



Proximity Model



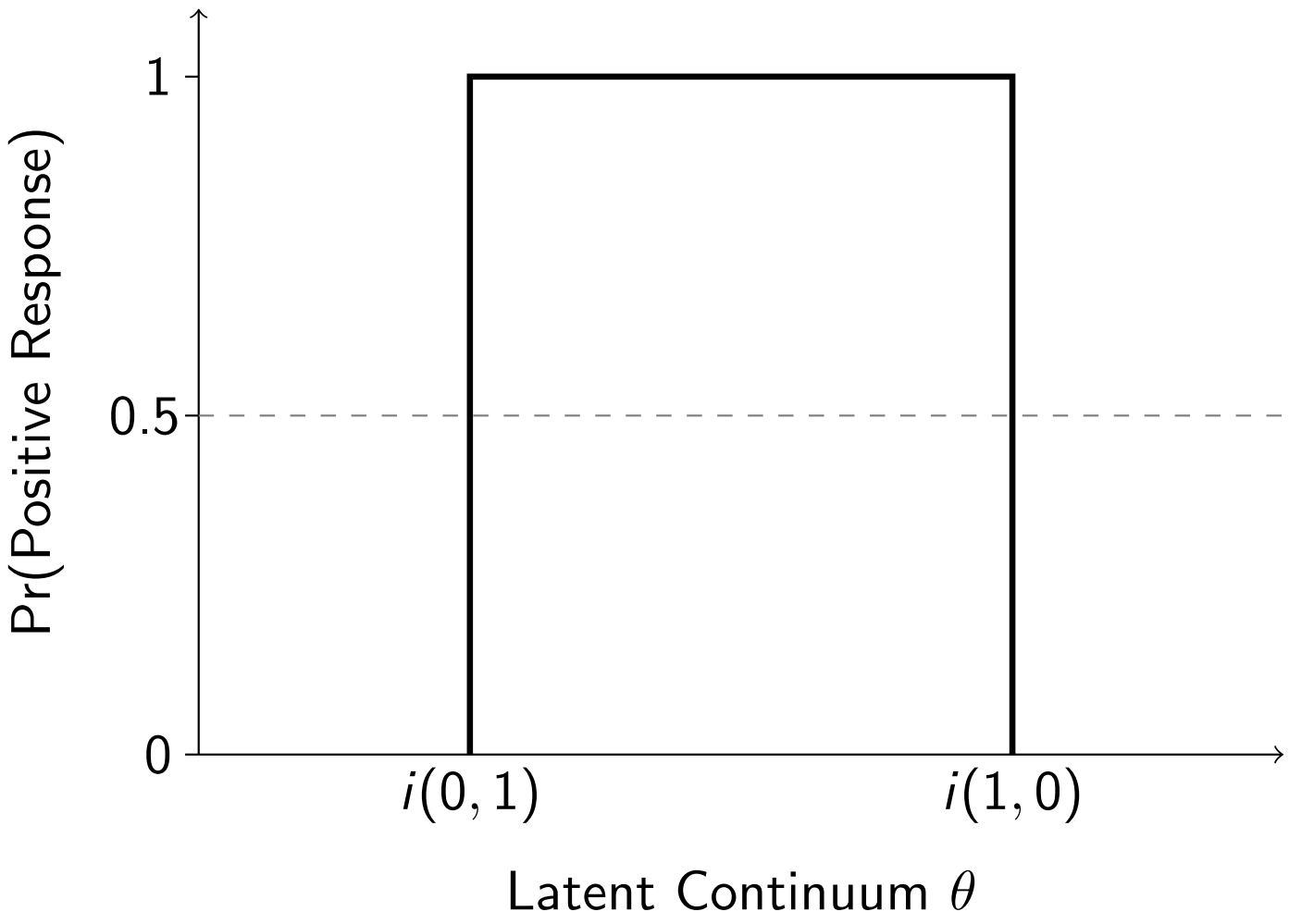
Since $2 < 3$, X votes for candidate B

The Deterministic Model

- Subjects only respond positively to items that represented close to their position on the latent continuum
- They respond negatively to items that are far away from that position
- Since they respond positively to all items close to their own position, they will respond positively only to items that are adjacent to each other
  - In other words, they have single-peaked preferences

The Deterministic Model

A "step function" with two steps: the left-sided step, in which the probability of a positive response increases from 0 to 1, and the right-sided step, in which the probability decreases from 1 to 0



Error-Free Response Pattern

Matrix of (rearranged) data that forms perfect (deterministic) unfolding scale:

Column Objects							
A	B	C	D	E	F	G	H
1	1	1	1	1	0	0	0
0	1	1	1	1	1	0	0
0	0	1	1	1	1	1	0
0	0	0	1	1	1	1	1

Unfolding originally called "parallelogram analysis" by Coombs (1964)

Assigning Scale Scores

1. First, column objects are given a rank score in the order in which they form an unfolding scale, using only the odd numbers
   - The rank score is assigned by rearranging rows and columns to construct (the closest approximation of) a parallelogram of 1's
2. Second, as their scale value, row objects are assigned the median value of the column scores of the column objects to which they responded positively

Error-Free Response Pattern

Matrix of (rearranged) data that forms perfect (deterministic) unfolding scale, with row and column scale scores:

	Column Objects								
Subjects	A	B	C	D	E	F	G	H	Row Score
1	1	1	1	1	1	0	0	0	5
2	0	1	1	1	1	1	0	0	7
3	0	0	1	1	1	1	1	0	9
4	0	0	0	1	1	1	1	1	11
Column Score	1	3	5	7	9	11	13	15	

A More Realistic Example

A close approximation of a parallelogram, but with obvious errors

Column Objects							
A	B	C	D	E	F	G	H
1	0	1	1	1	0	0	0
0	1	1	1	0	1	0	0
0	0	1	1	0	0	1	0
0	0	0	1	0	1	1	1

Problems: 1) how do we quantify error, and 2) how do we assign scale scores if we are comfortable that observed error is sufficiently negligible?

Assessing and Dealing with Error

- Like with the cumulative scaling model, we can use Loevinger's H coefficient
  - We compare the number of observed errors to the number of errors expected under statistical independence (i.e., the case where the data do not form unfolding scale)
  - $H = 1 - \frac{E(obs)}{E(exp)}$
  - Here, the observed errors are according to the unfolding model, rather than the cumulative scaling model
  - H still bound between 0 and 1, where 1 is perfect model fit
- This could help assess fit in the deterministic model, but couldn't help with dealing with erroneous response patterns
  - How could you assign scale values to response patterns with errors?
  - Most of the time, practitioners would simply drop rows with erroneous response patterns as long as a majority of the data was left

The Nonparametric Model

- Like with cumulative scaling, we're next going to consider a nonparametric version of the unfolding model
- Several reasons for doing so:
  - Like with Mokken Scaling, the nonparametric approach takes assumptions, and checking of those assumptions, very seriously
  - Assumes only nonmonotonic single-peaked preferences, rather than IRFs of a particular shape (e.g., Gaussian, quadratic, step function)
  - If you understand the process with the nonparametric formulation, you can easily understand the parametric one
- Important caveat:
  - Not many software packages can reliably estimate a A) unidimensional B) parametric unfolding model
  - "smacof" in R is unreliable...PROC MDS in SAS and ALSCAL in SPSS are best
  - Even though unidimensional parametric models are difficult, multidimensional unfolding is not

The Nonparametric Model

1. We will consider the ordinal nonparametric unidimensional unfolding model developed by van Schuur (1984)
   - Frequently referred to as the MUDFOLD model
2. Very recently (December 2017) written into an R package called "mudfold"
   - Does not *yet* have ability to deal with non-dichotomous data (though it does include a function for dichotomizing data if that seems worthwhile/appropriate)
   - However, this means that not many people have had a chance to implement it in published research yet
   - So, lots of opportunities to do something totally unique in your field

Constructing a Nonparametric Ordinal Unfolding Scale

1. Begin with automated item selection procedure based on item and scale H coefficients
   - Assess which column objects form a strong (usually $H \geq 0.30$) unfolding scale and proceed with only those column objects beginning with the best three column objects (if present)
   - Note that while the unfolding model is not inherently more restrictive than the cumulative scaling model, generally speaking fewer datasets fit this model
2. Assign scale values to row/column objects
3. Further check assumptions of the model that would allow one to order the column objects along the latent dimension, as well as row objects:
   - Construct and examine empirical patterns in the:
     - Adjacency Matrix
     - Conditional Adjacency Matrix
     - Dominance Matrix
   - If some column objects appear "problematic," consider removal from the scale

Automated Item Selection Procedure

1. Need to find the set of three items – a "triple" – that forms a scale with an H coefficient of 0.30 or above
   - Consider three column objects, A, B, and C, that are unfoldable in that order
   - Expected errors: $p(A) \times [1 - p(B)] \times p(C) \times n$
   - $p(A)$, $p(B)$, and $p(C)$ are the relative frequencies with which the positive response to these objects was given, and n is the sample size of the dataset
   - Observed error: frequency with which objects A and C get the positive response but object B gets the negative response from the same subject
   - Need best "unique" triple:
     - A triple with a positive $H(ijk)$ in one permutation, but negative $H(jik)$ and $H(ikj)$ values for the other two permutations
     - Highest proportion of 111, 110, and 011 response patterns

Automated Item Selection Procedure

2. Next, add the item which forms the quadruple that maximizes the scale H coefficient
   - If no additional item produces a scale with an H coefficient of at least 0.30, terminate procedure
3. Repeat step 2, moving on to the quintuple, and so on, until the H value for the scale plus the additional item is less than 0.30
   - Terminate item selection procedure and proceed with analysis

Assign Scale Values

- For imperfect response patterns – those in which the 0 (negative) response occurs between two or more 1 responses – a decision must be made about whether the 0 response means 0 or 2 item steps have been passed
- Decision is based on the relative number of 1 responses to the left and right of the 0 response in an ordered response pattern
  - If the majority of 1 responses is to the left of the 0 response, we assume that item step (01) has not been passed
  - If the majority of 1 responses is to the right of the 0 response, we assume that both item steps (01) and (10) have been passed
  - If the number of 1 responses to the left and the right of the 0 response is the same, we assume that one item step has been passed
  - Subjects who do not respond positively to any items cannot be assigned a scale score

Assign Scale Values & Error

TABLE 2. Scale Values for Perfect and Imperfect Response Patterns in Terms of the Number of Item Steps Passed and the Number of Errors in Terms of the Number of 101 Triples in the Response Pattern

	Item					Number of Item Steps Passed						
Subject	A	B	C	D	E	A	B	C	D	E	Scale Value	Number of Errors
1	1	1	0	0	0	1	1	0	0	0	2	0
2	0	1	1	1	0	2	1	1	1	0	5	0
3	0	0	0	0	1	2	2	2	2	1	9	0
4	1	0	1	1	1	1	2	1	1	1	6	3
5	1	1	1	0	1	1	1	1	0	1	4	3
6	1	1	0	1	1	1	1	1	1	1	5	4
7	1	0	0	1	1	1	2	2	1	1	7	4
8	1	0	1	0	1	1	2	1	0	1	5	4

Error = any triple displaying 101 pattern

(Stealing) An Example

- van Schuur (1992) uses attitudes about five politicians in 1980:
  - Carter (A), Kennedy (B), John B. Anderson (C), Reagan (D), H. W. Bush (E)
- Original data values recoded so that 1=(very) favorable and 0=neutral/unfavorable
- Scale order appears above from A-E
- Instead of assuming ideological dimension, we use the favorability items to estimate it

A First Look

TABLE 3. Percentages of Positive Scores to Each Unfoldable Item by Subjects with Different Scale Values

Scale Value	N	Kennedy		Carter		Anderson		Bush		Reagan	
		0	1	0	1	0	1	0	1	0	1
1–2	31	10	90	0	100	100	0	100	0	100	0
3	36	36	64	22	78	31	69	94	6	97	3
4	30	67	33	30	70	30	70	67	33	100	0
5–6	20	65	35	40	60	35	65	30	70	60	40
7	23	100	0	78	22	26	74	13	87	4	96
8–9	190	100	0	100	0	100	0	1	99	0	100
N	330	244	86	251	79	254	76	95	235	109	221

Looking from left to right for each scale value group, the expected single peakedness of the probabilities of the 1 responses can be observed

Order: Carter, Kennedy, Anderson, Reagan, Bush

Model Checking: Dominance Matrix

- Dominance matrix constructed as follows:
  - Square matrix in which rows and columns are made up of the items in their unfoldable order
  - Cell ij contains the percentage of subjects who gave the positive response to the row item i and the negative response to the column item j
- Under the deterministic unfolding model, this percentage should be low for adjacent items, but should increase with the distance between items
- The dominance matrix should, then, show for each row that the percentages decrease from the leftmost column to the diagonal, and increase from the diagonal to the rightmost column

Example: Dominance Matrix

TABLE 4. Dominance Matrix for Five Unfolding Items

	B	A	C	E	D
(B) KENNEDY	—	11	13	22	23
(A) CARTER	13	—	15	18	23
(C) ANDERSON	12	12	—	15	17
(E) BUSH	69	63	63	—	8
(D) REAGAN	66	64	61	3	—

In each row the percentages decrease from the leftmost column to the diagonal, and increase from the diagonal to the rightmost column

Model Checking: Adjacency Matrix

- Adjacency matrix constructed as follows:
  - Square matrix in which rows and columns are made up of the items in their unfoldable order
  - Cell ij contains the percentage of subjects who give the positive response to both items i and j
- Under the deterministic unfolding model, this percentage should be "high(er)" for adjacent items, but should decrease with increasing distance between the items
- The matrix should show that the percentages decrease from the diagonal both to the left on the same row, and down on the same column

Example: Adjacency Matrix

TABLE 5. Adjacency Matrix for Five Unfolding Items

	B	A	C	E	D
(B) KENNEDY	—				
(A) CARTER	13	—			
(C) ANDERSON	10	10	—		
(E) BUSH	2	7	8	—	
(D) REAGAN	1	3	6	63	—

Percentages decrease from the diagonal both to the left on the same row, and down on the same column

Model Checking: Conditional Adjacency Matrix

- Conditional adjacency matrix constructed as follows:
  - Square matrix in which rows and columns contain the items in their unfoldable order
  - The value of each cell (ij) – the number of subjects who give the positive response to both items i and j – is specified relative to the number of subjects who give the positive response to the row item (item i)
- Maximum cell value is found in cells that "move" from top left to bottom right in the matrix

Example: Conditional Adjacency Matrix

TABLE 6. Conditional Adjacency Matrix for Five Unfolding Items

	B	A	C	D	E
(B) KENNEDY	—	0.51	0.47	0.03	0.02
	—	(.054)	(.057)	(.011)	(.009)
(A) CARTER	0.56	—	0.46	0.11	0.05
	(.056)	—	(.057)	(.020)	(.015)
(C) ANDERSON	0.41	0.46	—	0.11	0.09
	(.053)	(.056)	—	(.021)	(.019)
(E) BUSH	0.09	0.30	0.36	—	0.95
	(.032)	(.050)	(.055)	—	(.015)
(D) REAGAN	0.05	0.13	0.26	0.89	—
	(.026)	(.036)	(.051)	(.020)	—

Maximum cell values found in cells adjacent to the diagonal in each row

Model Checking: Correlation Matrix

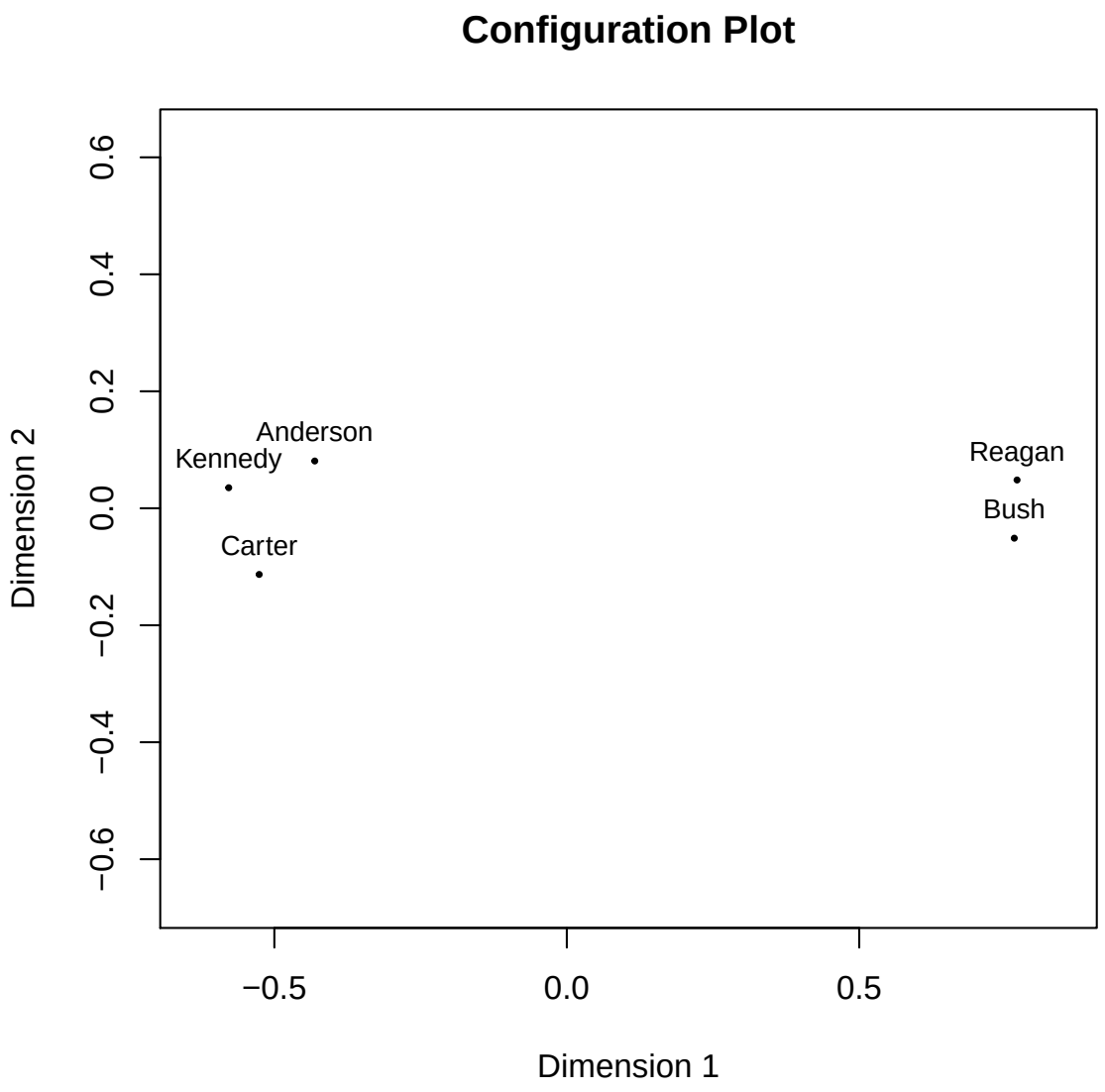
- Correlation matrix constructed as follows:
  - Regular old product-moment correlations between variables, *reordered according to estimated item ordering*
- Correlation matrix of the items should exhibit two sign changes at most: from negative to positive to negative
- Should form a "simplex-like" pattern

Example: Correlation Matrix

TABLE 8. Correlation Matrix for Five Unfolding Items

	B	A	C	E	D
(B) KENNEDY	1.00	0.38	0.30	−0.77	−0.74
(A) CARTER	0.38	1.00	0.25	−0.54	−0.68
(C) ANDERSON	0.30	0.25	1.00	−0.43	−0.47
(E) BUSH	−0.77	−0.54	−0.43	1.00	0.75
(D) REAGAN	−0.74	−0.68	−0.47	0.75	1.00

Example: Correlations as Distances Between Points



Summary

- Unfolding model fits the politician attitude data very well
  - No violations of model assumptions checks (various matrices)
- Can use distribution of individuals along the latent dimension as a DV or IV in a multivariate statistical model
  - In this example, scale scores could be used to test the spatial theory of voting by regressing vote choice on scale scores
  - Validation of the unfolding scale might also come in the form of correlations with party identification, ideology, and issue attitudes (basically anything that might be indicative of general left-right political orientations)

Example: Attitudes Toward Capitol Punishment

- 54 students were asked whether they agreed (1) or disagreed (0) with the following 8 statements:
  - `HIDEOUS`: "Capital punishment is one of the most hideous practices of our time"
  - `LIFESACRED`: "The state cannot teach the sacredness of human life by destroying it"
  - `INEFFECTIV`: "Capital punishment is not an effective deterrent to crime"
  - `DONTBELIEV`: "I do not believe in capital punishment but i am not sure it is not necessary"
  - `WISHNOTNEC`: "I think capital punishment is necessary but i wish it were not"
  - `MUSTHAVEIT`: "Until we find a more civilized way to prevent crime we must have capital punishment"
  - `DETERRENT`: "Capital punishment is justified because it does act as a deterrent to crime"
  - `CRIMDESERV`: "Capital punishment gives the criminal what he deserves"

Scalability Coefficient H

```
Individuals: 54

Items: 8

Item set: HIDEOUS LIFESACRED INEFFECTIV DONTBELIEV WISHNOTNEC MUSTHAVEIT
DETERRENT CRIMDESERV
Best unique triple: INEFFECTIV DONTBELIEV DETERRENT
Iterations in the second step: 5
Unidimensional Mudfold scale: LIFESACRED INEFFECTIV HIDEOUS DONTBELIEV
WISHNOTNEC MUSTHAVEIT CRIMDESERV DETERRENT
Total number of observed errors for the Mudfold scale: 131
Total number of expected errors for the Mudfold scale: 362.5686
scalability H for the Mudfold scale: 0.6386891
Iso statistic for the Mudfold scale: 0.4267544
```

Item Information and Statistics

	index	samp.size	freq	prop	std.err	Obs.err	Exp.err	Iso	Scalab.H
1	LIFESACRED	54	35	0.65	0.09	60	143.42	0.23	0.58
2	INEFFECTIV	54	36	0.67	0.09	55	146.67	0.00	0.63
3	HIDEOUS	54	24	0.44	0.06	64	137.87	0.00	0.54
4	DONTBELIEV	54	25	0.46	0.06	53	137.16	0.01	0.61
5	WISHNOTNEC	54	26	0.48	0.07	39	136.71	0.03	0.71
6	MUSTHAVEIT	54	24	0.44	0.06	38	135.60	0.01	0.72
7	CRIMDESERV	54	19	0.35	0.05	35	125.14	0.00	0.72
8	DETERRENT	54	19	0.35	0.05	49	125.14	0.13	0.61
9	total	54	NA	NA	NA	131	362.57	0.43	0.64

Dominance Matrix

	LIFESACRED	INEFFECTIV	HIDEOUS	DONTBELIEV	WISHNOTNEC	MUSTHAVEIT	CRIMDESERV	DETERRENT
LIFESACRED	NA	0.074	0.278	0.333	0.481	0.519	0.611	0.537
INEFFECTIV	0.093	NA	0.278	0.315	0.481	0.519	0.593	0.611
HIDEOUS	0.074	0.056	NA	0.222	0.333	0.352	0.389	0.389
DONTBELIEV	0.148	0.111	0.241	NA	0.241	0.278	0.315	0.315
WISHNOTNEC	0.315	0.296	0.370	0.259	NA	0.056	0.167	0.185
MUSTHAVEIT	0.315	0.296	0.352	0.259	0.019	NA	0.130	0.167
CRIMDESERV	0.315	0.278	0.296	0.204	0.037	0.037	NA	0.093
DETERRENT	0.241	0.296	0.296	0.204	0.056	0.074	0.093	NA

In each row the percentages *should* decrease from the leftmost column to the diagonal, and increase from the diagonal to the rightmost column in a perfect scale

We observe a couple violations – should move on to next test before taking action

Adjacency Matrix

	LIFESACRED	INEFFECTIV	HIDEOUS	DONTBELIEV	WISHNOTNEC	MUSTHAVEIT	CRIMDESERV	DETERRENT
LIFESACRED	NA	NA	NA	NA	NA	NA	NA	NA
INEFFECTIV	0.574	NA	NA	NA	NA	NA	NA	NA
HIDEOUS	0.370	0.389	NA	NA	NA	NA	NA	NA
DONTBELIEV	0.315	0.352	0.222	NA	NA	NA	NA	NA
WISHNOTNEC	0.167	0.185	0.111	0.222	NA	NA	NA	NA
MUSTHAVEIT	0.130	0.148	0.093	0.185	0.426	NA	NA	NA
CRIMDESERV	0.037	0.074	0.056	0.148	0.315	0.315	NA	NA
DETERRENT	0.111	0.056	0.056	0.148	0.296	0.278	0.259	NA

Percentages *should* decrease from the diagonal both to the left on the same row, and down on the same column

Again, we observe a couple of seemingly minor violations

Conditional Adjacency Matrix

	LIFESACRED	INEFFECTIV	HIDEOUS	DONTBELIEV	WISHNOTNEC	MUSTHAVEIT	CRIMDESERV	DETERRENT
LIFESACRED	NA	0.861	0.833	0.68	0.346	0.292	0.105	0.316
INEFFECTIV	0.886	NA	0.875	0.76	0.385	0.333	0.211	0.158
HIDEOUS	0.571	0.583	NA	0.48	0.231	0.208	0.158	0.158
DONTBELIEV	0.486	0.528	0.500	NA	0.462	0.417	0.421	0.421
WISHNOTNEC	0.257	0.278	0.250	0.48	NA	0.958	0.895	0.842
MUSTHAVEIT	0.200	0.222	0.208	0.40	0.885	NA	0.895	0.789
CRIMDESERV	0.057	0.111	0.125	0.32	0.654	0.708	NA	0.737
DETERRENT	0.171	0.083	0.125	0.32	0.615	0.625	0.737	NA

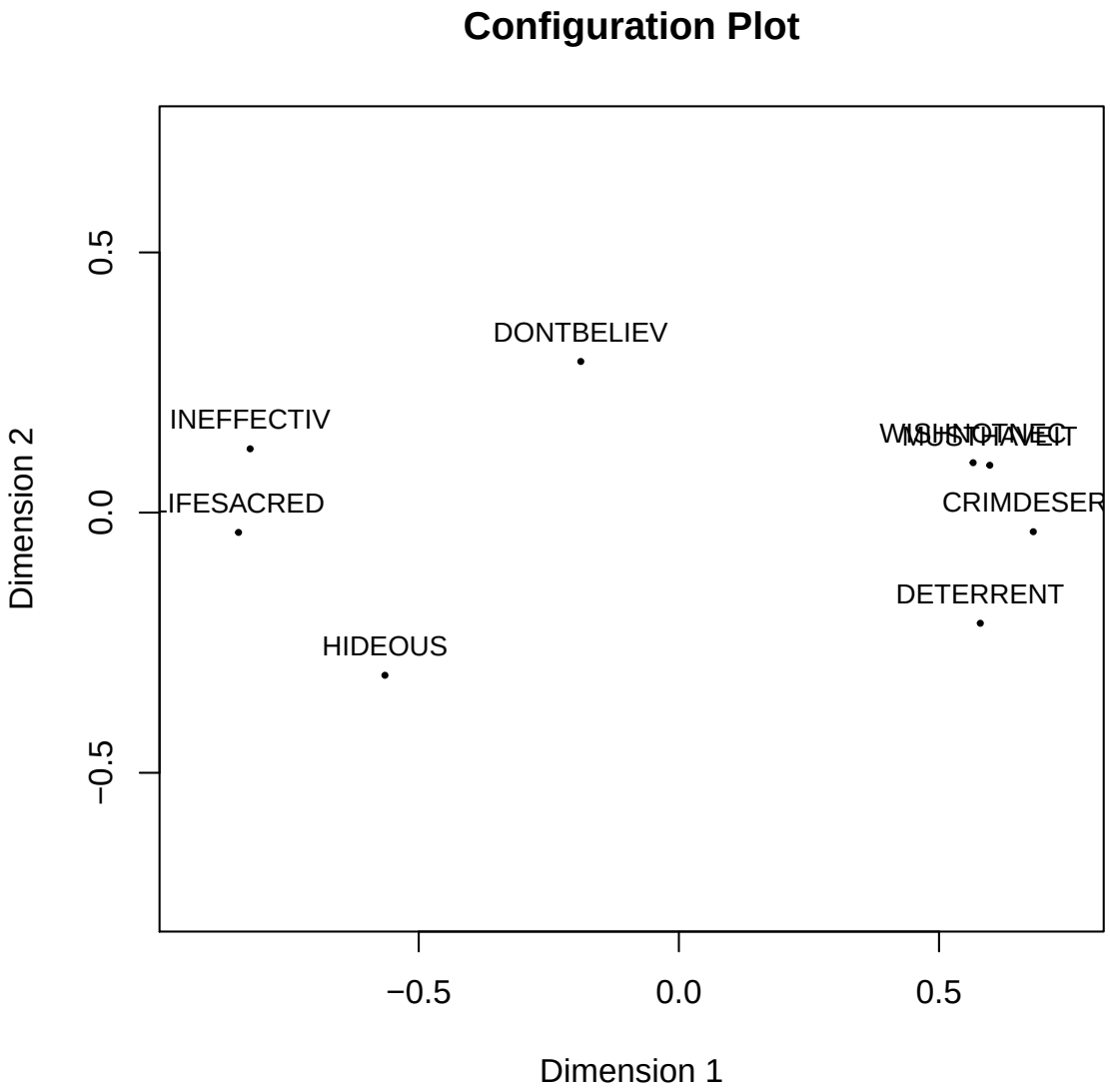
Maximum cell values *should be* found in cells adjacent to the diagonal in each row

Only one violation (DONTBELIEV, HIDEOUS)

Correlation Matrix

	LIFESACRED	INEFFECTIV	HIDEOUS	DONTBELIEV	WISHNOTNEC	MUSTHAVEIT	CRIMDESERV	DETERRENT
LIFESACRED	1.000	0.631	0.347	0.062	-0.609	-0.668	-0.838	-0.513
INEFFECTIV	0.631	1.000	0.395	0.184	-0.577	-0.632	-0.713	-0.795
HIDEOUS	0.347	0.395	1.000	0.066	-0.414	-0.425	-0.425	-0.425
DONTBELIEV	0.062	0.184	0.066	1.000	-0.003	-0.083	-0.062	-0.062
WISHNOTNEC	-0.609	-0.577	-0.414	-0.003	1.000	0.854	0.609	0.532
MUSTHAVEIT	-0.668	-0.632	-0.425	-0.083	0.854	1.000	0.668	0.512
CRIMDESERV	-0.838	-0.713	-0.425	-0.062	0.609	0.668	1.000	0.594
DETERRENT	-0.513	-0.795	-0.425	-0.062	0.532	0.512	0.594	1.000

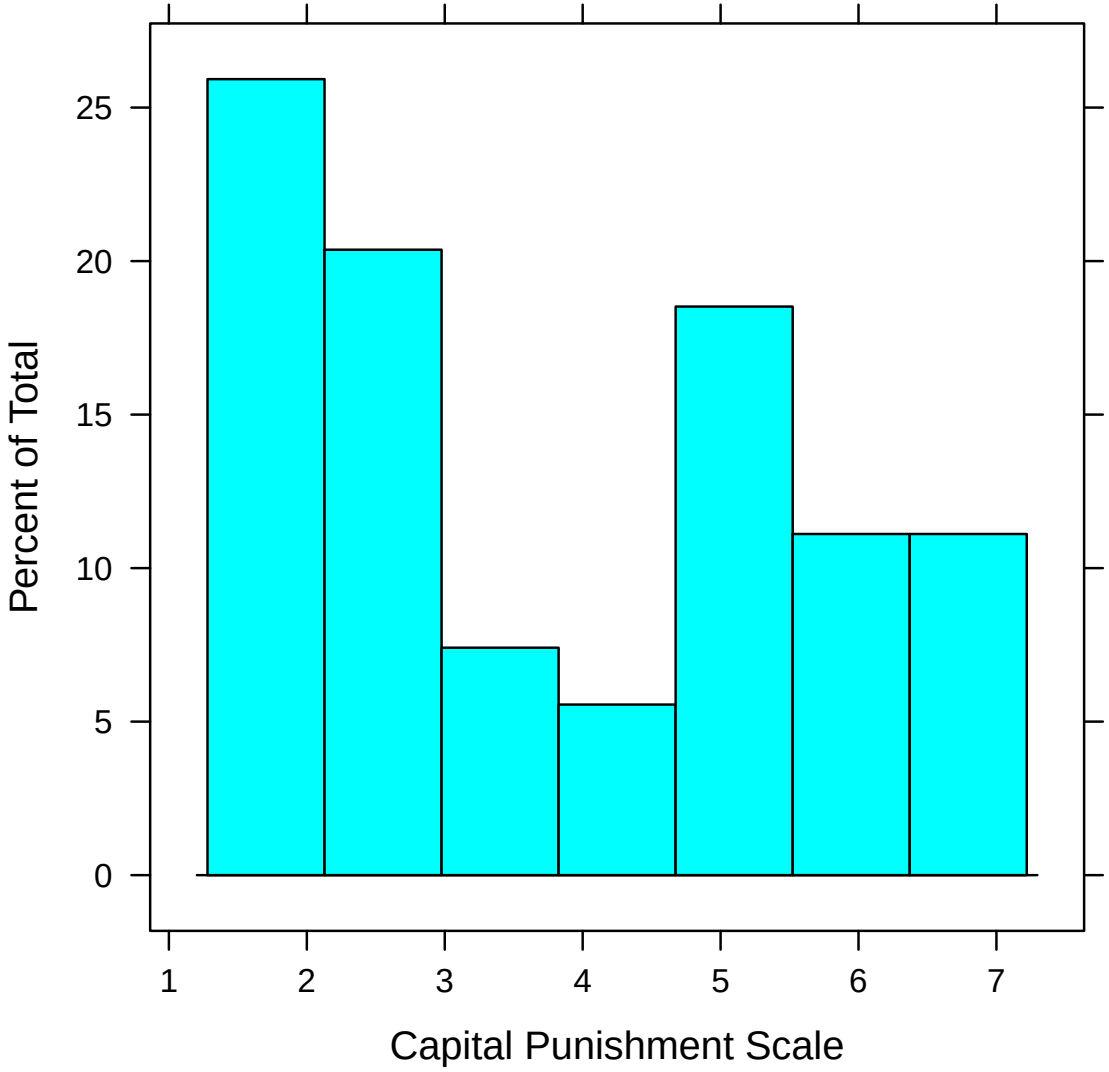
Modeling Correlations as Distances



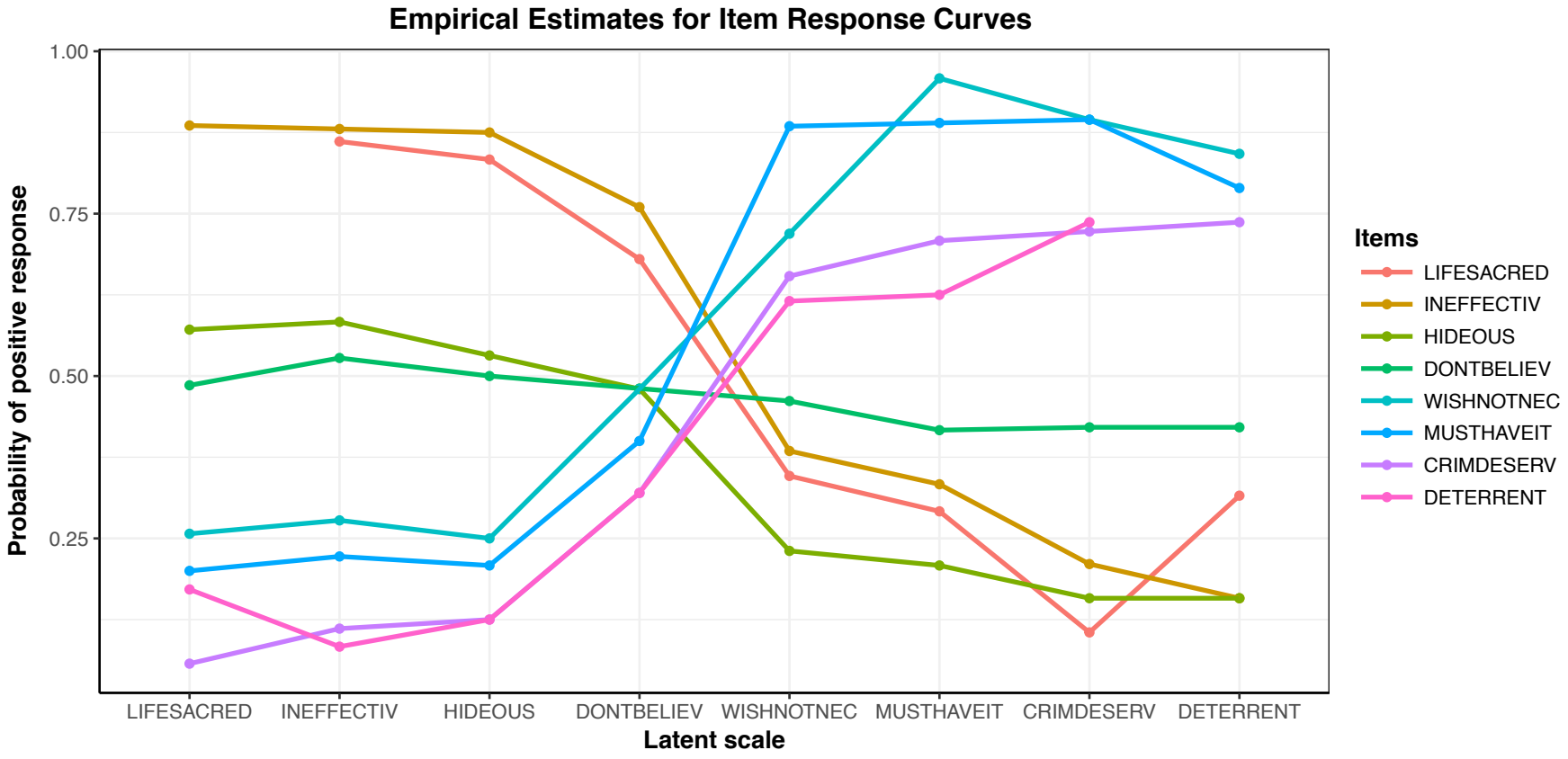
Example: Attitudes Toward Capitol Punishment

- Items in order along latent dimension:
  - `LIFESACRED`: "The state cannot teach the sacredness of human life by destroying it"
  - `INEFFECTIV`: "Capital punishment is not an effective deterrent to crime"
  - `HIDEOUS`: "Capital punishment is one of the most hideous practices of our time"
  - `DONTBELIEV`: "I do not believe in capital punishment but i am not sure it is not necessary"
  - `WISHNOTNEC`: "I think capital punishment is necessary but i wish it were not"
  - `MUSTHAVEIT`: "Until we find a more civilized way to prevent crime we must have capital punishment"
  - `CRIMDESERV`: "Capital punishment gives the criminal what he deserves"
  - `DETERRENT`: "Capital punishment is justified because it does act as a deterrent to crime"

Distribution of Scale Scores

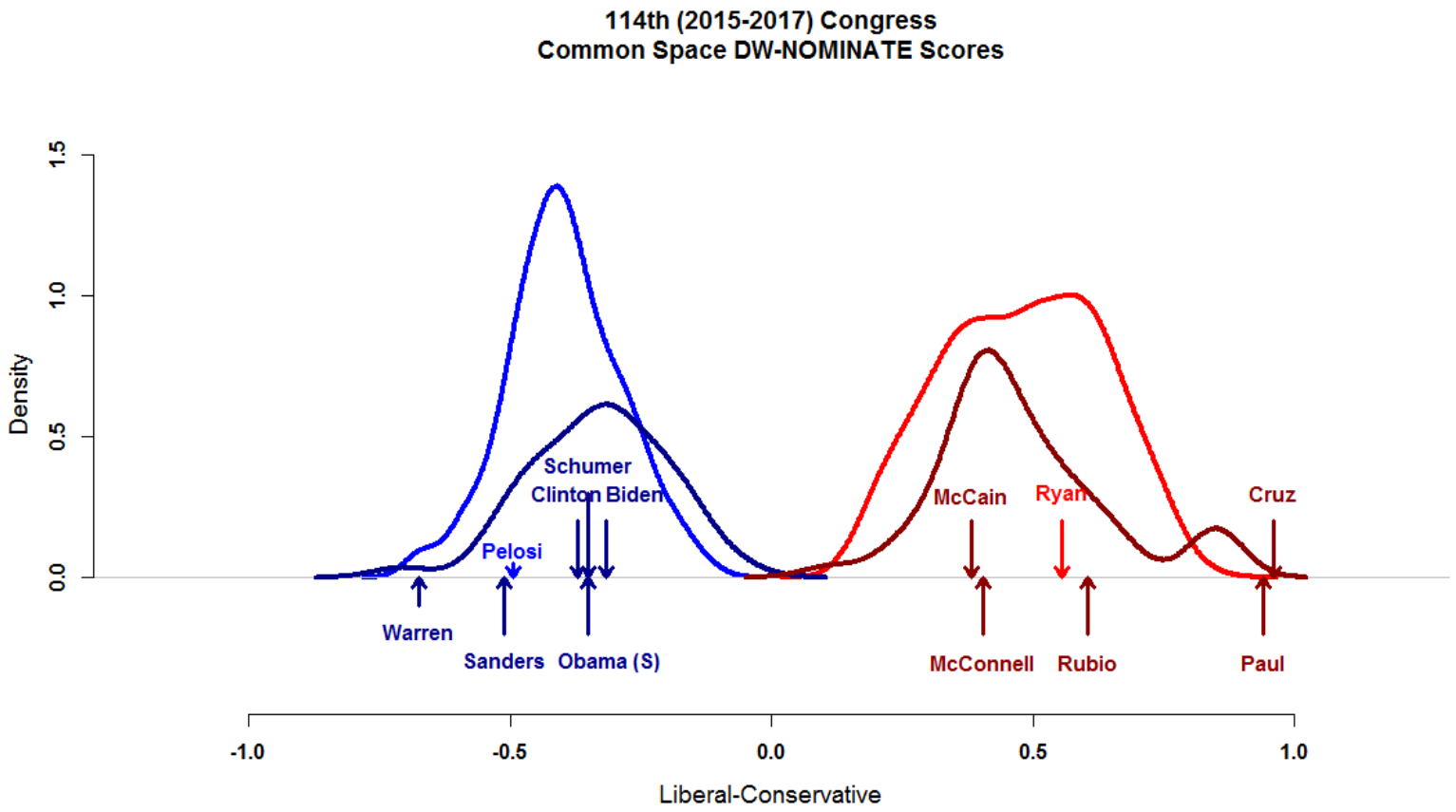


Estimated Item Response Functions



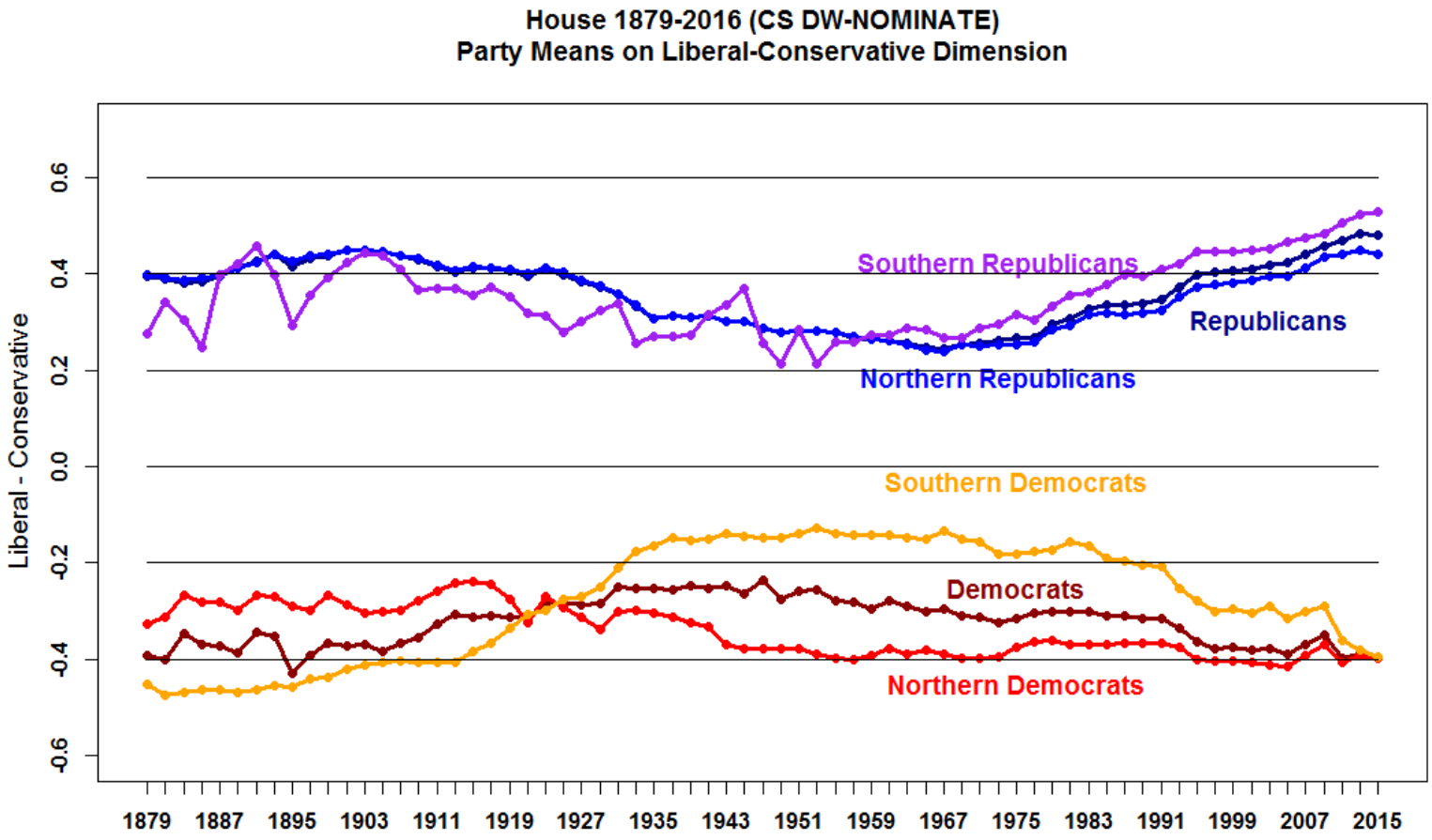
Example: Congressional Ideology

Distribution of ideal points along first dimension in 2015-2017 House of Representatives:



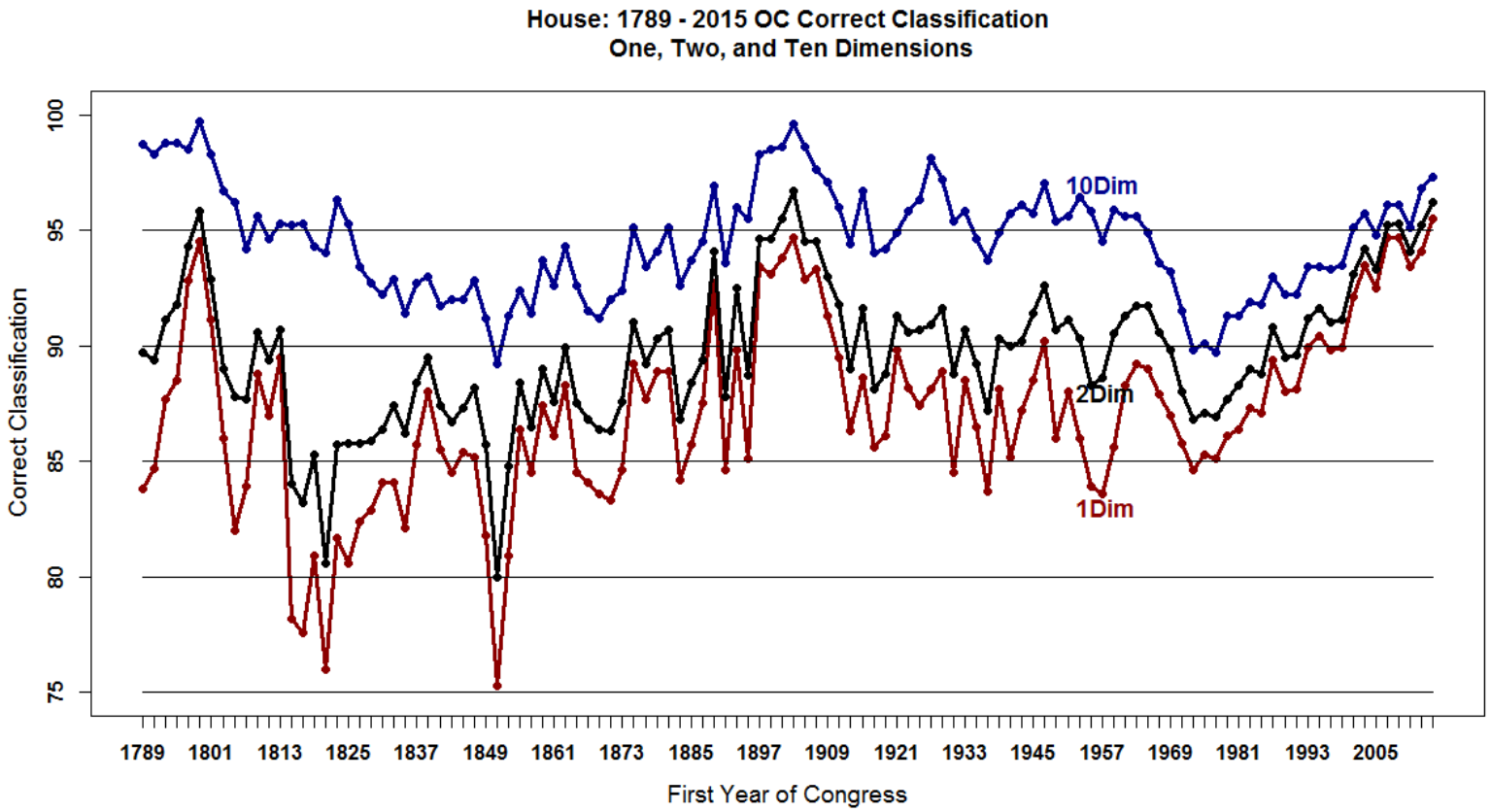
Example: Congressional Ideology

Average ideal point of (sub-)party along single dimension over time



Example: Congressional Ideology

Though the model is frequently estimated in multiple dimensions, only the first really matters much:



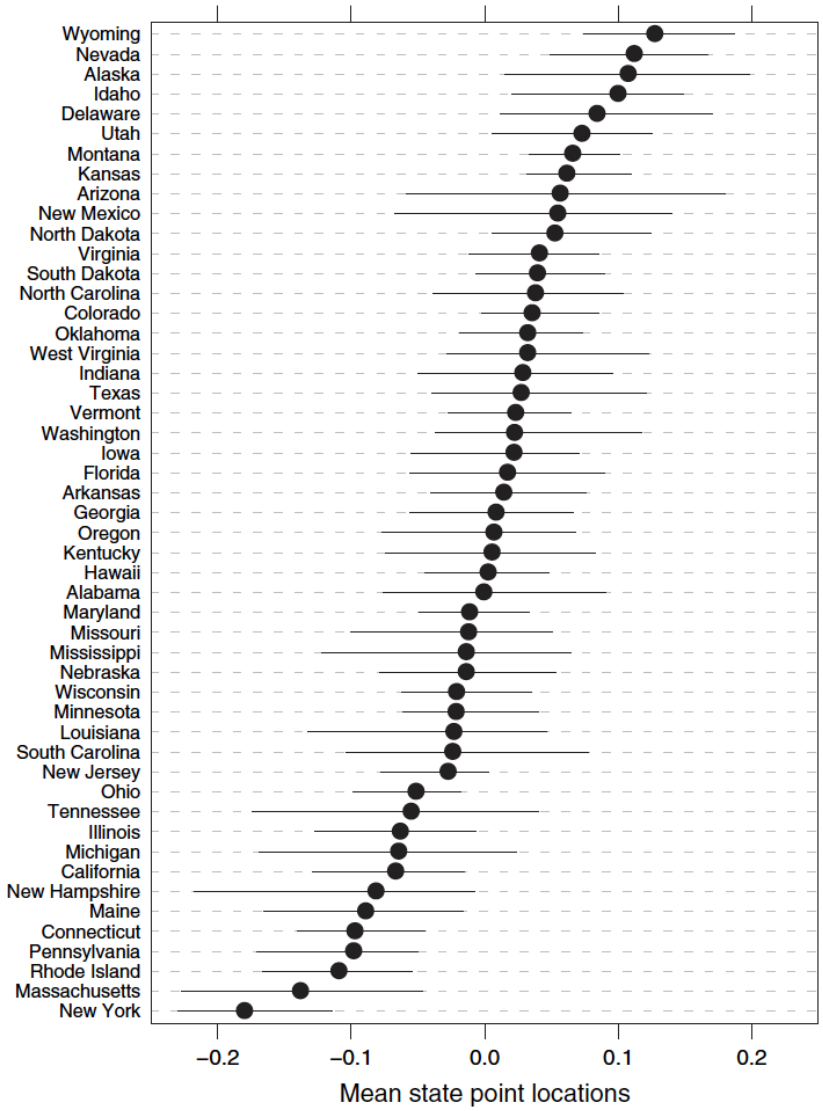
Example: State Spending Priorities

Locations of policy spending areas along latent dimension:

Interpretation: spending on particularized benefits vs. spending on collective goods

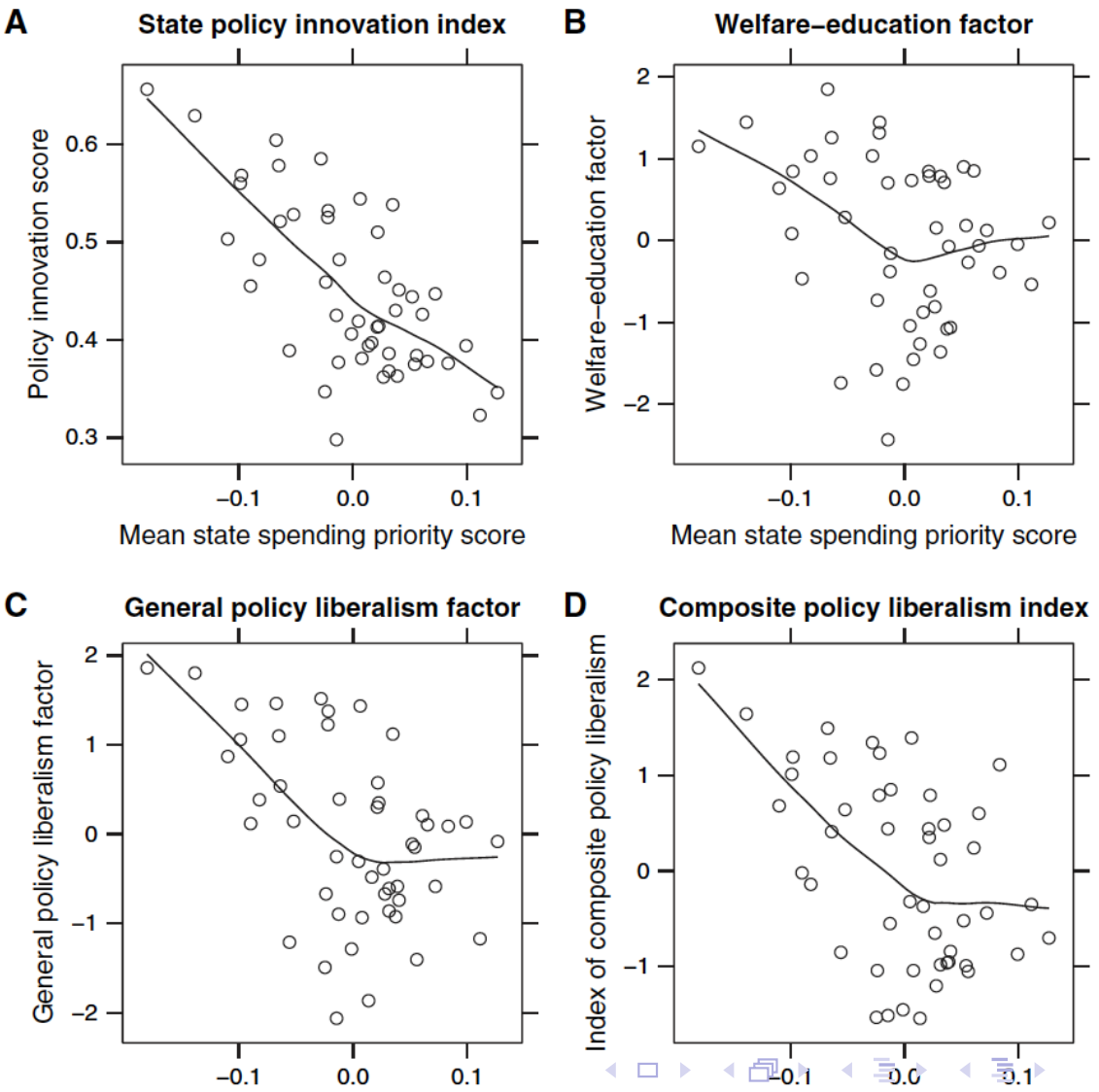
Example: State Spending Priorities

Average locations of states on latent continuum over many years:



Example: State Spending Priorities

(Criterion) validation by examining relationships with other indicators of state spending and ideology:



Parametric Unfolding

- Just like the cumulative model (MSA) has a family of parametric analogues, so too does the unfolding model
- Note: some think of the unfolding model as a particular form of IRT model
  - In a way that makes sense...the IRF is just a different shape
  - But, (cumulative) IRT models are **dominance** models, and unfolding is a **proximity** model
  - Thus, many psychometricians think of them as different
- The "mirt" package includes two IRT-based formulations of the unfolding model: the "dichotomous ideal point model" and the "generalized graded unfolding model" (GGUM)
- Can estimate these just like other IRT models, and use same person and item fit statistics to assess model fit

Other Options for Estimation

1. Optimal Classification, developed by Poole (2000, 2005)
   - R package called "oc"
   - Pros: can estimate unidimensional model, nonparametric
   - Cons: only dichotomous data, programmed in language of legislators/votes
2. Ordinal Optimal Classification, Hare et al. (2018)
   - Pros: nonparametric, ordinal data, can estimate unidimensional model
   - Con: not fully implemented in package yet
3. Smacof, developed by de Leeuw (lots of papers)
   - Pros: R package, ordinal and interval input data
   - Cons: won't fit unidimensional models, no dichotomous data
   - In a pinch could fit 2-dimensional model to unidimensional data and just use first dimension coordinates

Unfolding vs. Factor Analysis

Hypothetical dataset

	Policy		
	A	B	C
s_1	10	5	0
s_2	9	6	1
s_3	8	7	2
s_4	7	8	3
s_5	6	9	4
s_6	5	10	5
s_7	4	9	6
s_8	3	8	7
s_9	2	7	8
s_{10}	1	6	9
s_{11}	0	5	10

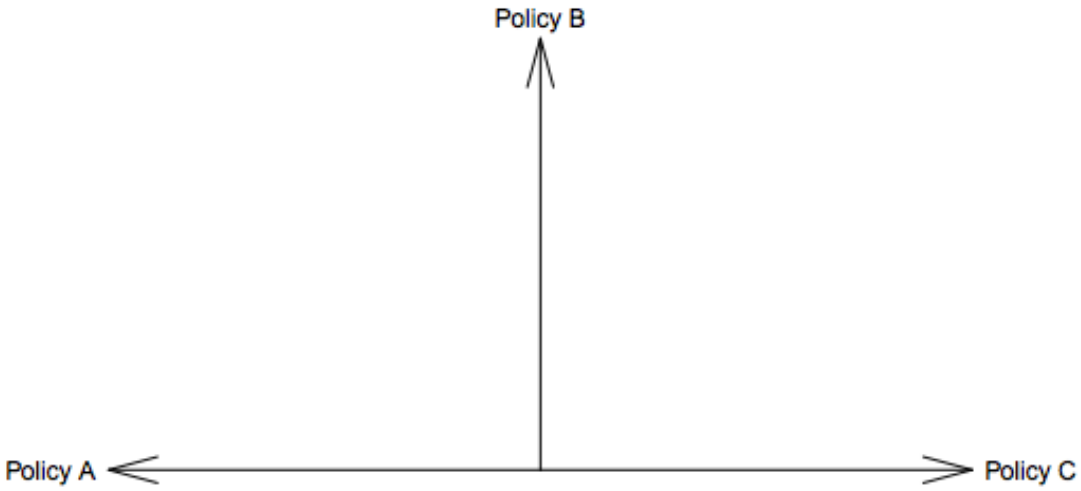
Unfolding vs. Factor Analysis

Correlation matrix

	A	B	C
A	1.00	0.00	-1.00
B	0.00	1.00	0.00
C	-1.00	0.00	1.00

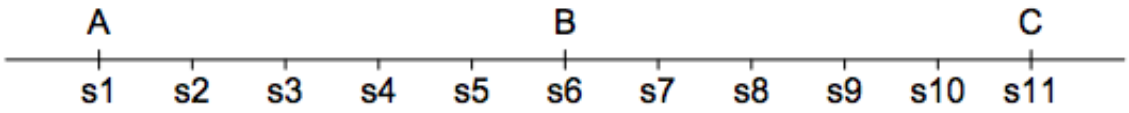
Unfolding vs. Factor Analysis

Results of factor analysis (which models correlations)



Unfolding vs. Factor Analysis

Results of unfolding analysis



Unfolding vs. Factor Analysis

- Bipolar constructs, in particular, can be "difficult" for factor analysis
- Oftentimes the two "halves" of a bipolar continuum end up being represented by two distinct latent factors
- Why?
  - Unfolding is a model of distance, or proximity
  - Factor analysis is a model of correlations, which correspond to the angular separation between pairs of variable vectors
- My point: dimensionality is much more theoretical and flexible than we tend to think
  - The most powerful measurement of a latent construct will start with some serious thinking about the DGP, and then finding a model that corresponds to it (and the data "type")