

# Thinking Outside the Diagonal

## Exploring Agreement in Rectangular Contingency Matrices

Tzur M. Karelitz  
PACE Center, Tufts University

David V. Budescu  
University of Illinois, Urbana-Champaign



# Square Contingency Matrices

		Judge B				
		1	2	3	4	5
Judge A	1					
	2					
	3					
	4					
	5					

Agreement is defined by the diagonal entries

# Rectangular Contingency Matrices

## Scale A

1. Very easy
2. Easy
3. Somewhat easy
4. Undecided
5. Somewhat difficult
6. Difficult
7. Very difficult

		Judge B				
		1	2	3	4	5
Judge A	1					
	2					
	3					
	4					
	5					
	6					
	7					

## Scale B

1. Very easy
2. Easy
3. Undecided
4. Difficult
5. Very difficult

# Rectangular Contingency Matrices

## Scale A

1. Very easy
2. Easy
3. Somewhat easy
4. Undecided
5. Somewhat difficult
6. Difficult
7. Very difficult

		Judge B				
		1	2	3	4	5
Judge A	1					
	2					
	3					
	4					
	5					
	6					
	7					

## Scale B

1. Very easy
2. Easy
3. Undecided
4. Difficult
5. Very difficult

# Rectangular Contingency Matrices

## Scale A

1. Very easy
2. Easy
3. Somewhat easy
4. Undecided
5. Somewhat difficult
6. Difficult
7. Very difficult

		Judge B				
		1	2	3	4	5
Judge A	1					
	2					
	3					
	4					
	5					
	6					
	7					

## Scale B

1. Very easy
2. Easy
3. Undecided
4. Difficult
5. Very difficult

# Rectangular Contingency Matrices

## Scale A

1. Very easy
2. Easy
3. Somewhat easy
4. Undecided
5. Somewhat difficult
6. Difficult
7. Very difficult

		Judge B				
		1	2	3	4	5
Judge A	1					
	2					
	3					
	4					
	5					
	6					
	7					

## Scale B

1. Very easy
2. Easy
3. Undecided
4. Difficult
5. Very difficult

# Rectangular Contingency Matrices

- Challenges:
  - The diagonal is ill-defined.
  - Common agreement measures cannot be computed.

*However, agreement information does exist in the data.*

- Naïve solutions:
  - “Square” the matrix by summing adjacent rows.
  - “Square” the matrix by adding zero columns.
    - Trick the software to think the matrix is square.

# Square Contingency Matrices

		Judge B				
		1	2	3	4	5
Judge A	1					
	2					
	3					
	4					
	5					

## Scale A

1. Strongly disagree
2. Disagree
3. Undecided
4. Agree
5. Strongly agree

## Scale B

1. Strongly disagree
2. Disagree
3. Undecided
4. Agree
5. Strongly agree

# Effectively Non-Square Contingency Matrices

		Senior				
		1	2	3	4	5
Junior	1	0	10	1	1	0
	2	0	20	3	1	0
	3	0	1	15	4	2
	4	0	2	4	10	5
	5	0	1	3	5	15

## Scale A

1. Strongly disagree
2. Disagree
3. Undecided
4. Agree
5. Strongly agree

## Scale B

1. *Strongly disagree*
2. Disagree
3. Undecided
4. Agree
5. Strongly agree

# Off-diagonal Agreement

		Senior				
		1	2	3	4	5
Junior	1	3	12	1	1	0
	2	1	2	23	1	0
	3	2	1	5	14	2
	4	0	2	4	4	15
	5	0	0	1	3	2

## Scale A

1. Strongly disagree
2. Disagree
3. Undecided
4. Agree
5. Strongly agree

## Scale B

1. Strongly disagree
2. Disagree
3. Undecided
4. Agree
5. Strongly agree

# Research Question

- How could agreement be measured when the diagonal is ill-defined?
  - Structurally non-square matrices (i.e., rectangular)
  - Effectively non-square matrices
  - Off-diagonal agreement patterns

# Common Measures of Agreement

$$\kappa = \frac{\sum n_{ii} - \sum m_{ii}}{N - \sum m_{ii}}$$

$n$  = observed frequency

$m$  = expected frequency (under independence)

$N$  = total

Cohen's Kappa (1960)

		Judge B				
		1	2	3	4	5
Judge A	1					
	2					
	3					
	4					
	5					

# Common Measures of Agreement

$$\kappa = \frac{\sum n_{ii} - \sum m_{ii}}{N - \sum m_{ii}}$$

Cohen's Kappa (1960)

$$\kappa_w = \frac{\sum_i^C \sum_j^C w_{ij} n_{ij} - \sum_i^C \sum_j^C w_{ij} m_{ij}}{N - \sum_i^C \sum_j^C w_{ij} m_{ij}}$$

Cohen's weighted Kappa (1968)

		Judge B				
		1	2	3	4	5
Judge A	1	1	0.75	0.50	0.25	0
	2	0.75	1	0.75	0.50	0.25
	3	0.50	0.75	1	0.75	0.50
	4	0.25	0.50	0.75	1	0.75
	5	0	0.25	0.50	0.75	1

$w$  = weights

$C$  = # categories

# Hierarchical Agreement

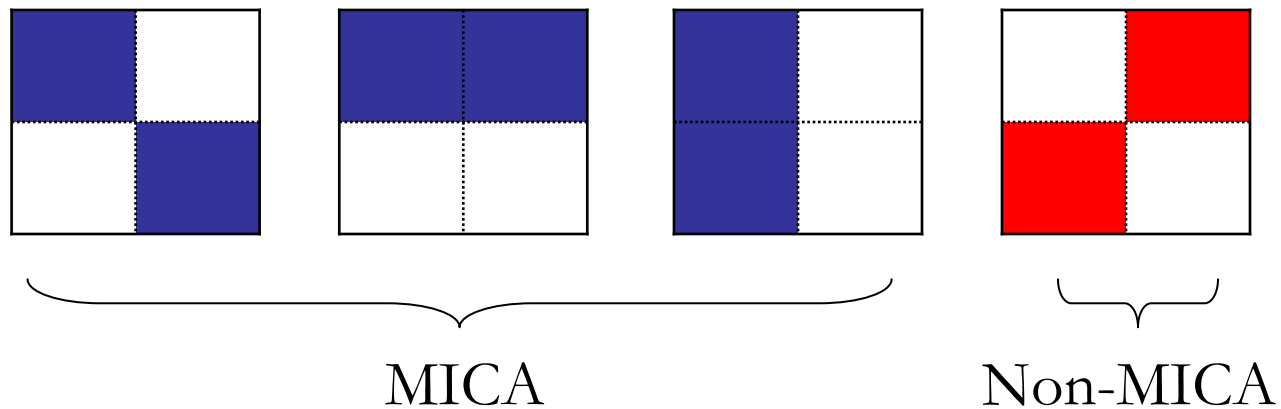
Hierarchical agreement weights (Landis & Koch, 1977, p 166)

		Judge B															
		$W_1$				$W_2$				$W_3$				$W_4$			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Judge A	1	1	0	0	0	1	1	0	0	1	1	0	0	1	1	0	0
	2	0	1	0	0	1	1	0	0	1	1	0	0	1	1	1	0
	3	0	0	1	0	0	0	1	0	0	0	1	1	0	1	1	1
	4	0	0	0	1	0	0	0	1	0	0	1	1	0	0	1	1

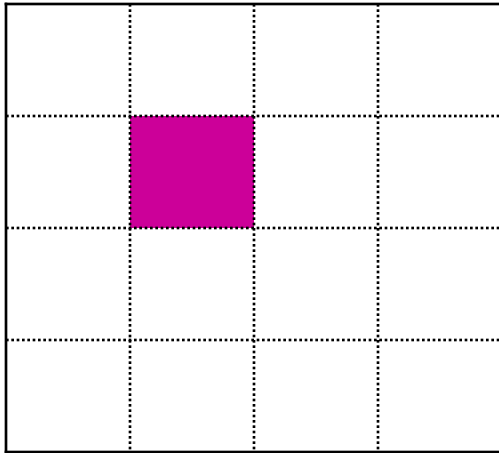
# Principles of the Proposed Framework

- There are multiple patterns that reflect *generalized agreement*.
- There are multiple measures of generalized agreement.
- Evidence for generalized agreement is not restricted to the diagonal.
- Generalized agreement should depend on the frequencies of the cells included in the pattern.
- Agreement patterns consist of *contiguous and monotonically increasing patterns* within the matrix.

# Monotonically Increasing & Contiguous Agreement (MICA) Patterns

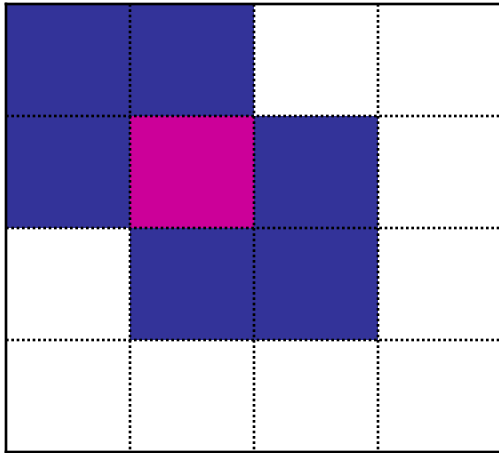


# Monotonically Increasing & Contiguous Agreement (MICA) Patterns



Starting point

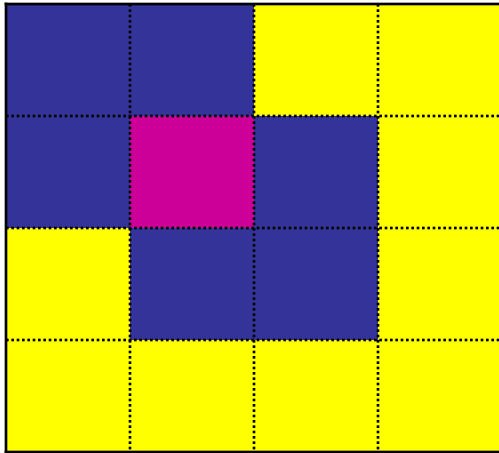
# Monotonically Increasing & Contiguous Agreement (MICA) Patterns



Starting point

Possible next cells

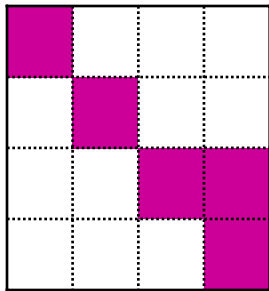
# Monotonically Increasing & Contiguous Agreement (MICA) Patterns



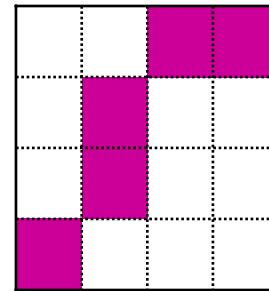
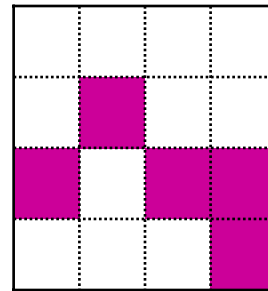
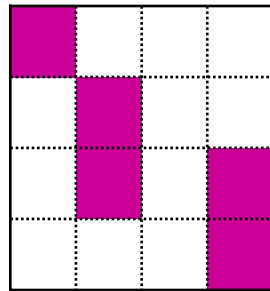
Starting point

Possible next cells

Non-admissible next cells



MICA



Non- MICA

# The MICA Algorithm

Step 1 – Choose a cell such that criterion  $Q$  is maximized.

...

Step  $t$  – Choose a cell such that:

- a. it maintains a *monotonically increasing and contiguous* pattern with the existing cells, and
- b. criterion  $Q$  is maximized.

# Maximization Criteria

- Observed Value:  $n_{ij}$
- Likelihood Ratio:  $n_{ij} \cdot \log(n_{ij} / m_{ij})$

*Other possibilities*

- Difference:  $n_{ij} - m_{ij}$
- Pearson:  $(n_{ij} - m_{ij})^2 / m_{ij}$

*There could be more...*

# Measure of Generalized Agreement

- At step  $t$ , agreement is measured based only on the cells included in pattern at that step (i.e., their weights is 1 and all other cells have 0 weights).

Generalized Kappa for the pattern at step  $t$  -

$$\kappa_t = \frac{\sum n_t - \sum m_t}{N - \sum m_t}$$

$n_t$  = observed frequency of pattern cells at step  $t$

$m_t$  = observed frequency of pattern cells at step  $t$

# Hypothetical Example: Rectangular Matrix

<i>n</i> (m)	1	2	3
1	<b>11</b> (4.5)	<b>2</b> (4.5)	<b>1</b> (5.0)
2	<b>16</b> (7.4)	<b>6</b> (7.4)	<b>1</b> (8.3)
3	<b>3</b> (8.0)	<b>18</b> (8.0)	<b>4</b> (9.0)
4	<b>2</b> (6.4)	<b>5</b> (6.4)	<b>13</b> (7.2)
5	<b>0</b> (5.8)	<b>1</b> (5.8)	<b>17</b> (6.5)

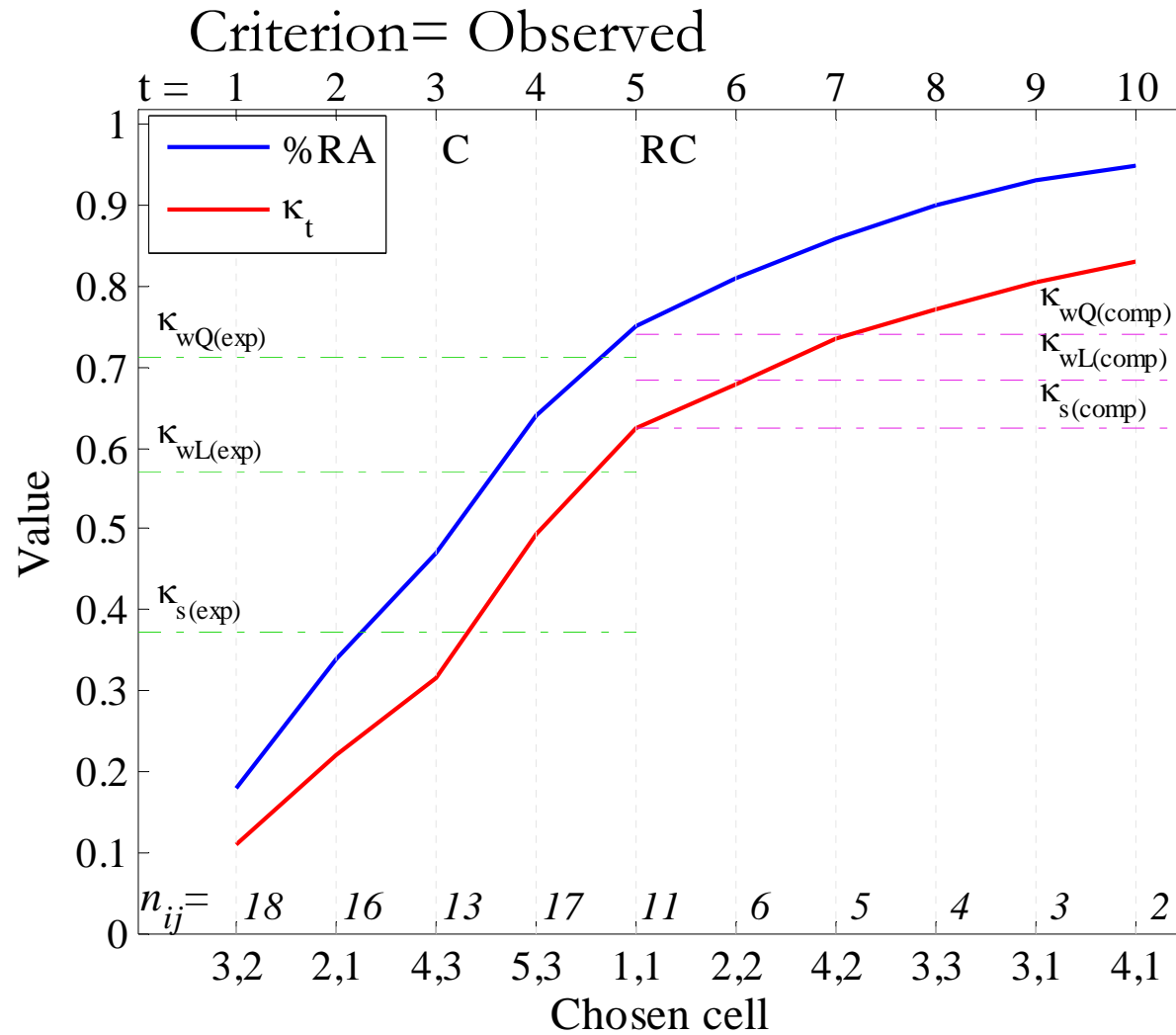
Best Expansion

		1	2		3
1	0	11	2	0	1
2	0	16	6	0	1
3	0	3	18	0	4
4	0	2	5	0	13
5	0	0	1	0	17

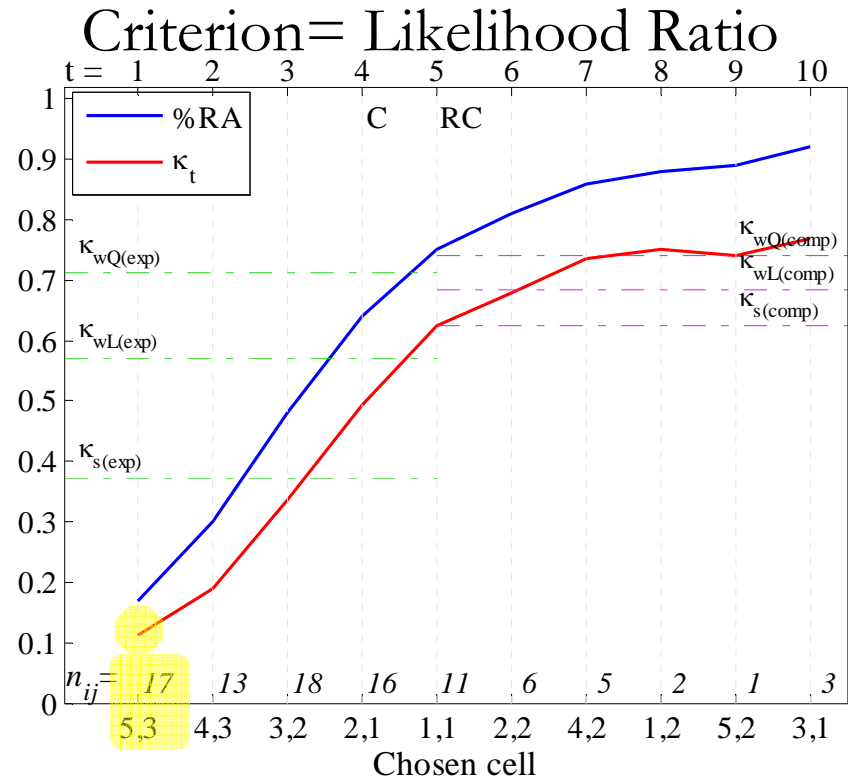
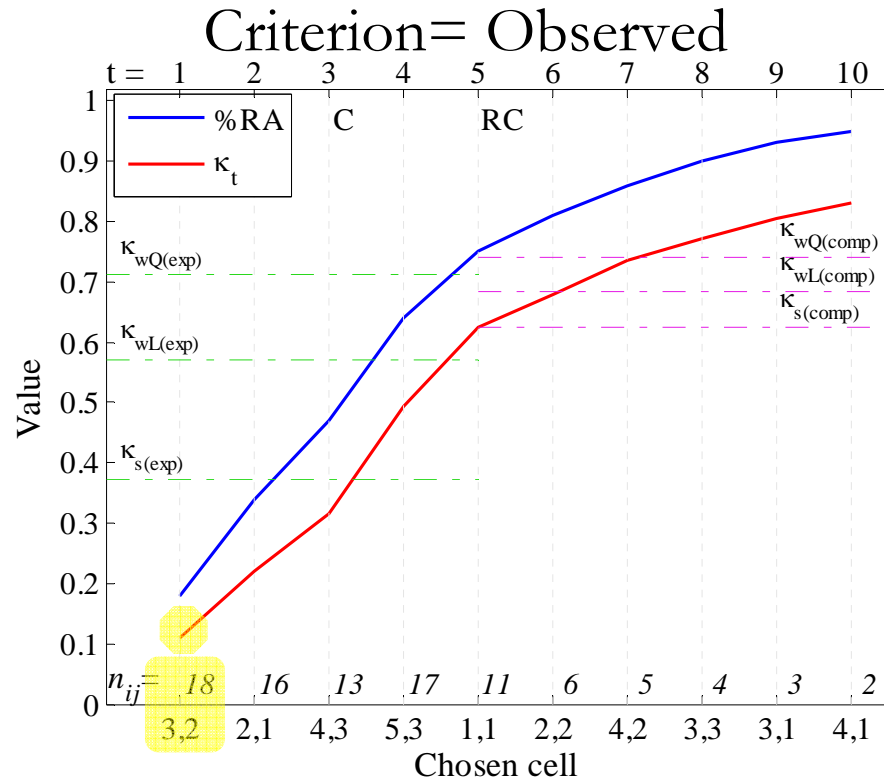
Best Compression

	1	2	3
1+2	27	8	2
3	3	18	4
4+5	2	6	30

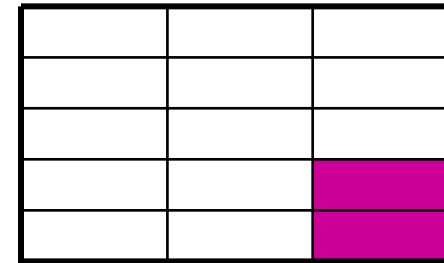
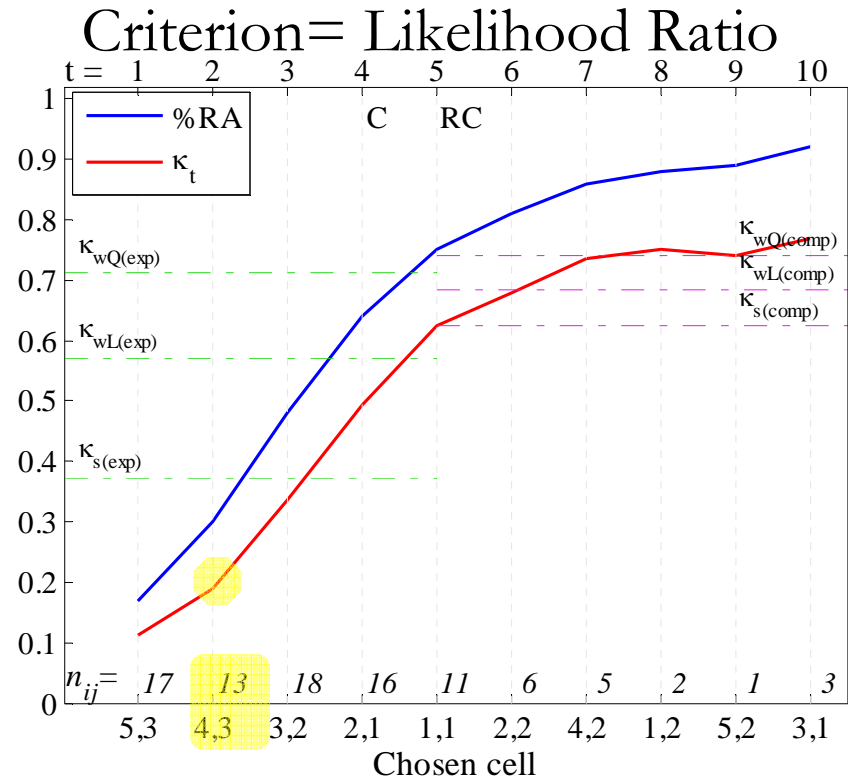
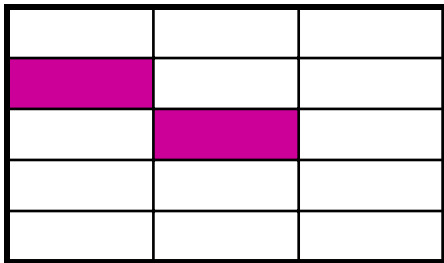
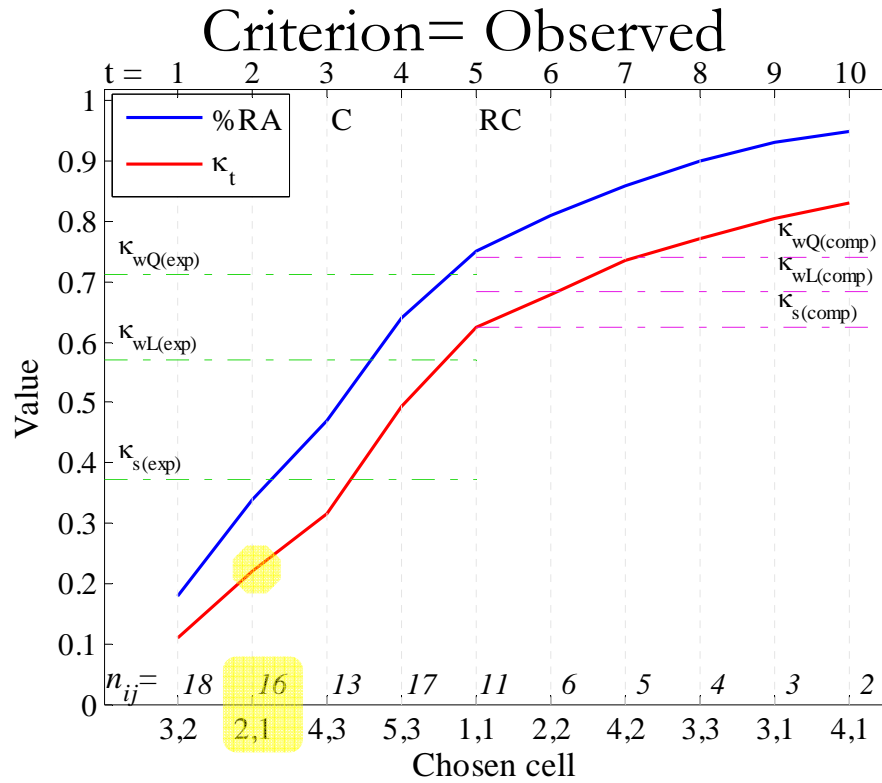
# Hypothetical Example Results



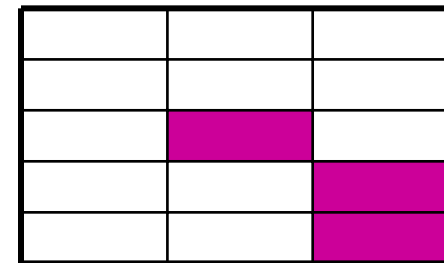
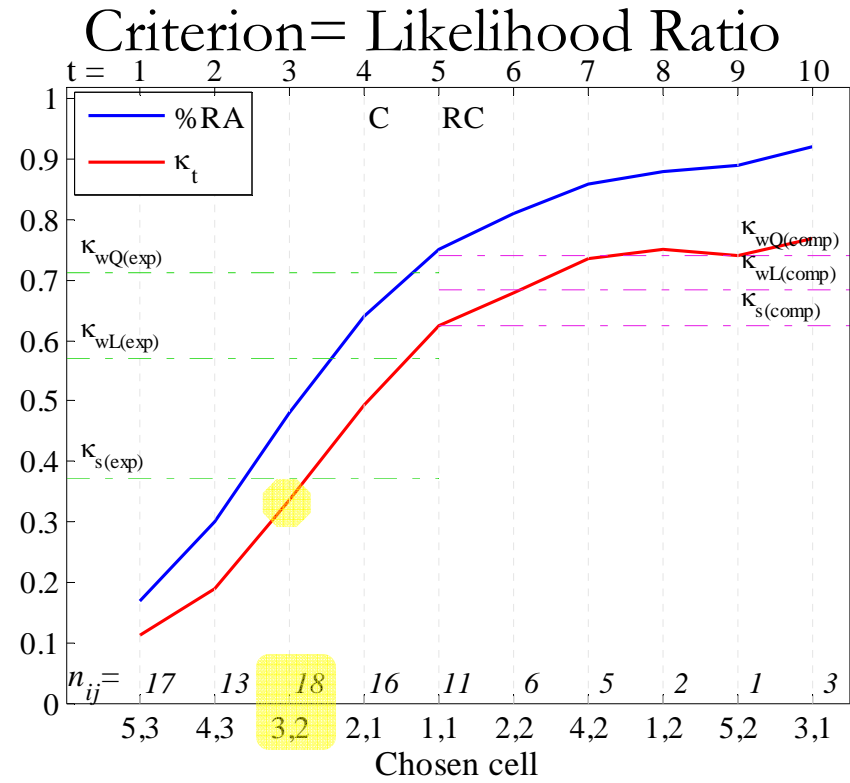
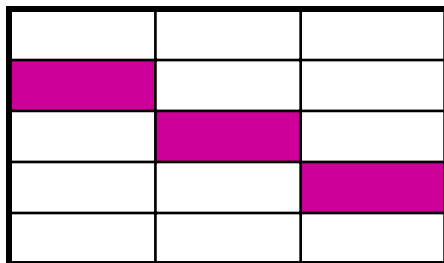
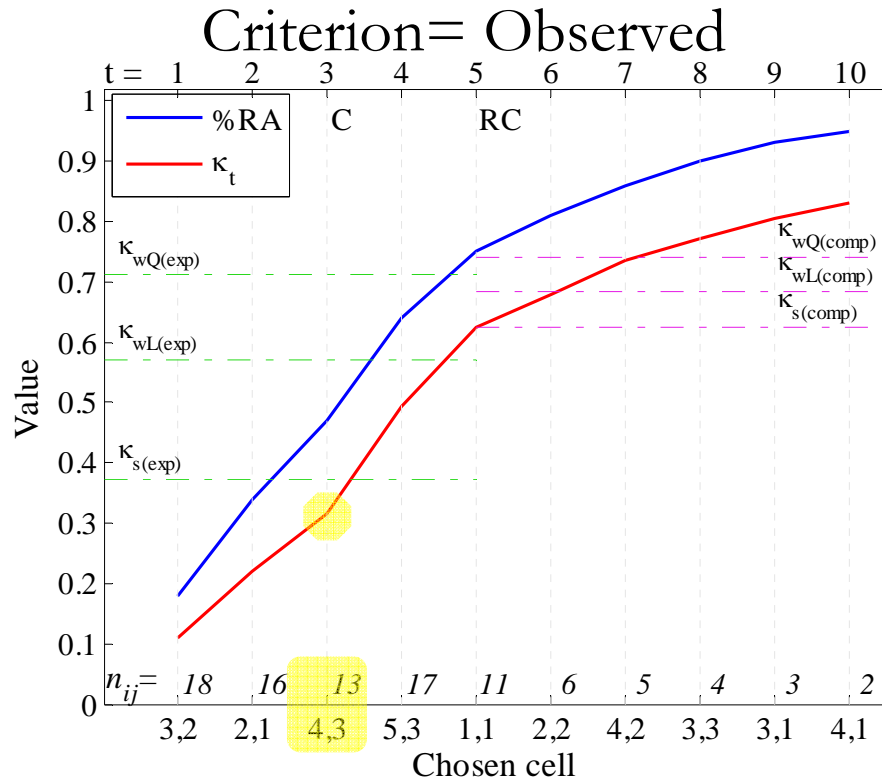
# Hypothetical Example Results



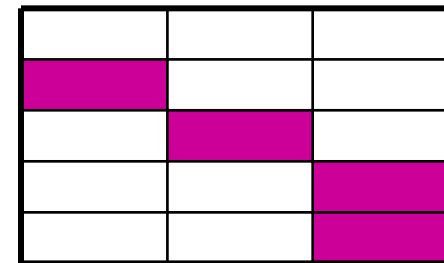
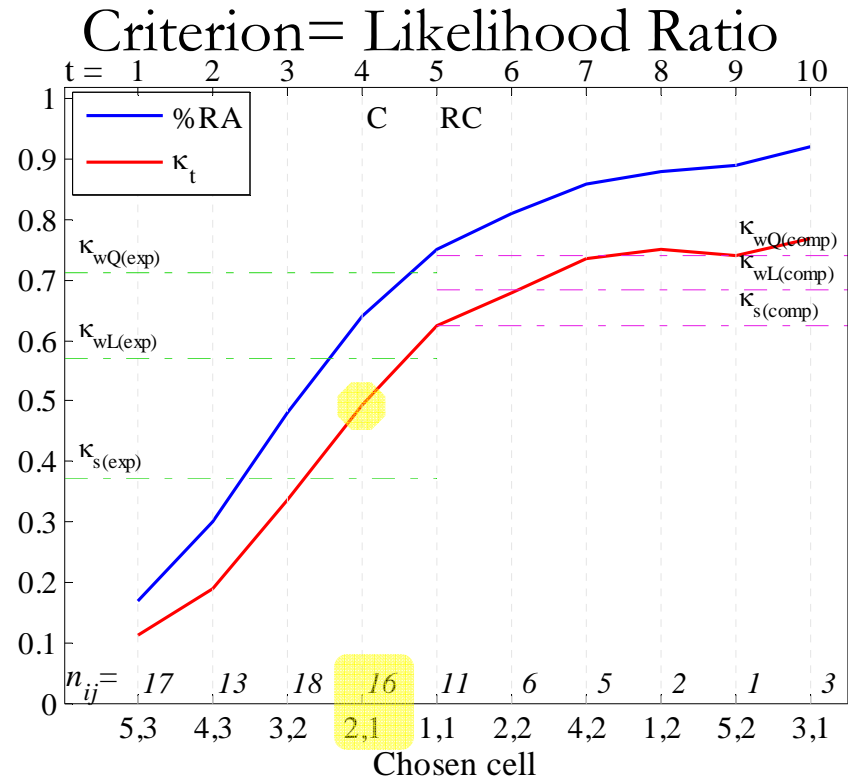
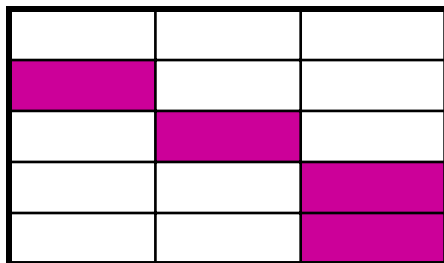
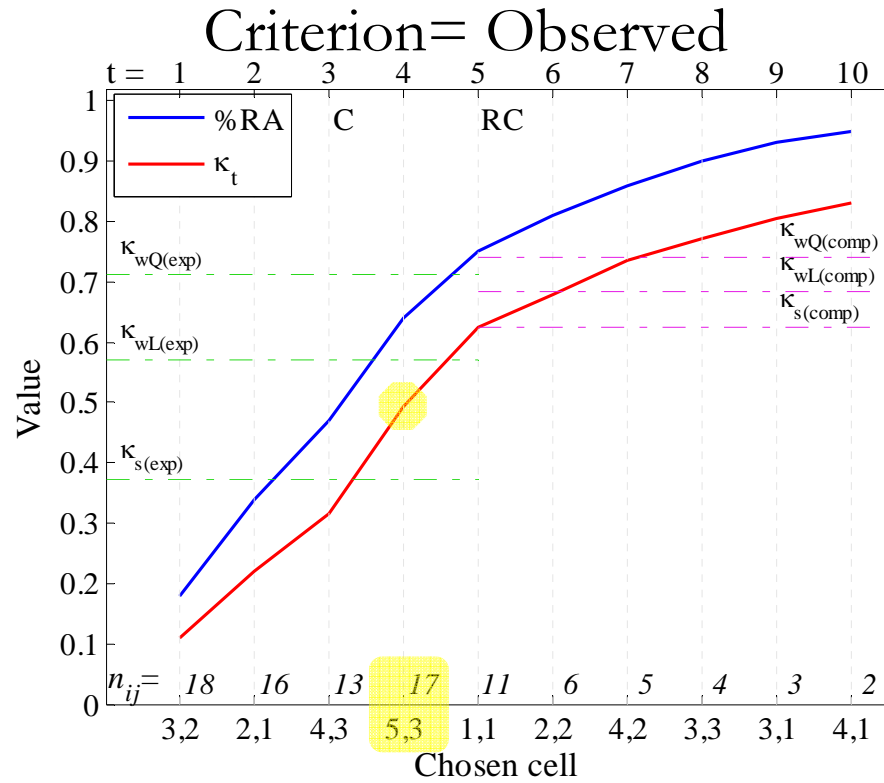
# Hypothetical Example Results



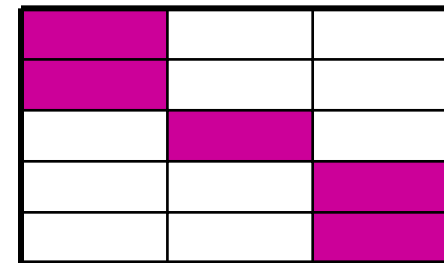
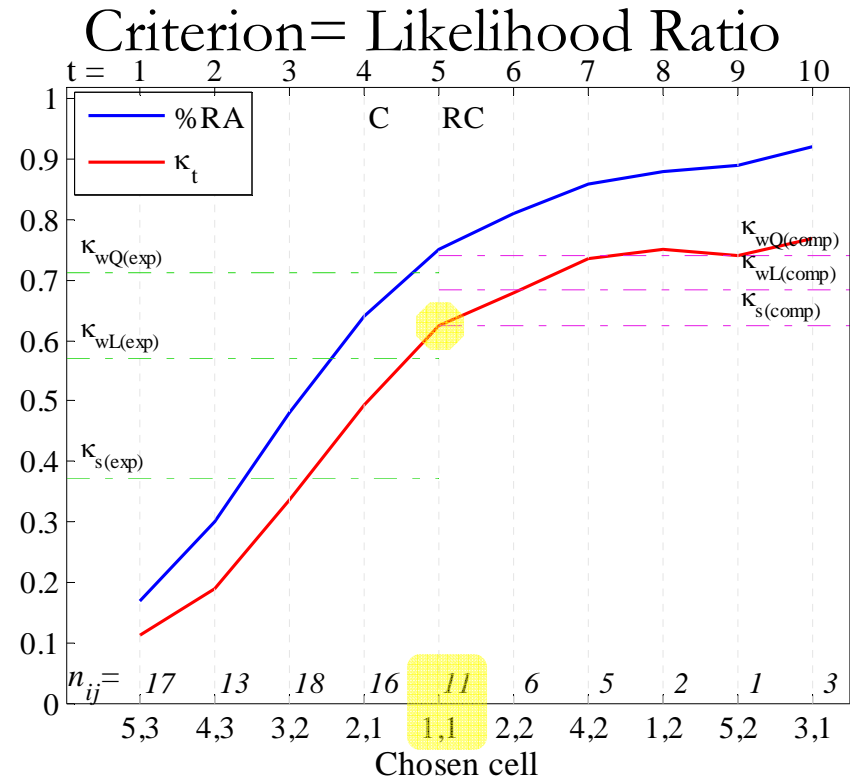
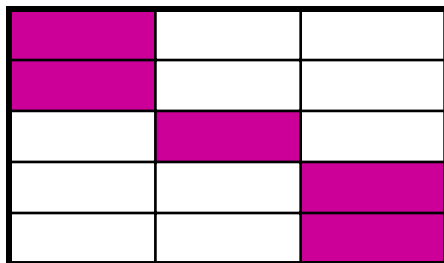
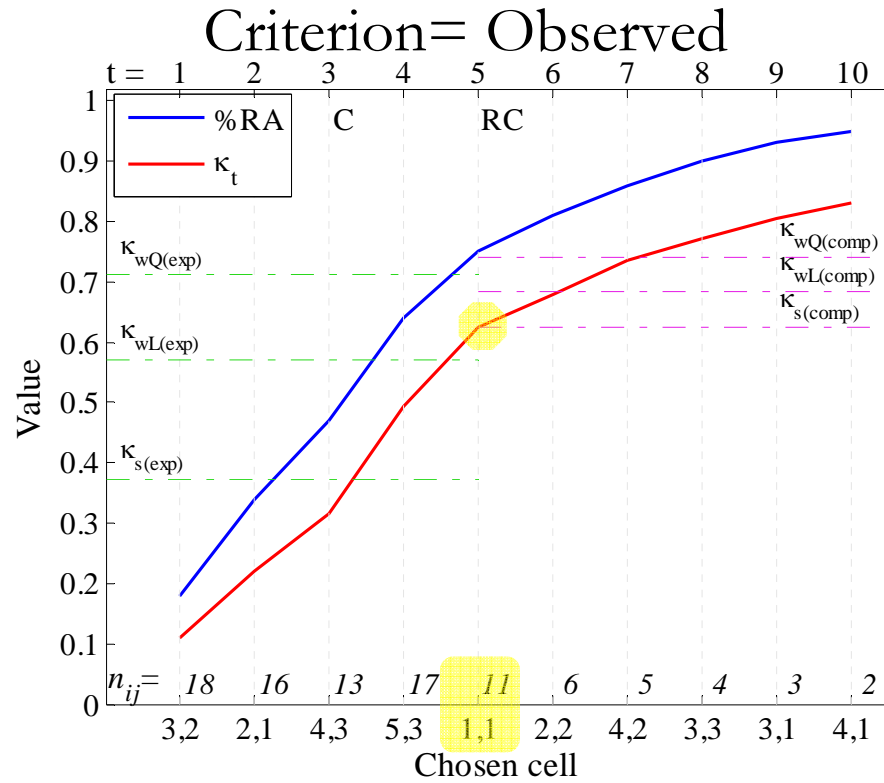
# Hypothetical Example Results



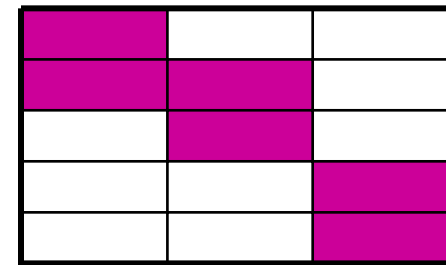
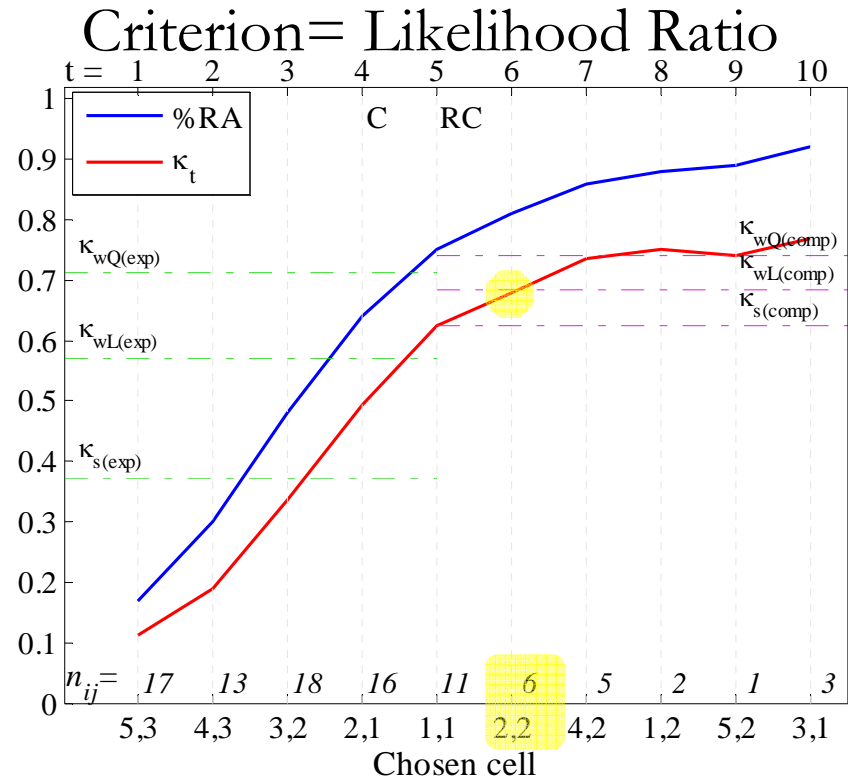
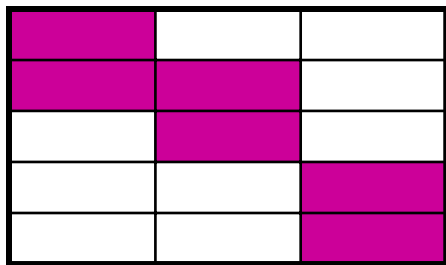
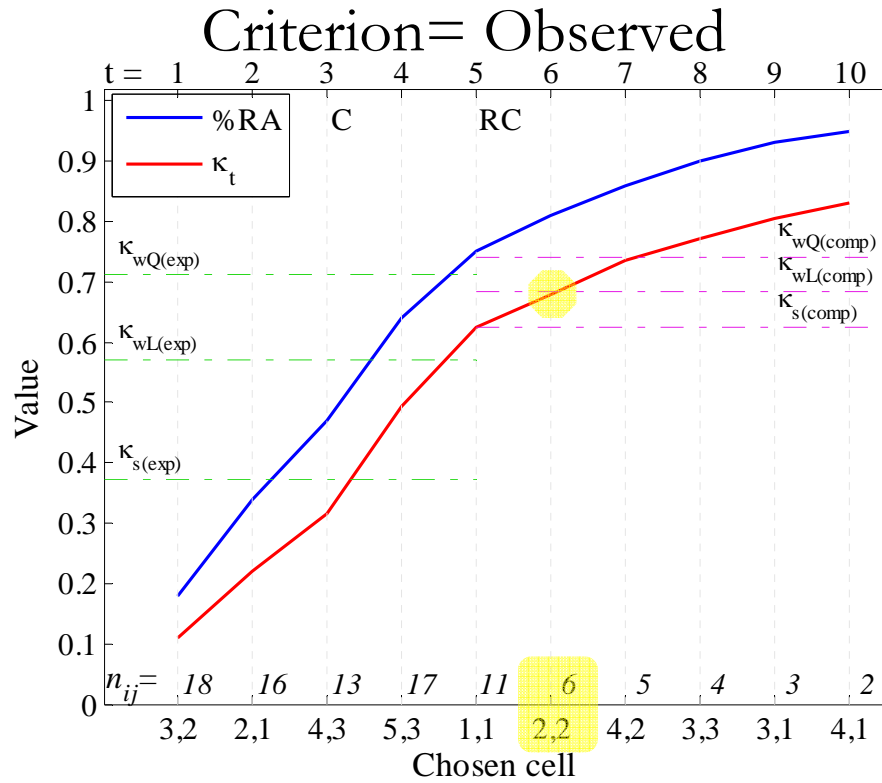
# Hypothetical Example Results



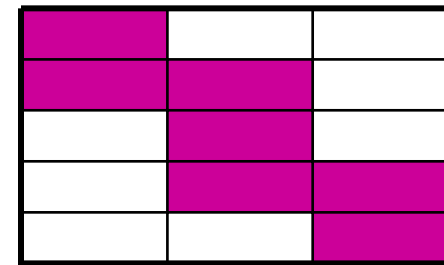
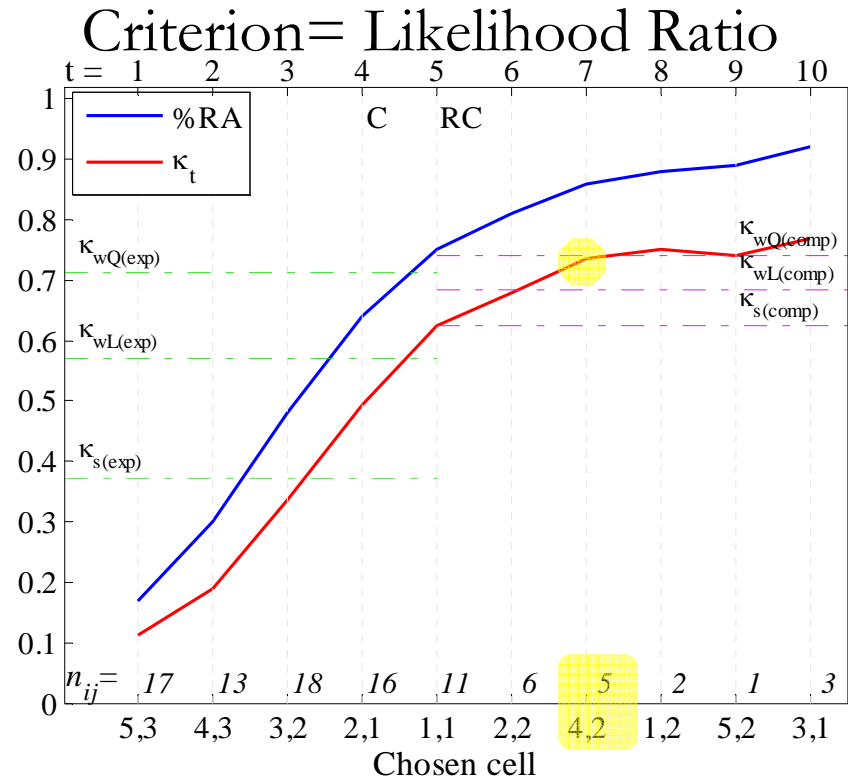
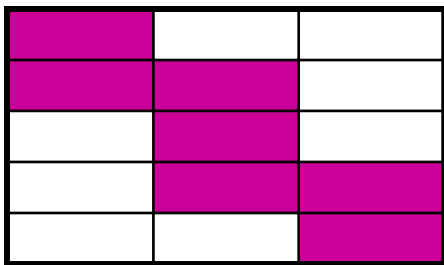
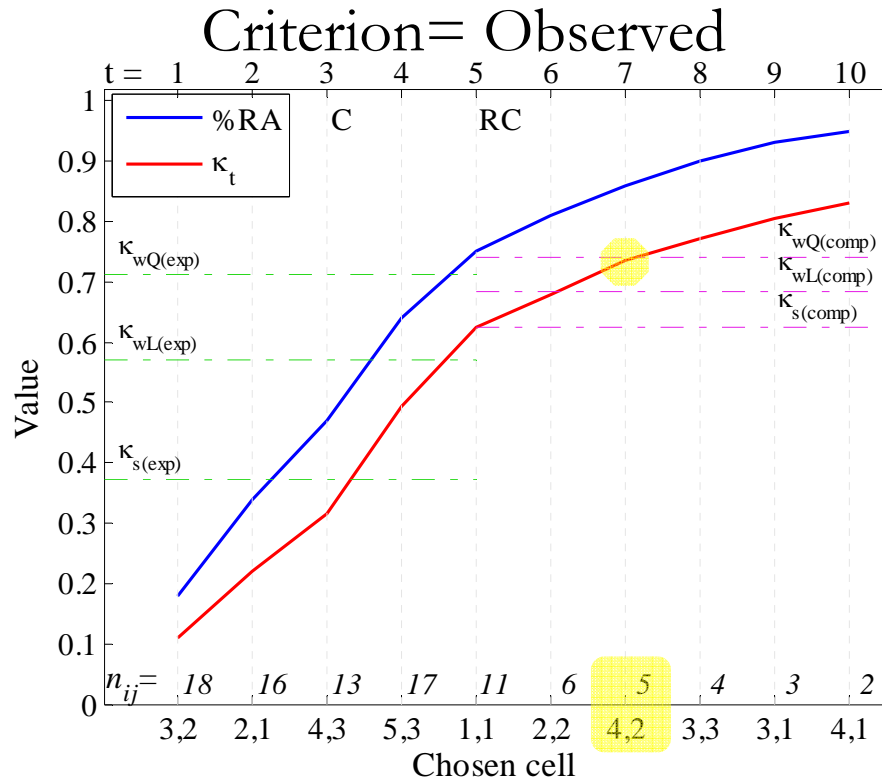
# Hypothetical Example Results



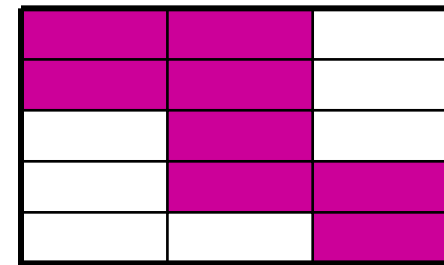
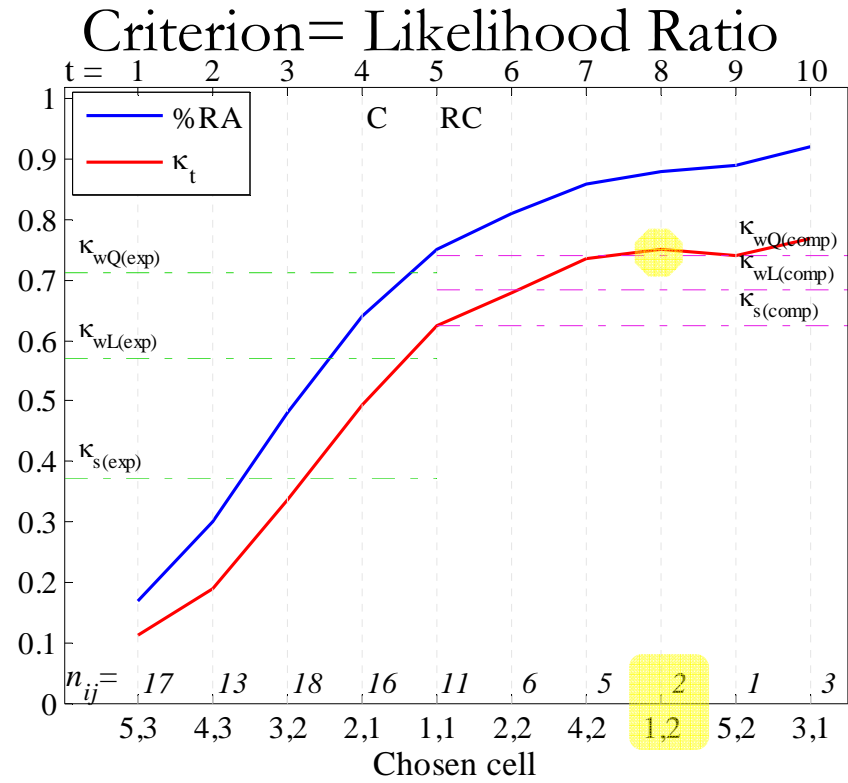
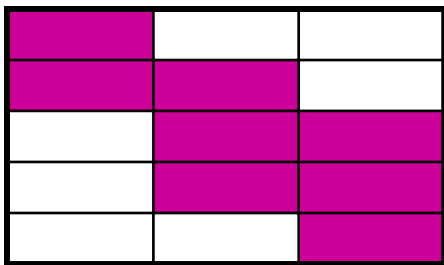
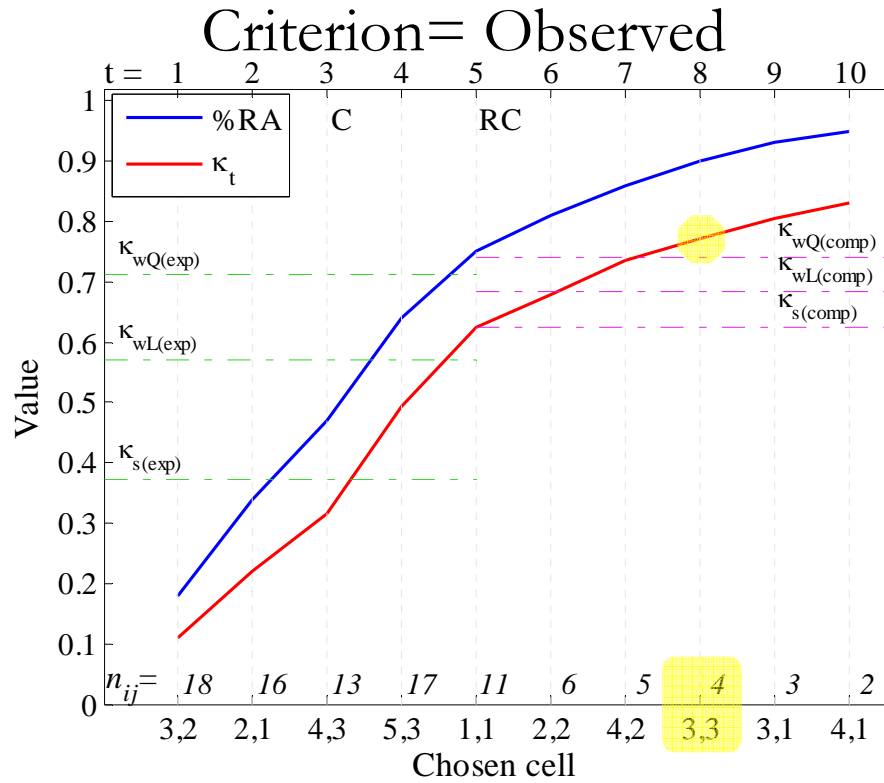
# Hypothetical Example Results



# Hypothetical Example Results



# Hypothetical Example Results



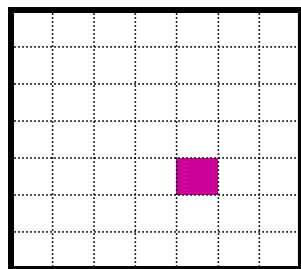
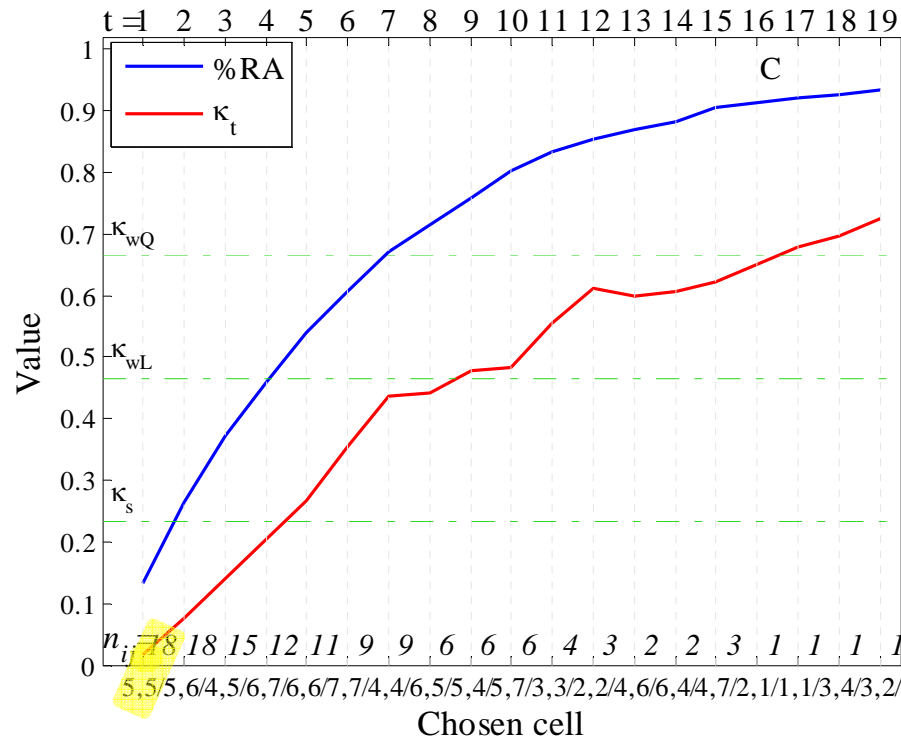
# Real Example: Square Matrix

Ratings on the Rhetorical Structure dimension  
for the Hebrew Proficiency Test

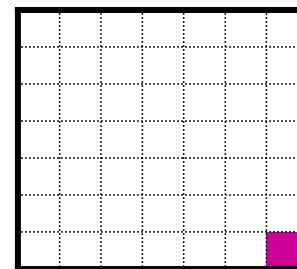
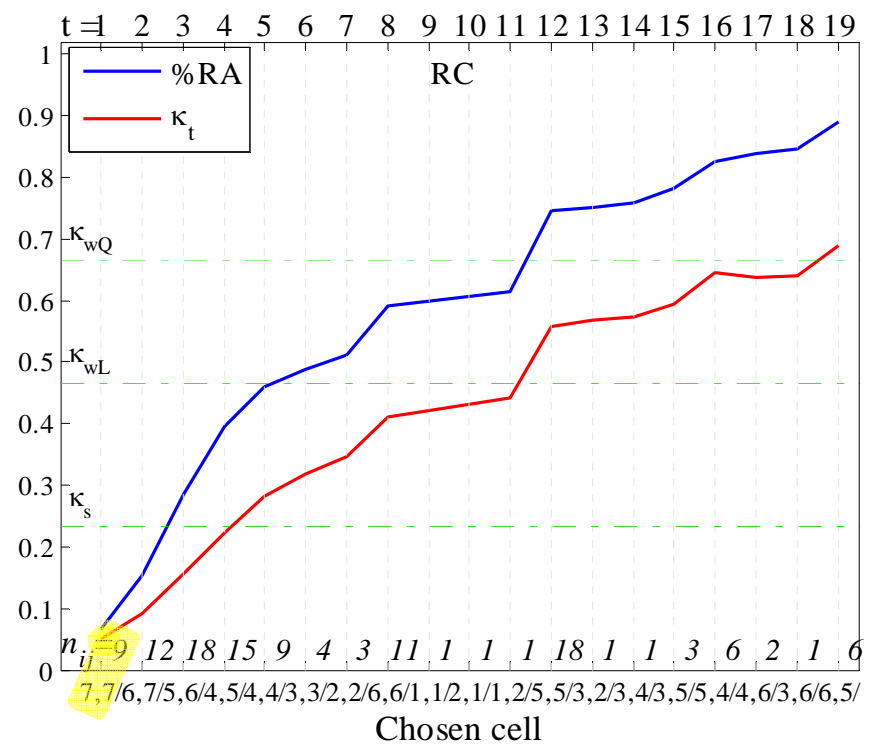
		Rater 2						
		<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>
Rater 1	<i>1</i>	<b>1</b>	<b>1</b>	0	0	0	0	0
	<i>2</i>	<b>1</b>	<b>3</b>	0	0	0	0	0
	<i>3</i>	0	<b>1</b>	<b>4</b>	<b>1</b>	<b>3</b>	<b>1</b>	0
	<i>4</i>	0	0	0	<b>9</b>	<b>15</b>	<b>2</b>	<b>3</b>
	<i>5</i>	0	0	<b>2</b>	<b>6</b>	<b>18</b>	<b>18</b>	<b>6</b>
	<i>6</i>	0	0	0	<b>2</b>	<b>6</b>	<b>11</b>	<b>12</b>
	<i>7</i>	0	0	0	0	<b>2</b>	0	<b>9</b>

# Results for Rhetorical Ratings

Criterion= Observed

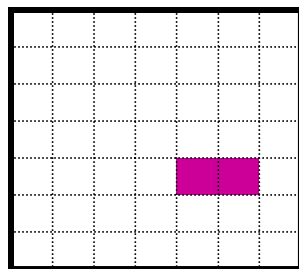
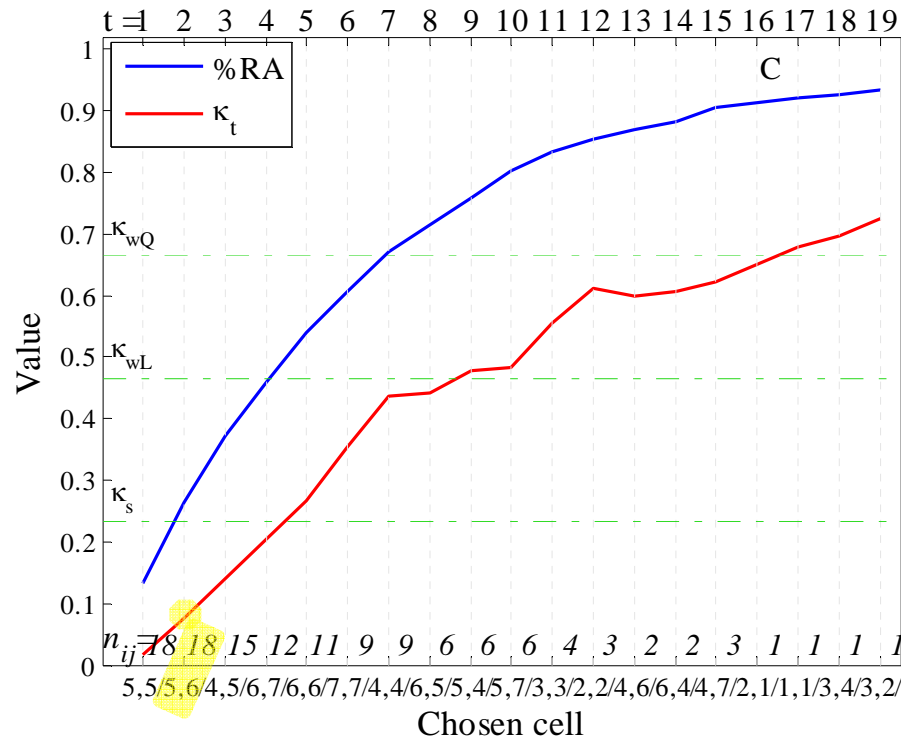


Criterion= Likelihood Ratio

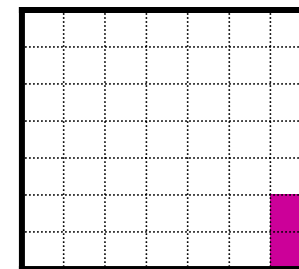
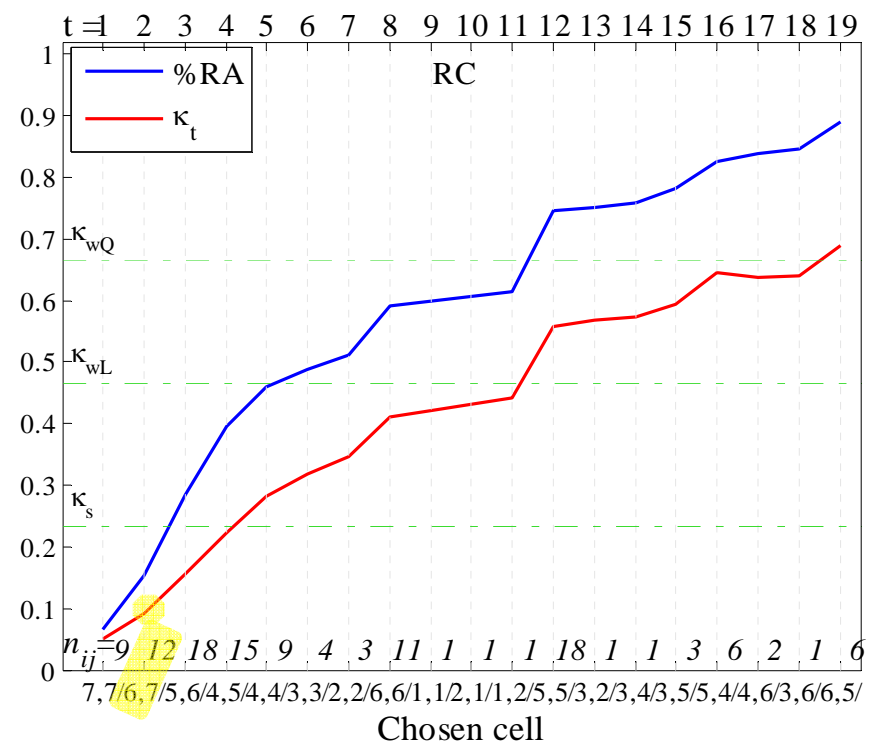


# Results for Rhetorical Ratings

Criterion= Observed

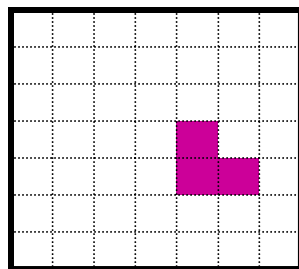
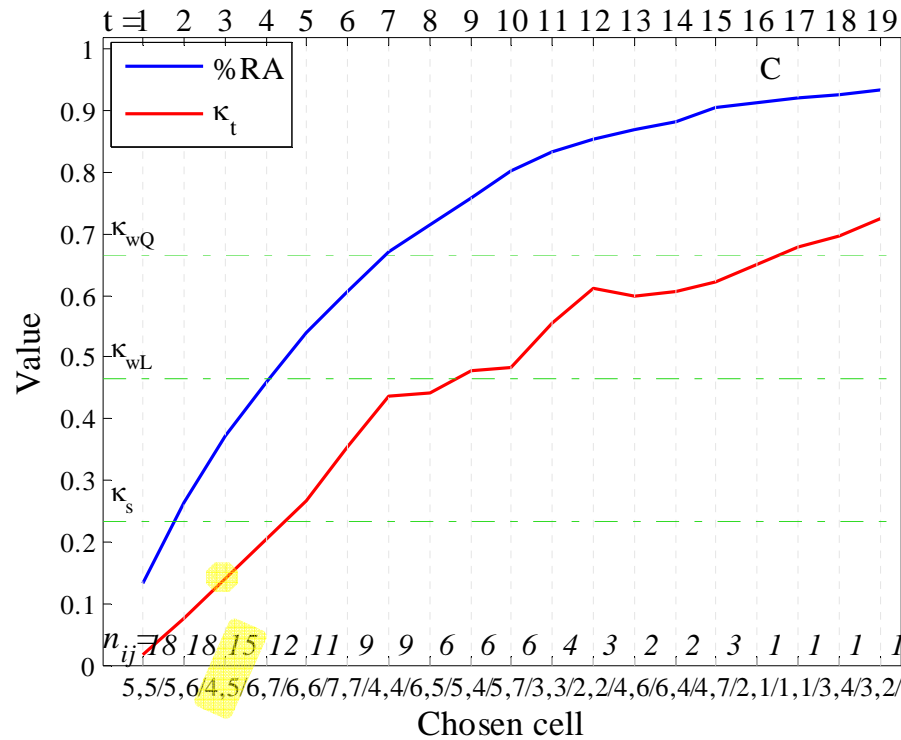


Criterion= Likelihood Ratio

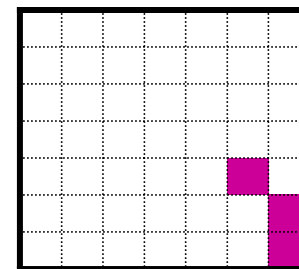
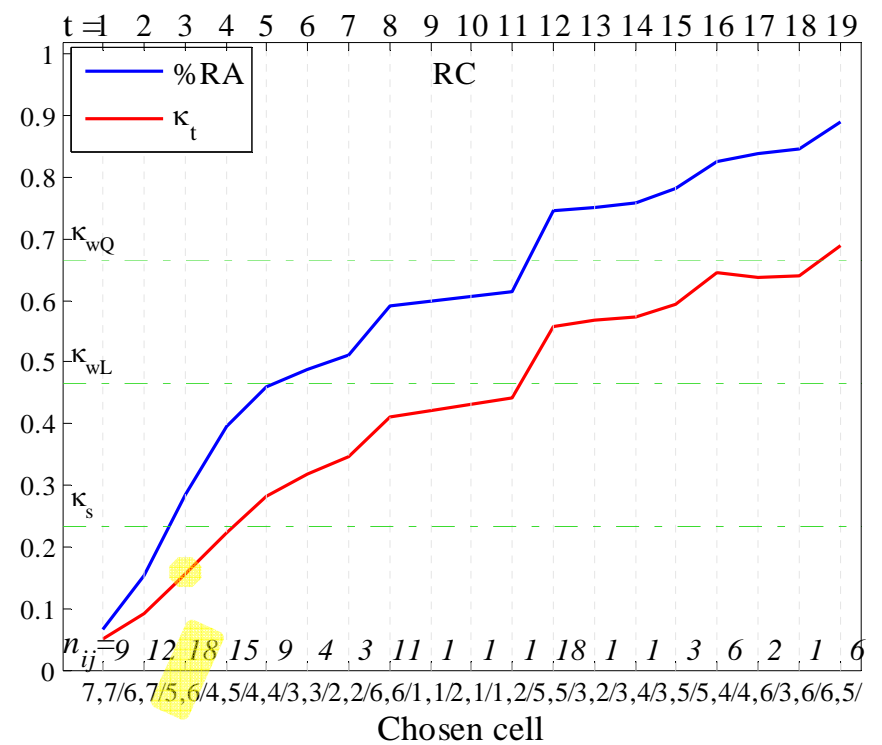


# Results for Rhetorical Ratings

Criterion= Observed

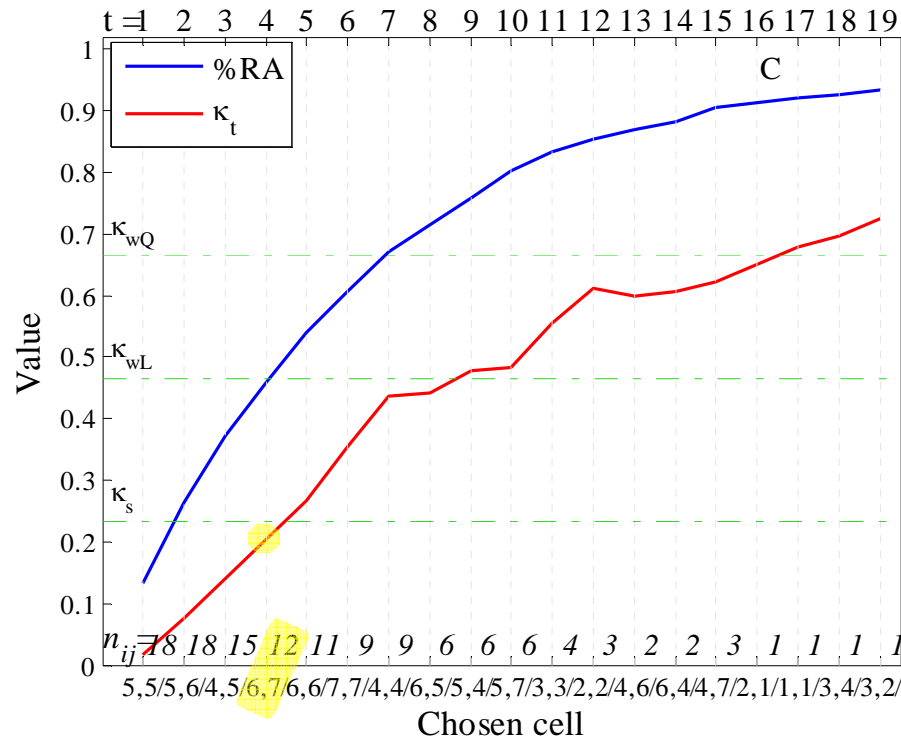


Criterion= Likelihood Ratio

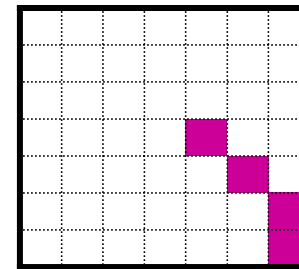
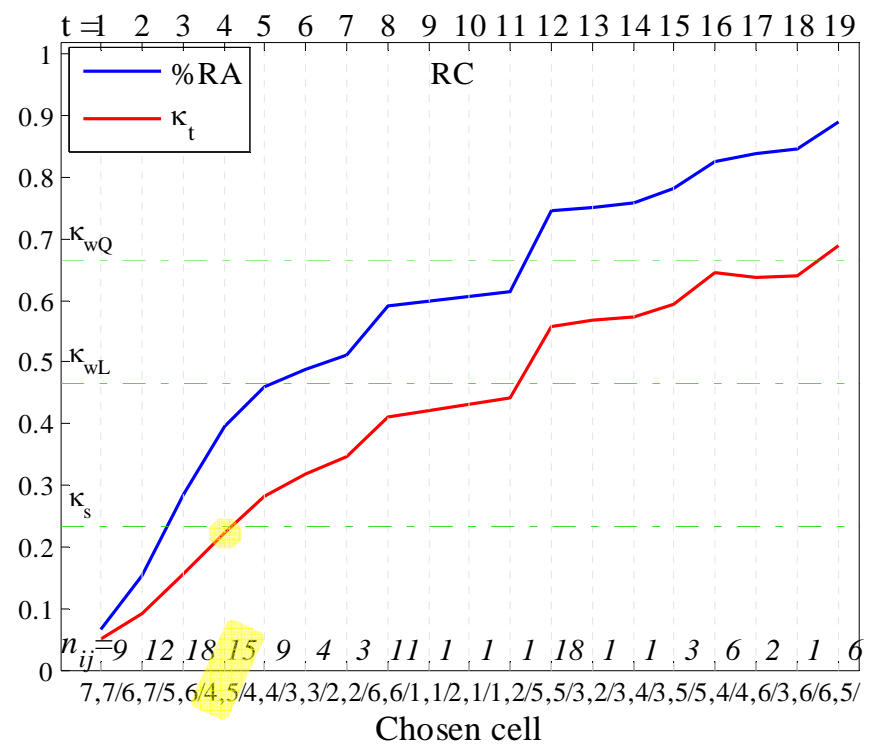


# Results for Rhetorical Ratings

Criterion= Observed

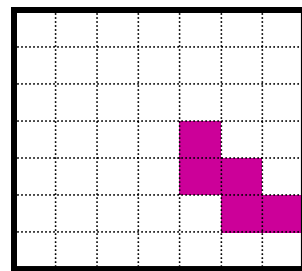
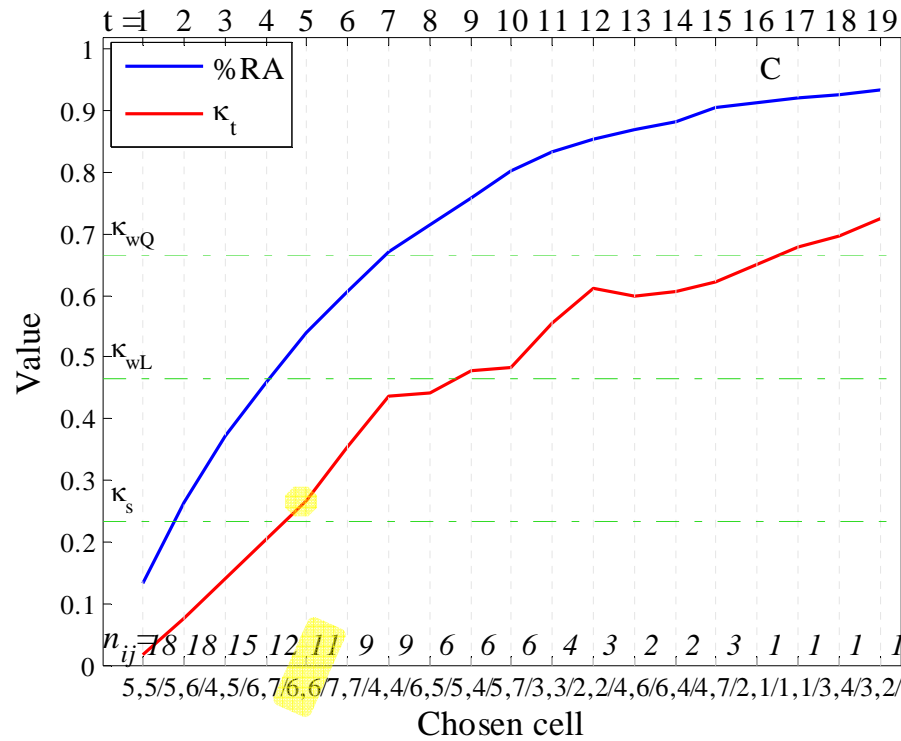


Criterion= Likelihood Ratio

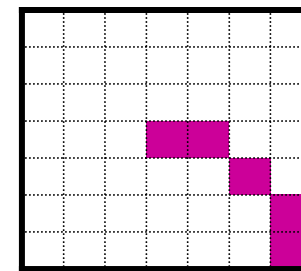
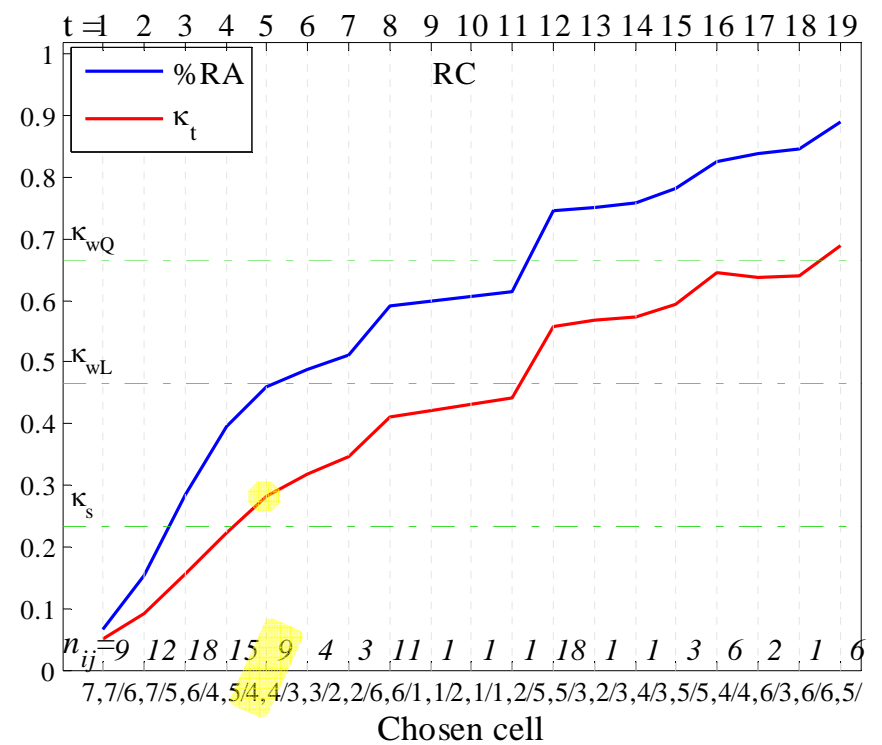


# Results for Rhetorical Ratings

Criterion= Observed

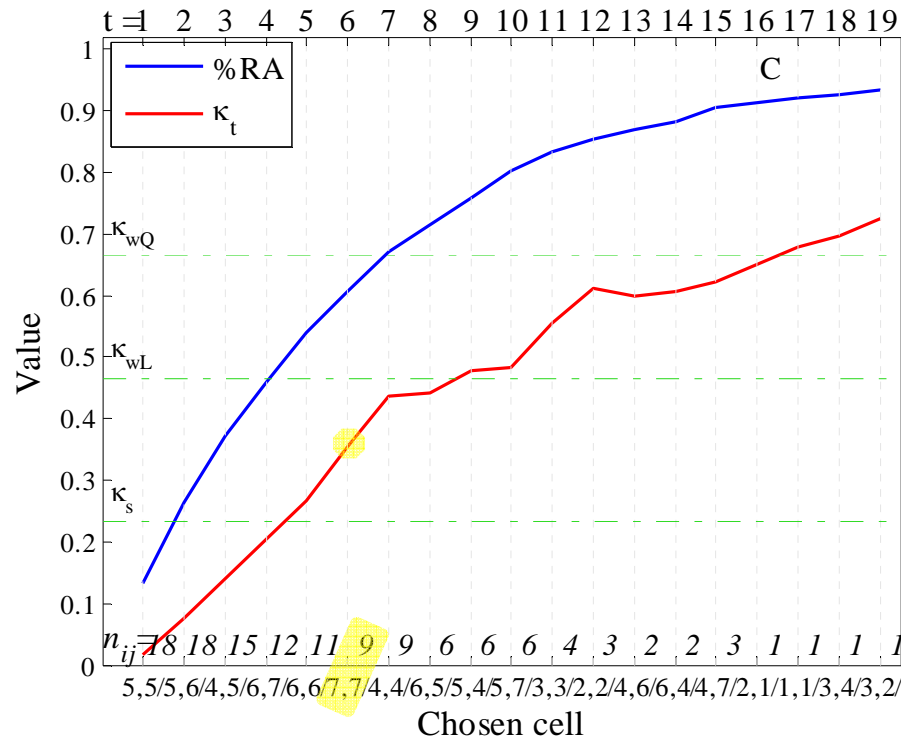


Criterion= Likelihood Ratio

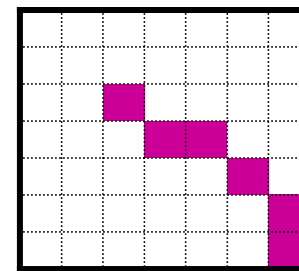
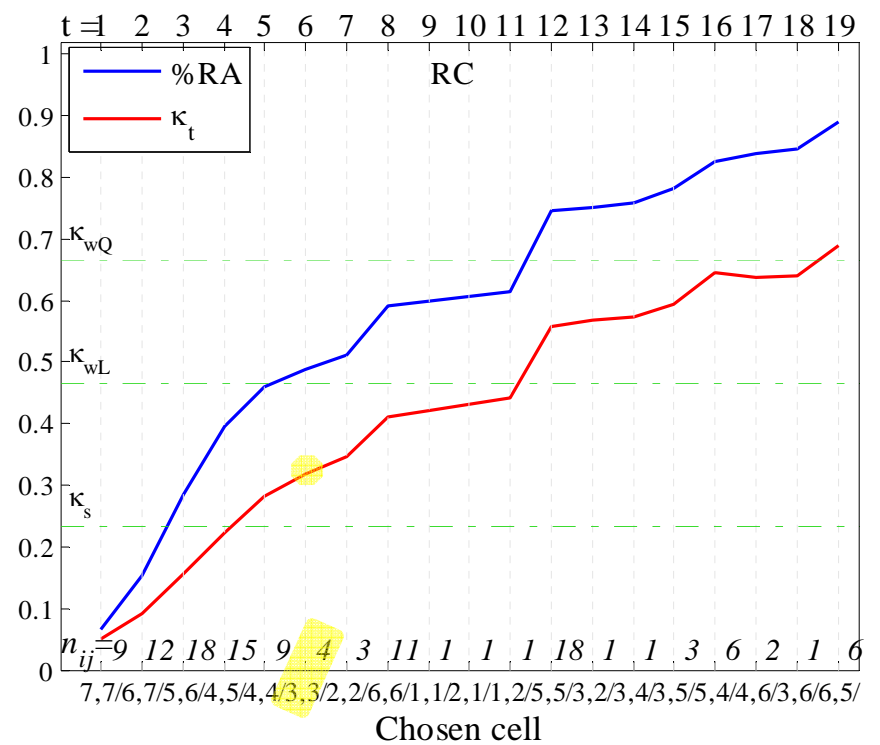


# Results for Rhetorical Ratings

Criterion= Observed

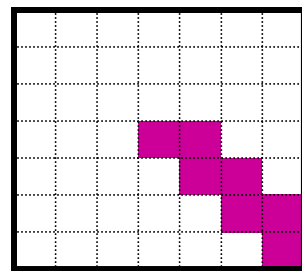
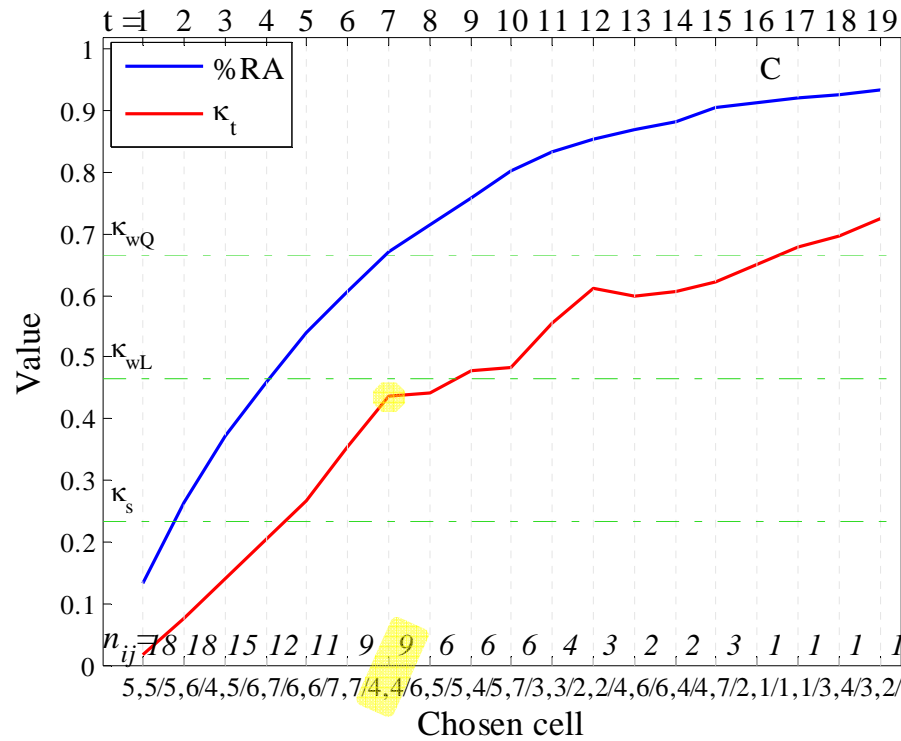


Criterion= Likelihood Ratio

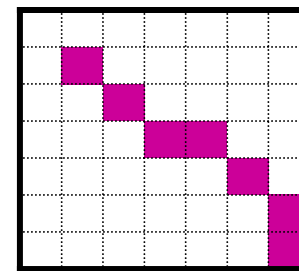
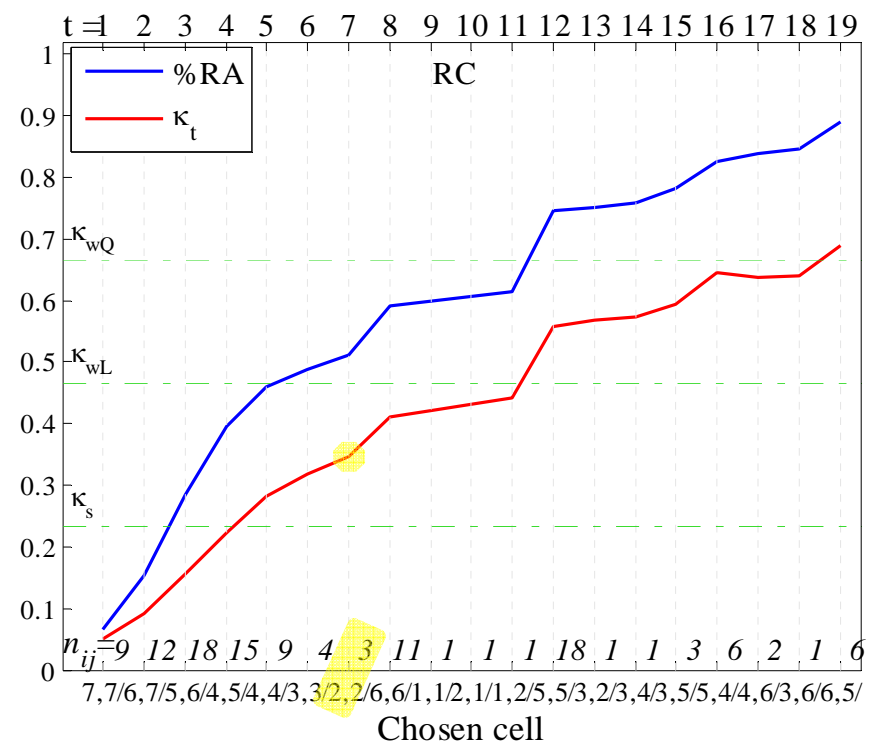


# Results for Rhetorical Ratings

Criterion= Observed



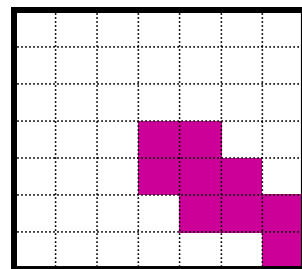
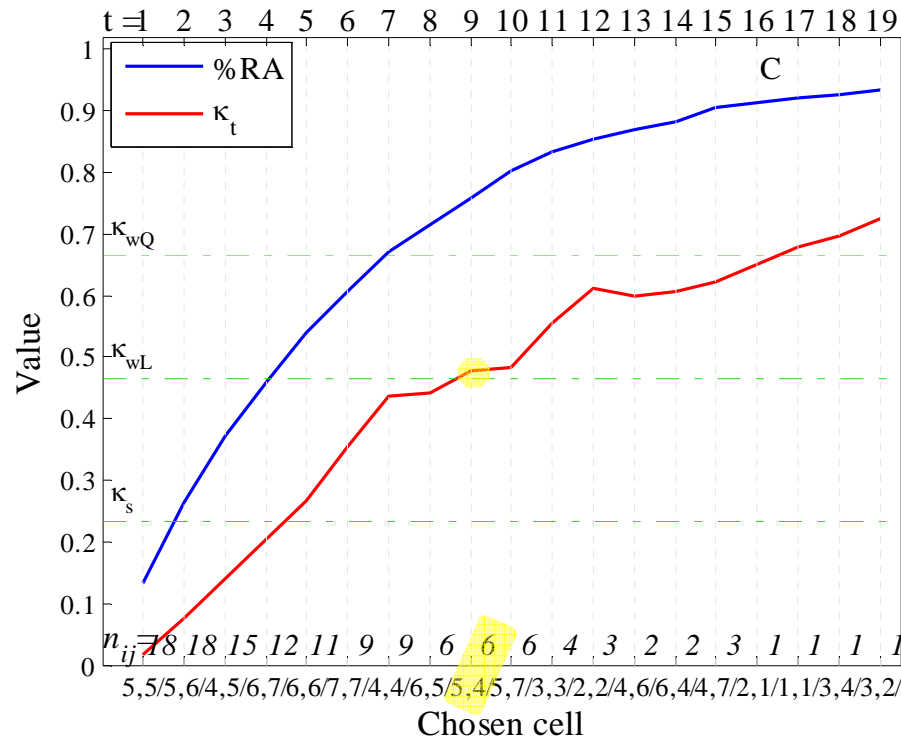
Criterion= Likelihood Ratio



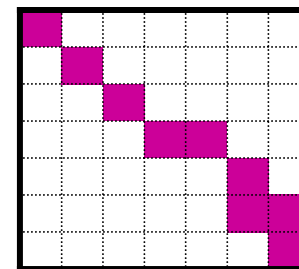
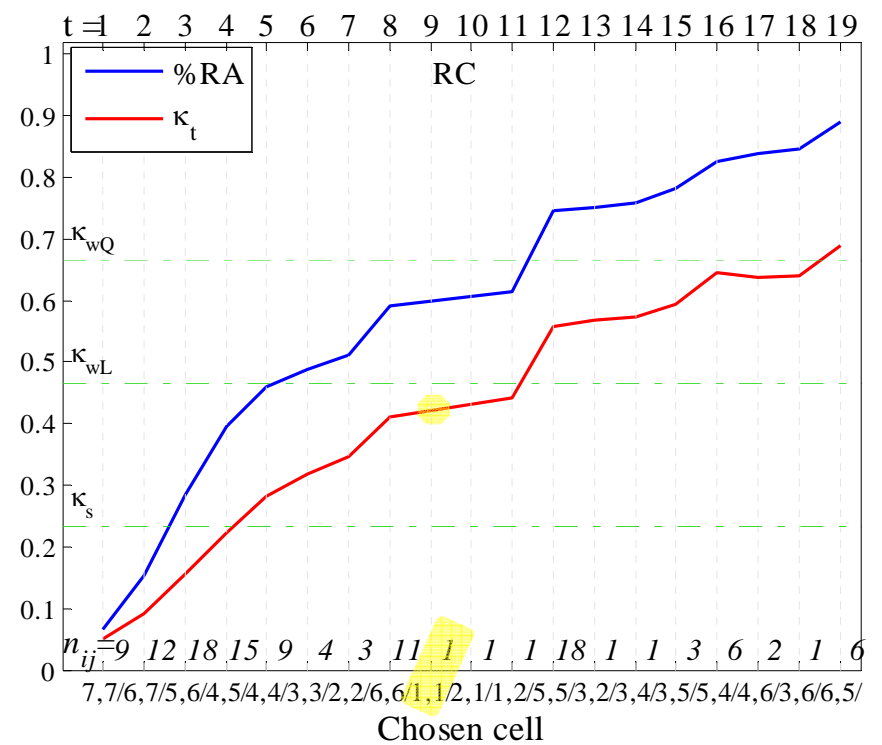


# Results for Rhetorical Ratings

Criterion= Observed



Criterion= Likelihood Ratio



# Conclusions

- The MICA algorithm explores agreement when conventional measures are inadequate- e.g.,
  - the assumption of symmetry is unrealistic, or
  - the emphasis on diagonal agreement is unjustified.
- The algorithm compares different maximization criteria for recovering agreement patterns.
- Recovered agreement patterns are compared to meaningful benchmarks.
- Users can explore the patterns of agreement in their data more effectively.