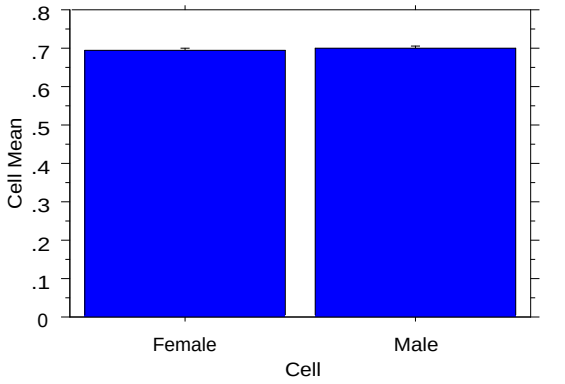
ANOVA Table for ratings
Row exclusion: output_noLLM1 LLM2 data

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Gender	1	.070	.070	1.677	.1955	1.677	.238
Subject(Group)	1510	62.800	.042				
rating type	1	22.692	22.692	712.951	<.0001	712.951	1.000
rating type * Gender	1	.011	.011	.342	.5586	.342	.088
rating type * Subject(Group)	1510	48.060	.032				
problem type	1	5.018	5.018	180.217	<.0001	180.217	1.000
problem type * Gender	1	.019	.019	.695	.4045	.695	.127
problem type * Subject(Group)	1510	42.048	.028				
rating type * problem type	1	.166	.166	11.872	.0006	11.872	.950
rating type * problem type * Gender	1	.002	.002	.148	.7008	.148	.067
rating type * problem type * Subject(Group)	1510	21.067	.014				

Means Table for ratings
Effect: Gender
Row exclusion: output_noLLM1 LLM2 data

	Count	Mean	Std. Dev.	Std. Err.
Female	2336	.694	.184	.004
Male	3712	.701	.183	.003

Interaction Bar Plot for ratings
Effect: Gender
Error Bars: ± 1 Standard Error(s)
Row exclusion: output_noLLM1 LLM2 data



Means Table for ratings
Effect: rating type
Row exclusion: output_noLLM1 LLM2 data

	Count	Mean	Std. Dev.	Std. Err.
confidence	3024	.761	.160	.003
opinion	3024	.636	.183	.003

Means Table for ratings

Effect: problem type
Row exclusion: output_noLLM1 LLM2 data

	Count	Mean	Std. Dev.	Std. Err.
standard	3024	.669	.185	.003
control	3024	.728	.176	.003

Means Table for ratings
Effect: rating type * problem type
Row exclusion: output_noLLM1 LLM2 data

	Count	Mean	Std. Dev.	Std. Err.
confidence, standard	1512	.737	.167	.004
confidence, control	1512	.785	.149	.004
opinion, standard	1512	.601	.177	.005
opinion, control	1512	.671	.183	.005

Means Table for ratings
Effect: rating type * Gender
Row exclusion: output_noLLM1 LLM2 data

	Count	Mean	Std. Dev.	Std. Err.
Female, confidence	1168	.758	.161	.005
Female, opinion	1168	.630	.183	.005
Male, confidence	1856	.763	.160	.004
Male, opinion	1856	.640	.183	.004

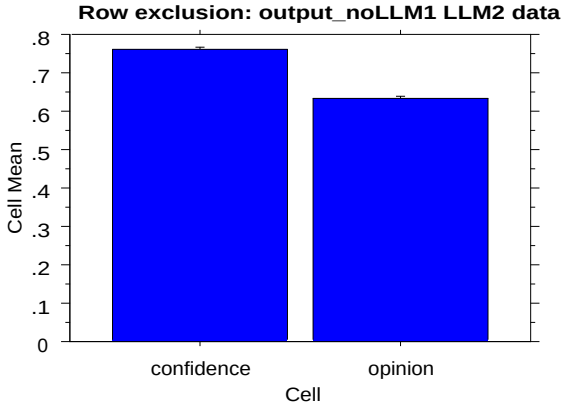
Means Table for ratings
Effect: problem type * Gender
Row exclusion: output_noLLM1 LLM2 data

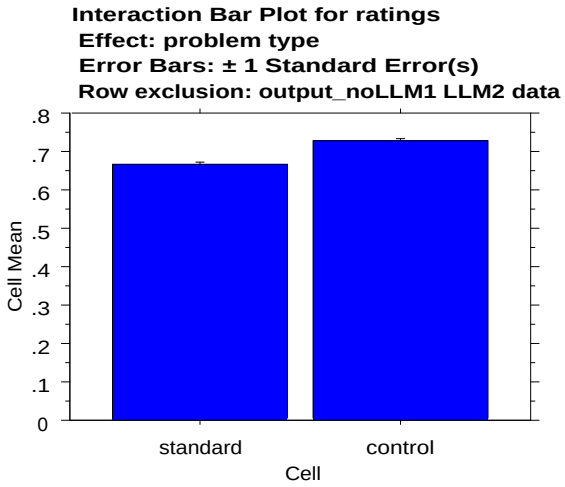
	Count	Mean	Std. Dev.	Std. Err.
Female, standard	1168	.663	.185	.005
Female, control	1168	.726	.177	.005
Male, standard	1856	.673	.186	.004
Male, control	1856	.729	.176	.004

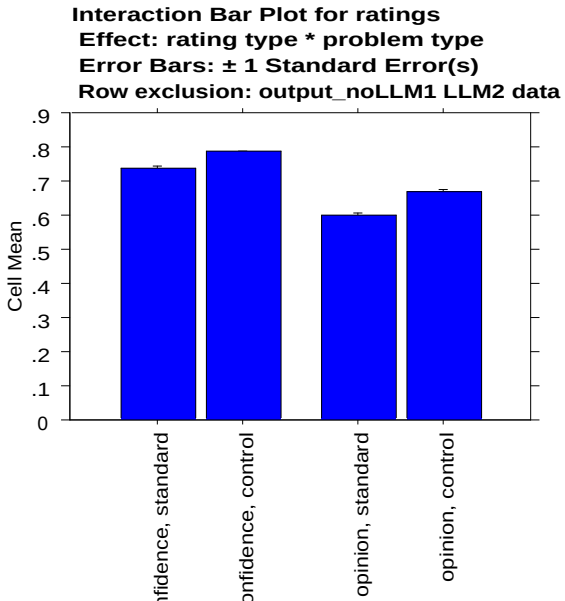
Means Table for ratings
Effect: rating type * problem type * Gender
Row exclusion: output_noLLM1 LLM2 data

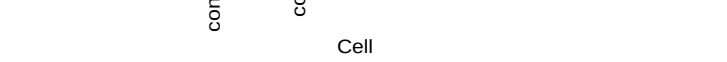
	Count	Mean	Std. Dev.	Std. Err.
Female, confidence, standard	584	.732	.166	.007
Female, confidence, control	584	.785	.150	.006
Female, opinion, standard	584	.594	.177	.007
Female, opinion, control	584	.666	.183	.008
Male, confidence, standard	928	.741	.168	.006
Male, confidence, control	928	.784	.149	.005
Male, opinion, standard	928	.606	.178	.006
Male, opinion, control	928	.673	.183	.006

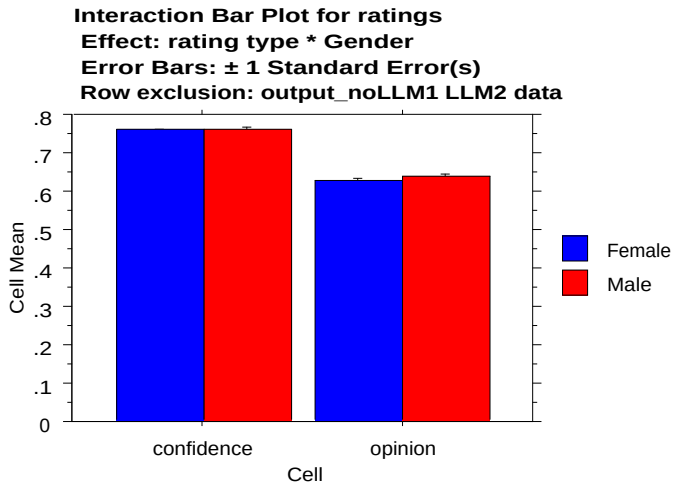
Interaction Bar Plot for ratings
Effect: rating type
Error Bars: ± 1 Standard Error(s)

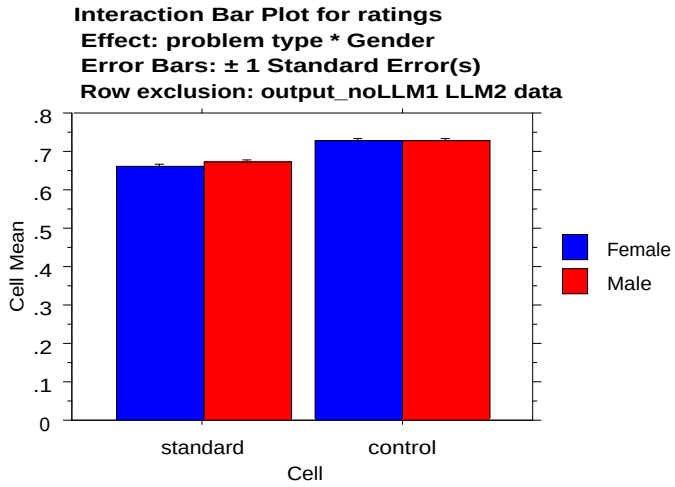

Row exclusion: output_noLLM1 LLM2 data
Cell Mean
.8
.7
.6
.5
.4
.3
.2
.1
0
confidence
opinion
Cell

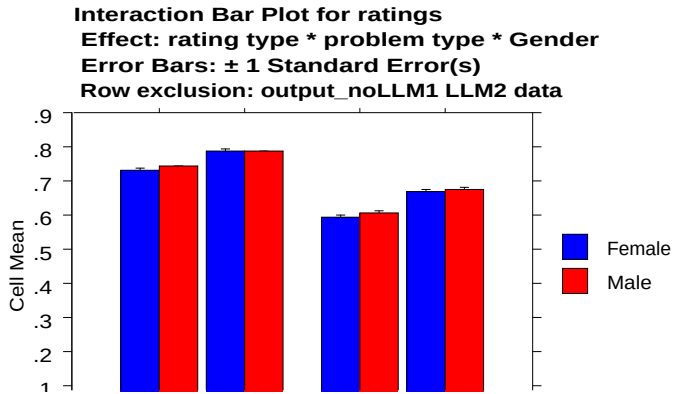

Interaction Bar Plot for ratings
Effect: problem type
Error Bars: ± 1 Standard Error(s)
Row exclusion: output_noLLM1 LLM2 data
Cell Mean
.8
.7
.6
.5
.4
.3
.2
.1
0
standard
control
Cell

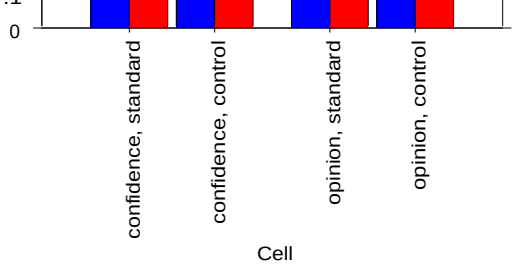

Interaction Bar Plot for ratings
Effect: rating type * problem type
Error Bars: ± 1 Standard Error(s)
Row exclusion: output_noLLM1 LLM2 data
Cell Mean
.9
.8
.7
.6
.5
.4
.3
.2
.1
0
nfidence, standard
onfidence, control
opinion, standard
opinion, control


Cell


Interaction Bar Plot for ratings
Effect: rating type * Gender
Error Bars: ± 1 Standard Error(s)
Row exclusion: output_noLLM1 LLM2 data
Cell Mean
.8
.7
.6
.5
.4
.3
.2
.1
0
confidence
opinion
Cell
Female
Male


Interaction Bar Plot for ratings
Effect: problem type * Gender
Error Bars: ± 1 Standard Error(s)
Row exclusion: output_noLLM1 LLM2 data
Cell Mean
.8
.7
.6
.5
.4
.3
.2
.1
0
standard
control
Cell
Female
Male


Interaction Bar Plot for ratings
Effect: rating type * problem type * Gender
Error Bars: ± 1 Standard Error(s)
Row exclusion: output_noLLM1 LLM2 data
Cell Mean
.9
.8
.7
.6
.5
.4
.3
.2
.1
Female
Male



Fisher's PLSD for ratings
Effect: Gender
Significance Level: 5 %
Row exclusion: output_noLLM1 LLM2 data

	Mean Diff.	Crit. Diff.	P-Value	
Female, Male	-.007	.011	.1955	

Fisher's PLSD for ratings
Effect: rating type
Significance Level: 5 %
Row exclusion: output_noLLM1 LLM2 data

	Mean Diff.	Crit. Diff.	P-Value	
confidence, opinion	.125	.009	<.0001	S

Fisher's PLSD for ratings
Effect: problem type
Significance Level: 5 %
Row exclusion: output_noLLM1 LLM2 data

	Mean Diff.	Crit. Diff.	P-Value	
standard, control	-.058	.008	<.0001	S