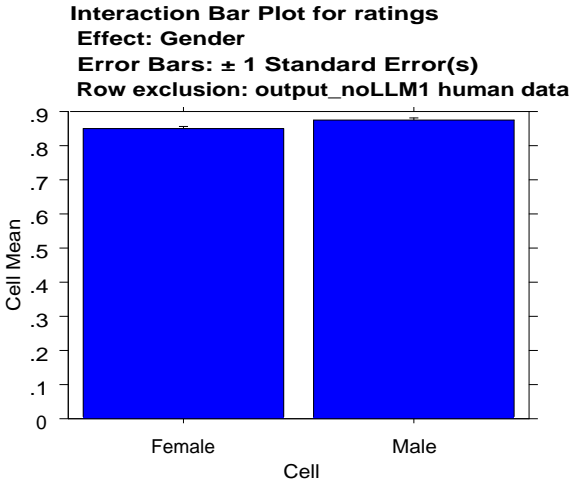


ANOVA Table for ratings
Row exclusion: output_noLLM1 human data

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Gender	1	.285	.285	3.721	.0545	3.721	.471
Subject(Group)	379	29.027	.077				
rating type	1	6.125	6.125	225.248	<.0001	225.248	1.000
rating type * Gender	1	.073	.073	2.682	.1023	2.682	.354
rating type * Subject(Group)	379	10.306	.027				
problem type	1	1.346	1.346	43.231	<.0001	43.231	1.000
problem type * Gender	1	.003	.003	.091	.7633	.091	.060
problem type * Subject(Group)	379	11.798	.031				
rating type * problem type	1	.282	.282	14.653	.0002	14.653	.982
rating type * problem type * Gender	1	3.946E-4	3.946E-4	.020	.8863	.020	.052
rating type * problem type * Subject(Group)	379	7.307	.019				

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

	Count	Mean	Std. Dev.	Std. Err.
Female	588	.847	.216	.009
Male	936	.875	.206	.007



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

	Count	Mean	Std. Dev.	Std. Err.
confidence	762	.931	.184	.007
opinion	762	.797	.214	.008

Means Table for ratings
Effect: problem type
Row exclusion: output_noLLM1 human data

	Count	Mean	Std. Dev.	Std. Err.
standard	762	.834	.225	.008
control	762	.894	.190	.007

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

	Count	Mean	Std. Dev.	Std. Err.
confidence, standard	381	.915	.201	.010
confidence, control	381	.947	.163	.008
opinion, standard	381	.753	.218	.011
opinion, control	381	.841	.200	.010

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

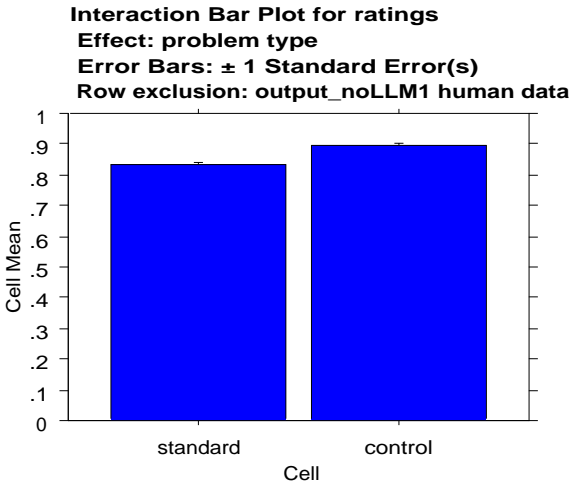
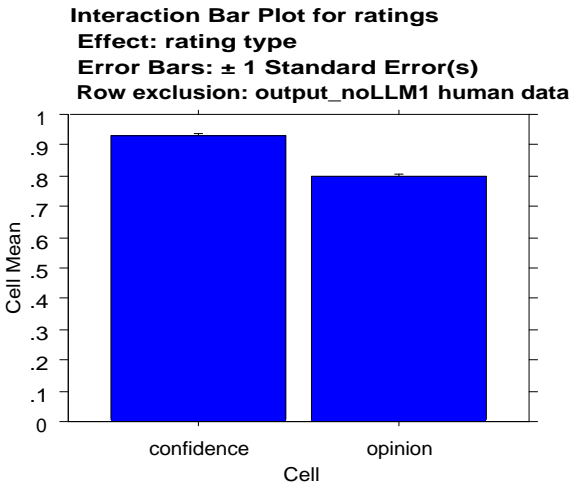
	Count	Mean	Std. Dev.	Std. Err.
Female, confidence	294	.905	.210	.012
Female, opinion	294	.789	.207	.012
Male, confidence	468	.947	.164	.008
Male, opinion	468	.803	.218	.010

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

	Count	Mean	Std. Dev.	Std. Err.
Female, standard	294	.815	.228	.013
Female, control	294	.879	.199	.012
Male, standard	468	.846	.222	.010
Male, control	468	.904	.184	.009

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

	Count	Mean	Std. Dev.	Std. Err.
Female, confidence, standard	147	.887	.222	.018
Female, confidence, control	147	.922	.196	.016
Female, opinion, standard	147	.742	.211	.017
Female, opinion, control	147	.835	.192	.016
Male, confidence, standard	234	.932	.186	.012
Male, confidence, control	234	.963	.137	.009
Male, opinion, standard	234	.760	.222	.015
Male, opinion, control	234	.845	.206	.013

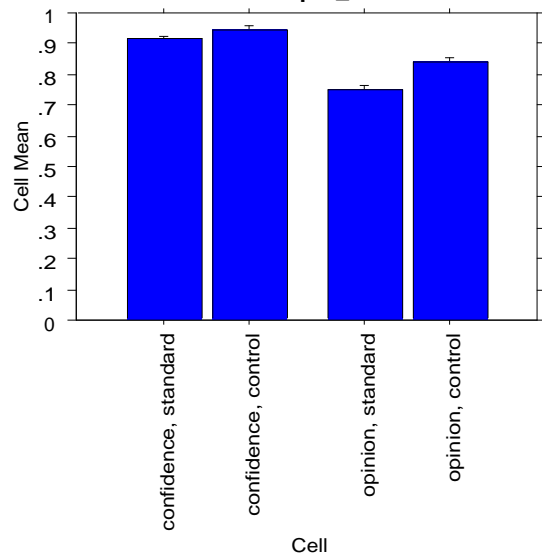


Interaction Bar Plot for ratings

Effect: rating type * problem type

Error Bars: ± 1 Standard Error(s)

Row exclusion: output_noLLM1 human data

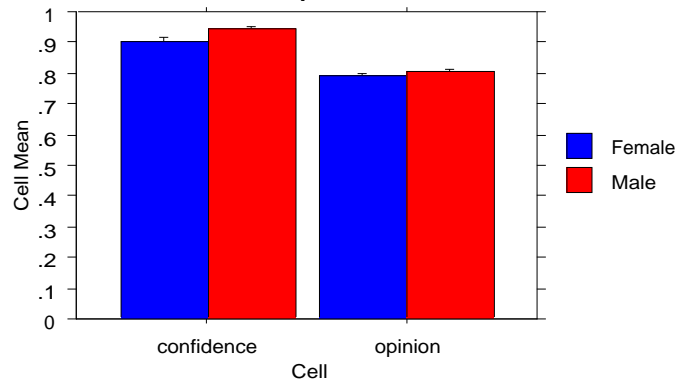


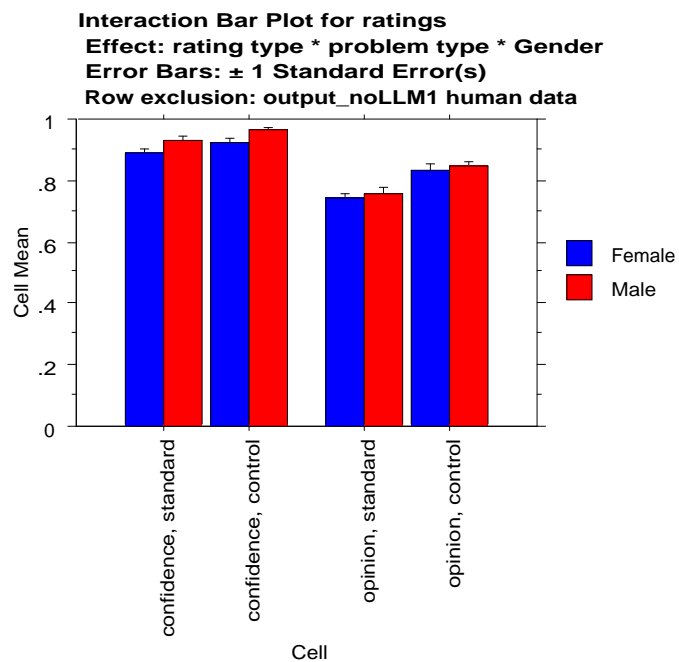
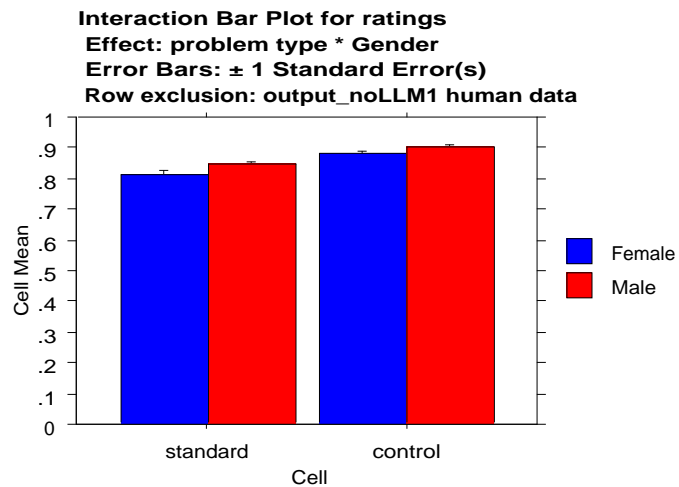
Interaction Bar Plot for ratings

Effect: rating type * Gender

Error Bars: ± 1 Standard Error(s)

Row exclusion: output_noLLM1 human data





Fisher's PLSD for ratings

Effect: Gender

Significance Level: 5 %

Row exclusion: output_noLLM1 human data

	Mean Diff.	Crit. Diff.	P-Value
Female, Male	-.028	.029	.0545

Fisher's PLSD for ratings

Effect: rating type

Significance Level: 5 %

Row exclusion: output_noLLM1 human data

	Mean Diff.	Crit. Diff.	P-Value	
confidence, opinion	.133	.017	<.0001	S

Fisher's PLSD for ratings

Effect: problem type

Significance Level: 5 %

Row exclusion: output_noLLM1 human data

	Mean Diff.	Crit. Diff.	P-Value	
standard, control	-.060	.018	<.0001	S