ANOVA Table for ratings
Row exclusion: output_all_MIX.HU dat3

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Group	1	7.448	7.448	146.280	<.0001	146.280	1.000
Gender	1	.326	.326	6.402	.0115	6.402	.720
Group * Gender	1	.145	.145	2.842	.0920	2.842	.374
Subject(Group)	1845	93.935	.051				
rating type	1	37.958	37.958	1513.947	<.0001	1513.947	1.000
rating type * Group	1	3.065	3.065	122.251	<.0001	122.251	1.000
rating type * Gender	1	.102	.102	4.065	.0439	4.065	.509
rating type * Group * Gender	1	.026	.026	1.047	.3063	1.047	.166
rating type * Subject(Group)	1845	46.259	.025				
problem type	1	7.935	7.935	262.180	<.0001	262.180	1.000
problem type * Group	1	.562	.562	18.570	<.0001	18.570	.997
problem type * Gender	1	.011	.011	.361	.5480	.361	.090
problem type * Group * Gender	1	9.576E-5	9.576E-5	.003	.9552	.003	.050
problem type * Subject(Group)	1845	55.843	.030				
rating type * problem type	1	1.614	1.614	103.977	<.0001	103.977	1.000
rating type * problem type * Group	1	.104	.104	6.732	.0095	6.732	.744
rating type * problem type * Gender	1	.003	.003	.214	.6438	.214	.074
rating type * problem type * Group * Gender	1	.009	.009	.557	.4554	.557	.112
rating type * problem type * Subject(Group)	1845	28.638	.016				

Means Table for ratings
Effect: Group
Row exclusion: output_all_MIX.HU dat3

	Count	Mean	Std. Dev.	Std. Err.
Human	1524	.864	.210	.005
LLM_MIX8X7B_T13	5872	.781	.213	.003

Interaction Bar Plot for ratings
Effect: Group
Error Bars: ± 1 Standard Error(s)
Row exclusion: output_all_MIX.HU dat3
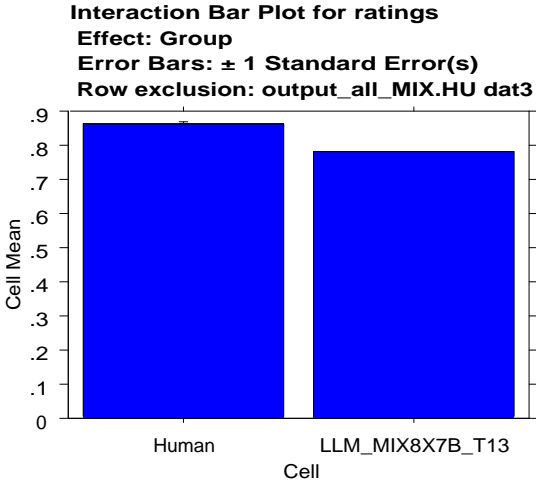
Means Table for ratings
Effect: Gender
Row exclusion: output_all_MIX.HU dat3

	Count	Mean	Std. Dev.	Std. Err.
Female	2852	.792	.215	.004
Male	4544	.802	.215	.003

Means Table for ratings
Effect: Group * Gender
Row exclusion: output_all_MIX.HU dat3

	Count	Mean	Std. Dev.	Std. Err.
Human, Female	588	.847	.216	.009
Human, Male	936	.875	.206	.007
LLM_MIX8X7B_T13, Female	2264	.778	.213	.004
LLM_MIX8X7B_T13, Male	3608	.783	.213	.004

Means Table for ratings
Effect: rating type
Row exclusion: output_all_MIX.HU dat3

	Count	Mean	Std. Dev.	Std. Err.
confidence	3698	.905	.152	.002
opinon	3698	.691	.216	.004

Means Table for ratings
Effect: problem type
Row exclusion: output_all_MIX.HU dat3

	Count	Mean	Std. Dev.	Std. Err.
standard	3698	.750	.233	.004
control	3698	.846	.184	.003

Means Table for ratings
Effect: rating type * problem type
Row exclusion: output_all_MIX.HU dat3

	Count	Mean	Std. Dev.	Std. Err.
confidence, standard	1849	.879	.166	.004
confidence, control	1849	.931	.130	.003
opinon, standard	1849	.622	.218	.005
opinon, control	1849	.761	.191	.004

Means Table for ratings
Effect: rating type * Group
Row exclusion: output_all_MIX.HU dat3

	Count	Mean	Std. Dev.	Std. Err.
Human, confidence	762	.931	.184	.007
Human, opinon	762	.797	.214	.008
LLM_MIX8X7B_T13, confidence	2936	.898	.141	.003
LLM_MIX8X7B_T13, opinon	2936	.664	.208	.004

Means Table for ratings
Effect: problem type * Group
Row exclusion: output_all_MIX.HU dat3

	Count	Mean	Std. Dev.	Std. Err.
Human, standard	762	.834	.225	.008
Human, control	762	.894	.190	.007
LLM_MIX8X7B_T13, standard	2936	.729	.230	.004
LLM_MIX8X7B_T13, control	2936	.833	.181	.003

Means Table for ratings
Effect: rating type * problem type * Group
Row exclusion: output_all_MIX.HU dat3

	Count	Mean	Std. Dev.	Std. Err.
Human, confidence, standard	381	.915	.201	.010
Human, confidence, control	381	.947	.163	.008
Human, opinon, standard	381	.753	.218	.011
Human, opinon, control	381	.841	.200	.010
LLM_MIX8X7B_T13, confidence, standard	1468	.870	.155	.004
LLM_MIX8X7B_T13, confidence, control	1468	.927	.120	.003
LLM_MIX8X7B_T13, opinon, standard	1468	.587	.204	.005
LLM_MIX8X7B_T13, opinon, control	1468	.740	.183	.005

Means Table for ratings
Effect: rating type * Gender
Row exclusion: output_all_MIX.HU dat3

	Count	Mean	Std. Dev.	Std. Err.
Female, confidence	1426	.895	.162	.004
Female, opinon	1426	.689	.213	.006
Male, confidence	2272	.912	.144	.003
Male, opinon	2272	.693	.218	.005

Means Table for ratings
Effect: problem type * Gender
Row exclusion: output_all_MIX.HU dat3

	Count	Mean	Std. Dev.	Std. Err.
Female, standard	1426	.742	.230	.006
Female, control	1426	.842	.187	.005
Male, standard	2272	.756	.234	.005
Male, control	2272	.849	.183	.004

Means Table for ratings
Effect: rating type * problem type * Gender
Row exclusion: output_all_MIX.HU dat3

	Count	Mean	Std. Dev.	Std. Err.
Female, confidence, standard	713	.865	.179	.007
Female, confidence, control	713	.925	.136	.005
Female, opinon, standard	713	.619	.209	.008
Female, opinon, control	713	.759	.194	.007
Male, confidence, standard	1136	.888	.157	.005
Male, confidence, control	1136	.935	.127	.004
Male, opinon, standard	1136	.623	.223	.007
Male, opinon, control	1136	.762	.189	.006

Means Table for ratings
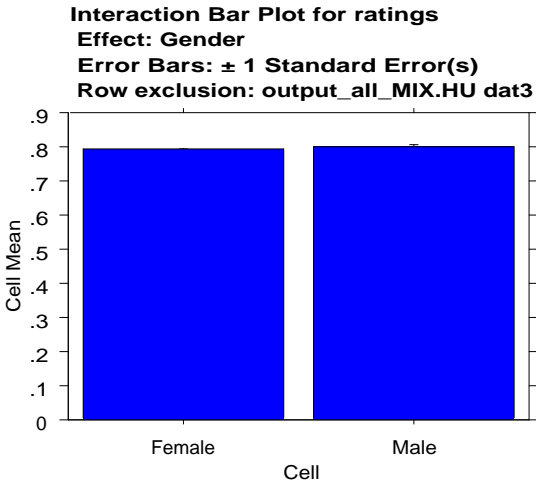**Effect: rating type * Group * Gender**
Row exclusion: output_all_MIX.HU dat3

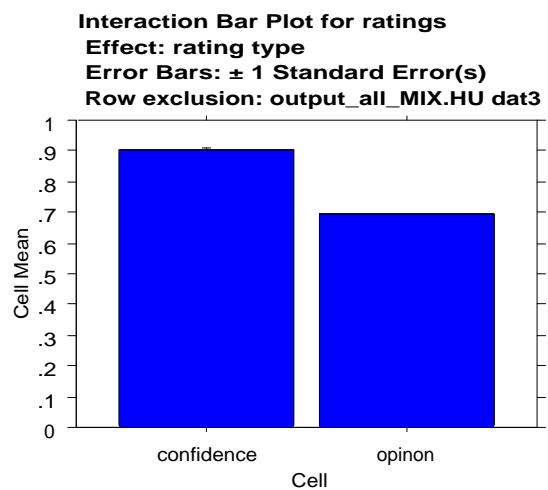|	Count	Mean	Std. Dev.	Std. Err.
Human, Female, confidence	294	.905	.210	.012
Human, Female, opinon	294	.789	.207	.012
Human, Male, confidence	468	.947	.164	.008
Human, Male, opinon	468	.803	.218	.010
LLM_MIX8X7B_T13, Female, confidence	1132	.892	.147	.004
LLM_MIX8X7B_T13, Female, opinon	1132	.663	.207	.006
LLM_MIX8X7B_T13, Male, confidence	1804	.902	.138	.003
LLM_MIX8X7B_T13, Male, opinon	1804	.664	.209	.005

Means Table for ratings
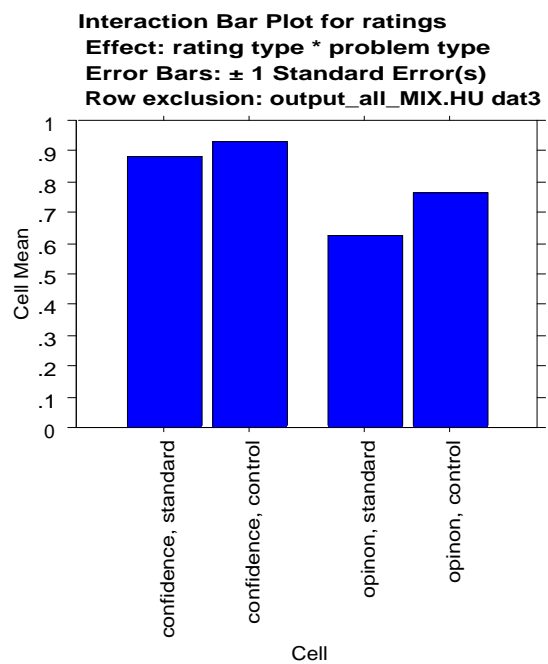**Effect: problem type * Group * Gender**
Row exclusion: output_all_MIX.HU dat3

	Count	Mean	Std. Dev.	Std. Err.
Human, Female, standard	294	.815	.228	.013
Human, Female, control	294	.879	.199	.012
Human, Male, standard	468	.846	.222	.010
Human, Male, control	468	.904	.184	.009
LLM_MIX8X7B_T13, Female, standard	1132	.723	.227	.007
LLM_MIX8X7B_T13, Female, control	1132	.832	.182	.005
LLM_MIX8X7B_T13, Male, standard	1804	.732	.232	.005
LLM_MIX8X7B_T13, Male, control	1804	.834	.179	.004

Means Table for ratings
Effect: rating type * problem type * Group * Gender
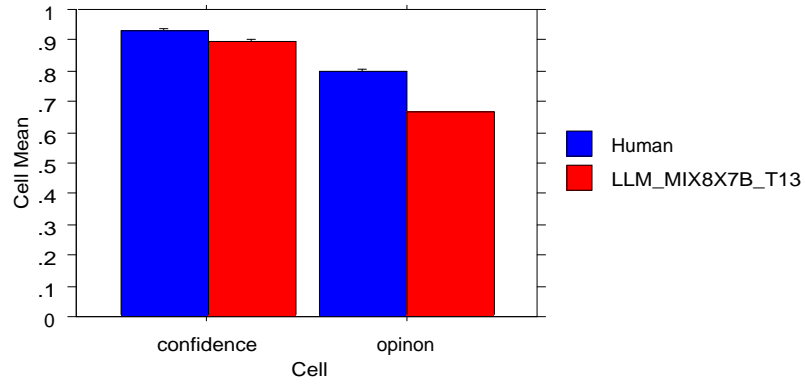**Row exclusion: output_all_MIX.HU dat3**

	Count	Mean	Std. Dev.	Std. Err.
Human, Female, confidence, standard	147	.887	.222	.018
Human, Female, confidence, control	147	.922	.196	.016
Human, Female, opinon, standard	147	.742	.211	.017
Human, Female, opinon, control	147	.835	.192	.016
Human, Male, confidence, standard	234	.932	.186	.012
Human, Male, confidence, control	234	.963	.137	.009
Human, Male, opinon, standard	234	.760	.222	.015
Human, Male, opinon, control	234	.845	.206	.013
LLM_MIX8X7B_T13, Female, confidence, st…	566	.859	.166	.007
LLM_MIX8X7B_T13, Female, confidence, co…	566	.925	.116	.005
LLM_MIX8X7B_T13, Female, opinon, standa…	566	.587	.197	.008
LLM_MIX8X7B_T13, Female, opinon, control	566	.739	.189	.008
LLM_MIX8X7B_T13, Male, confidence, stan…	902	.877	.147	.005
LLM_MIX8X7B_T13, Male, confidence, cont…	902	.928	.123	.004
LLM_MIX8X7B_T13, Male, opinon, standard	902	.587	.209	.007
LLM_MIX8X7B_T13, Male, opinon, control	902	.741	.178	.006

Interaction Bar Plot for ratings
Effect: Gender
Error Bars: ± 1 Standard Error(s)
Row exclusion: output_all_MIX.HU dat3

Interaction Bar Plot for ratings
Effect: Group * Gender
Error Bars: ± 1 Standard Error(s)
Row exclusion: output_all_MIX.HU dat3

Interaction Bar Plot for ratings
Effect: rating type
Error Bars: ± 1 Standard Error(s)
Row exclusion: output_all_MIX.HU dat3

Interaction Bar Plot for ratings
Effect: problem type
Error Bars: ± 1 Standard Error(s)
Row exclusion: output_all_MIX.HU dat3

Interaction Bar Plot for ratings
Effect: rating type * problem type
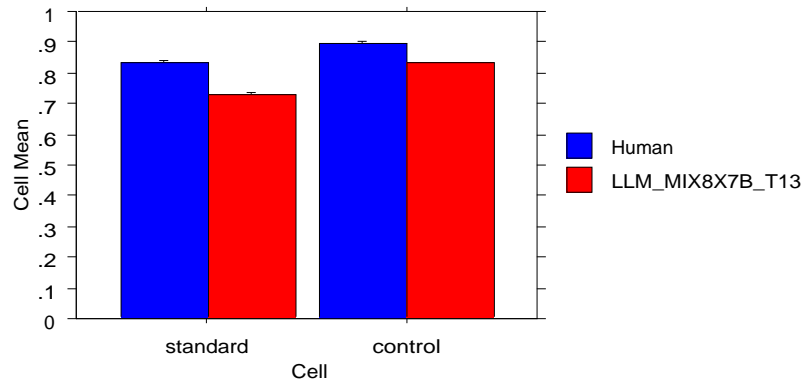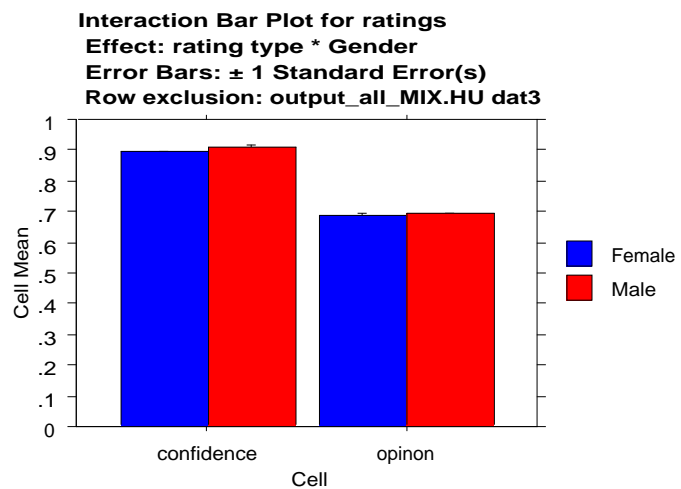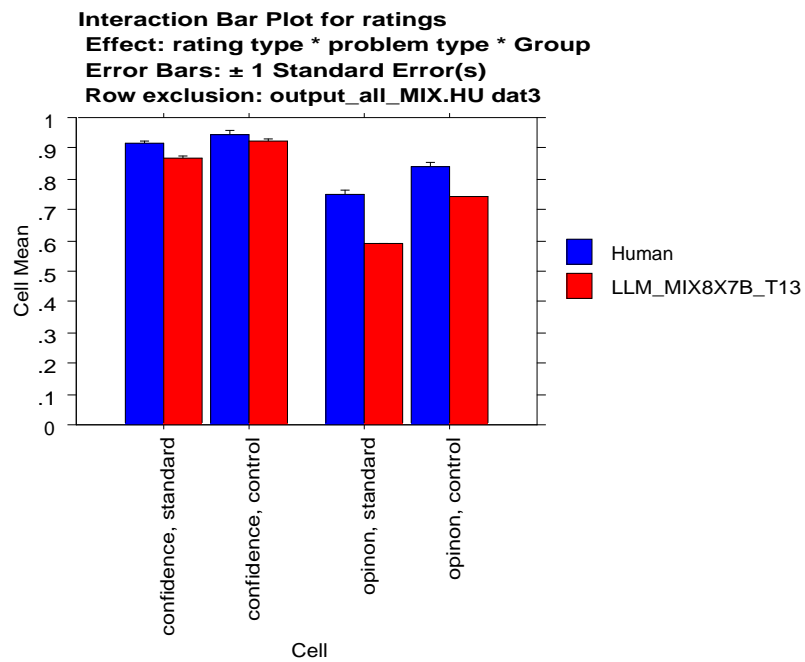**Error Bars: ± 1 Standard Error(s)**
Row exclusion: output_all_MIX.HU dat3

Interaction Bar Plot for ratings
Effect: rating type * Group
Error Bars: ± 1 Standard Error(s)
Row exclusion: output_all_MIX.HU dat3
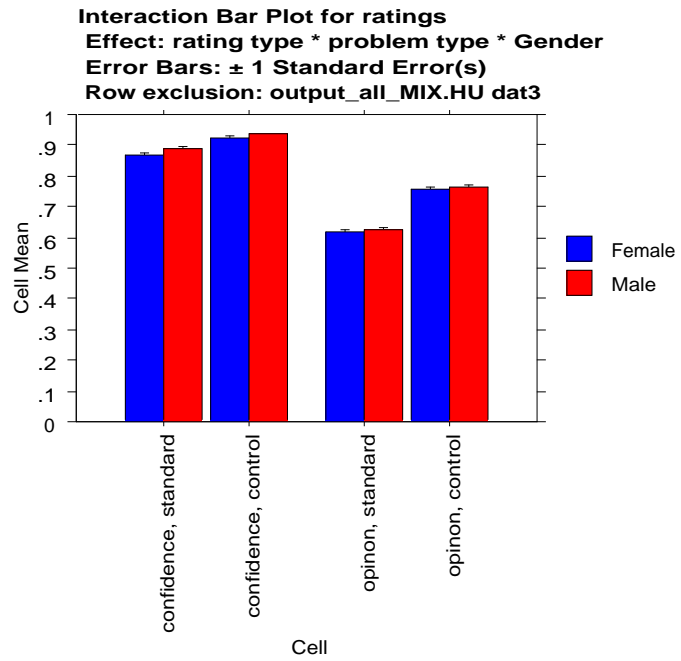
Cell Mean

Human
LLM_MIX8X7B_T13

confidence   opinon

Cell

Interaction Bar Plot for ratings
Effect: problem type * Group
Error Bars: ± 1 Standard Error(s)
Row exclusion: output_all_MIX.HU dat3

Cell Mean

Human
LLM_MIX8X7B_T13

standard   control

Cell

Interaction Bar Plot for ratings
 Effect: rating type * problem type * Group
 Error Bars: ± 1 Standard Error(s)
 Row exclusion: output_all_MIX.HU dat3

Interaction Bar Plot for ratings
 Effect: rating type * Gender
 Error Bars: ± 1 Standard Error(s)
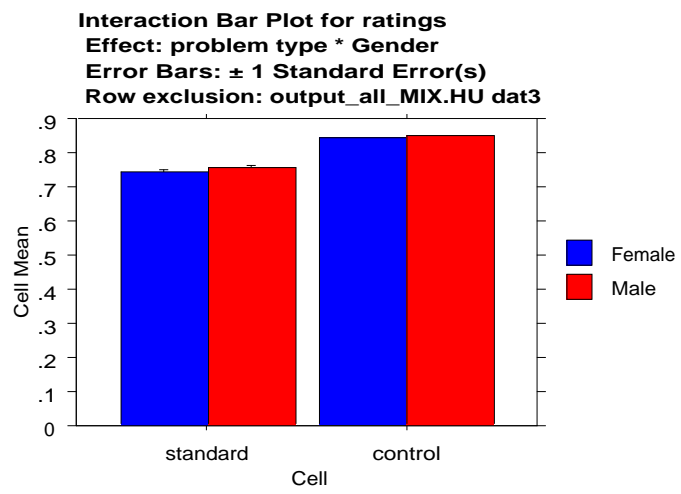 **Row exclusion: output_all_MIX.HU dat3**

Interaction Bar Plot for ratings
 Effect: problem type * Gender
 Error Bars: ± 1 Standard Error(s)
 Row exclusion: output_all_MIX.HU dat3

Interaction Bar Plot for ratings
 Effect: rating type * problem type * Gender
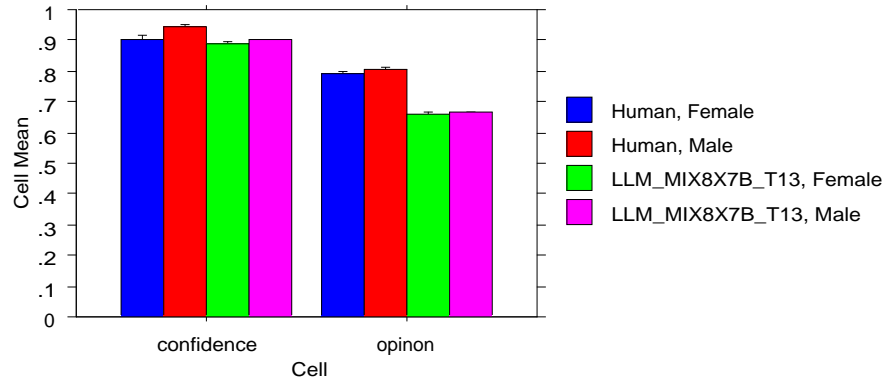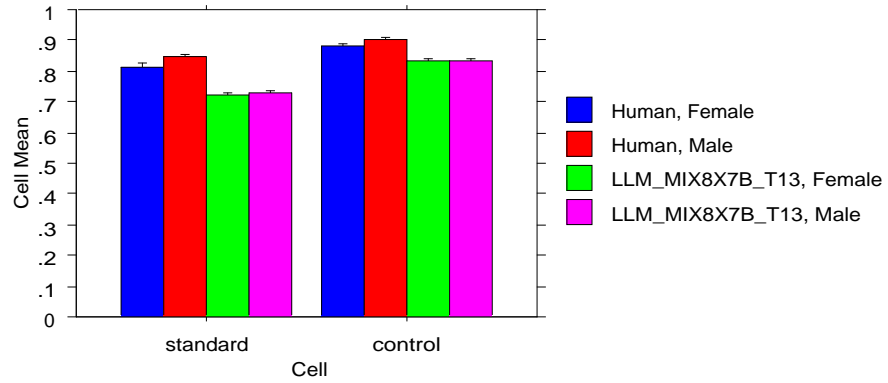 **Error Bars: ± 1 Standard Error(s)**
 Row exclusion: output_all_MIX.HU dat3

Interaction Bar Plot for ratings
Effect: rating type * Group * Gender
Error Bars: ± 1 Standard Error(s)
Row exclusion: output_all_MIX.HU dat3

Legend:
- Human, Female
- Human, Male
- LLM_MIX8X7B_T13, Female
- LLM_MIX8X7B_T13, Male

x-axis: Cell (confidence, opinon)
y-axis: Cell Mean

Interaction Bar Plot for ratings
Effect: problem type * Group * Gender
Error Bars: ± 1 Standard Error(s)
Row exclusion: output_all_MIX.HU dat3

Legend:
- Human, Female
- Human, Male
- LLM_MIX8X7B_T13, Female
- LLM_MIX8X7B_T13, Male

x-axis: Cell (standard, control)
y-axis: Cell Mean

Interaction Bar Plot for ratings
 Effect: rating type * problem type * Group * Gender
 Error Bars: ± 1 Standard Error(s)
 Row exclusion: output_all_MIX.HU dat3