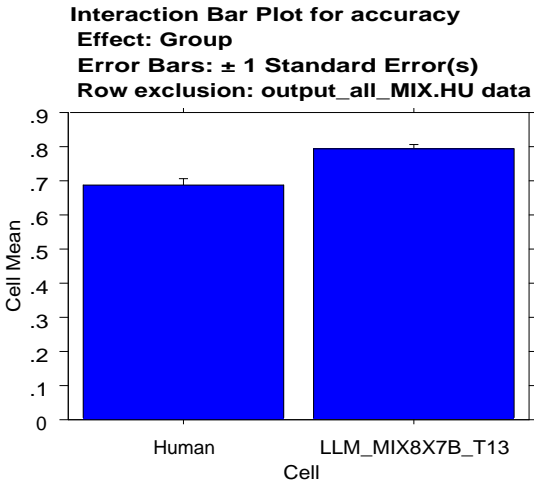


ANOVA Table for accuracy
Row exclusion: output_all_MIX.HU data

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Group	1	7.223	7.223	54.632	<.0001	54.632	1.000
CBType	3	3.058	1.019	7.709	<.0001	23.128	.994
Gender	1	.195	.195	1.477	.2244	1.477	.215
Group * CBType	3	2.047	.682	5.160	.0015	15.480	.936
Group * Gender	1	.046	.046	.344	.5575	.344	.088
CBType * Gender	3	.517	.172	1.304	.2716	3.911	.340
Group * CBType * Gender	3	.550	.183	1.386	.2452	4.159	.360
Subject(Group)	1889	249.761	.132				
problem type	1	116.647	116.647	925.324	<.0001	925.324	1.000
problem type * Group	1	6.323	6.323	50.160	<.0001	50.160	1.000
problem type * CBType	3	.656	.219	1.734	.1581	5.201	.445
problem type * Gender	1	.026	.026	.209	.6479	.209	.073
problem type * Group * CBType	3	2.590	.863	6.849	.0001	20.548	.986
problem type * Group * Gender	1	5.966E-5	5.966E-5	4.733E-4	.9826	4.733E-4	.050
problem type * CBType * Gender	3	.398	.133	1.052	.3684	3.156	.279
problem type * Group * CBType * Gender	3	.475	.158	1.255	.2882	3.766	.328
problem type * Subject(Group)	1889	238.129	.126				

Means Table for accuracy
Effect: Group
Row exclusion: output_all_MIX.HU data

	Count	Mean	Std. Dev.	Std. Err.
Human	762	.686	.464	.017
LLM_MIX8X7B_T13	3048	.797	.403	.007



Means Table for accuracy
Effect: CBType
Row exclusion: output_all_MIX.HU data

	Count	Mean	Std. Dev.	Std. Err.
1a	990	.758	.429	.014
1b	960	.741	.439	.014
2a	930	.753	.432	.014
2b	930	.849	.358	.012

Means Table for accuracy
Effect: Gender
Row exclusion: output_all_MIX.HU data

	Count	Mean	Std. Dev.	Std. Err.
Female	1470	.769	.422	.011
Male	2340	.778	.416	.009

Means Table for accuracy
Effect: Group * CBType
Row exclusion: output_all_MIX.HU data

	Count	Mean	Std. Dev.	Std. Err.
Human, 1a	198	.712	.454	.032
Human, 1b	192	.698	.460	.033
Human, 2a	186	.618	.487	.036
Human, 2b	186	.715	.453	.033
LLM_MIX8X7B_T13, 1a	792	.769	.422	.015
LLM_MIX8X7B_T13, 1b	768	.751	.433	.016
LLM_MIX8X7B_T13, 2a	744	.786	.410	.015
LLM_MIX8X7B_T13, 2b	744	.883	.322	.012

Means Table for accuracy
Effect: Group * Gender
Row exclusion: output_all_MIX.HU data

	Count	Mean	Std. Dev.	Std. Err.
Human, Female	294	.667	.472	.028
Human, Male	468	.699	.459	.021
LLM_MIX8X7B_T13, Female	1176	.794	.404	.012
LLM_MIX8X7B_T13, Male	1872	.798	.402	.009

Means Table for accuracy
Effect: CBType * Gender
Row exclusion: output_all_MIX.HU data

	Count	Mean	Std. Dev.	Std. Err.
1a, Female	310	.777	.417	.024
1a, Male	680	.749	.434	.017
1b, Female	360	.725	.447	.024
1b, Male	600	.750	.433	.018
2a, Female	400	.720	.450	.022
2a, Male	530	.777	.416	.018
2b, Female	400	.850	.358	.018
2b, Male	530	.849	.358	.016

Means Table for accuracy
Effect: Group * CBType * Gender
Row exclusion: output_all_MIX.HU data

	Count	Mean	Std. Dev.	Std. Err.
Human, 1a, Female	62	.726	.450	.057
Human, 1a, Male	136	.706	.457	.039
Human, 1b, Female	72	.708	.458	.054
Human, 1b, Male	120	.692	.464	.042
Human, 2a, Female	80	.588	.495	.055
Human, 2a, Male	106	.642	.482	.047
Human, 2b, Female	80	.662	.476	.053
Human, 2b, Male	106	.755	.432	.042
LLM_MIX8X7B_T13, 1a, Female	248	.790	.408	.026
LLM_MIX8X7B_T13, 1a, Male	544	.759	.428	.018
LLM_MIX8X7B_T13, 1b, Female	288	.729	.445	.026
LLM_MIX8X7B_T13, 1b, Male	480	.765	.425	.019
LLM_MIX8X7B_T13, 2a, Female	320	.753	.432	.024
LLM_MIX8X7B_T13, 2a, Male	424	.811	.392	.019
LLM_MIX8X7B_T13, 2b, Female	320	.897	.305	.017
LLM_MIX8X7B_T13, 2b, Male	424	.873	.334	.016

Means Table for accuracy
Effect: problem type
Row exclusion: output_all_MIX.HU data

	Count	Mean	Std. Dev.	Std. Err.
standard	1905	.580	.494	.011
control	1905	.970	.172	.004

Means Table for accuracy
Effect: problem type * Group
Row exclusion: output_all_MIX.HU data

	Count	Mean	Std. Dev.	Std. Err.
Human, standard	381	.407	.492	.025
Human, control	381	.966	.182	.009
LLM_MIX8X7B_T13, standard	1524	.623	.485	.012
LLM_MIX8X7B_T13, control	1524	.970	.169	.004

Means Table for accuracy
Effect: problem type * CBType
Row exclusion: output_all_MIX.HU data

	Count	Mean	Std. Dev.	Std. Err.
1a, standard	495	.535	.499	.022
1a, control	495	.980	.141	.006
1b, standard	480	.510	.500	.023
1b, control	480	.971	.168	.008
2a, standard	465	.576	.495	.023
2a, control	465	.929	.257	.012
2b, standard	465	.701	.458	.021
2b, control	465	.998	.046	.002

Means Table for accuracy
Effect: problem type * Gender
Row exclusion: output_all_MIX.HU data

	Count	Mean	Std. Dev.	Std. Err.
Female, standard	735	.573	.495	.018
Female, control	735	.965	.185	.007
Male, standard	1170	.584	.493	.014
Male, control	1170	.973	.163	.005

Means Table for accuracy
Effect: problem type * Group * CBType
Row exclusion: output_all_MIX.HU data

	Count	Mean	Std. Dev.	Std. Err.
Human, 1a, standard	99	.434	.498	.050
Human, 1a, control	99	.990	.101	.010
Human, 1b, standard	96	.448	.500	.051
Human, 1b, control	96	.948	.223	.023
Human, 2a, standard	93	.301	.461	.048
Human, 2a, control	93	.935	.247	.026
Human, 2b, standard	93	.441	.499	.052
Human, 2b, control	93	.989	.104	.011
LLM_MIX8X7B_T13, 1a, standard	396	.561	.497	.025
LLM_MIX8X7B_T13, 1a, control	396	.977	.149	.007
LLM_MIX8X7B_T13, 1b, standard	384	.526	.500	.026
LLM_MIX8X7B_T13, 1b, control	384	.977	.151	.008
LLM_MIX8X7B_T13, 2a, standard	372	.645	.479	.025
LLM_MIX8X7B_T13, 2a, control	372	.927	.260	.013
LLM_MIX8X7B_T13, 2b, standard	372	.766	.424	.022
LLM_MIX8X7B_T13, 2b, control	372	1.000	0.000	0.000

Means Table for accuracy
Effect: problem type * Group * Gender
Row exclusion: output_all_MIX.HU data

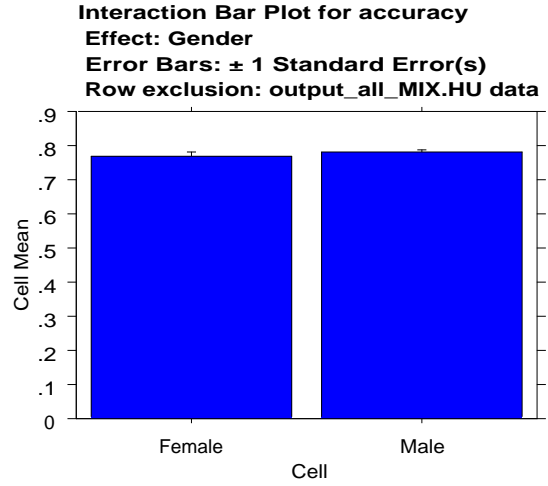
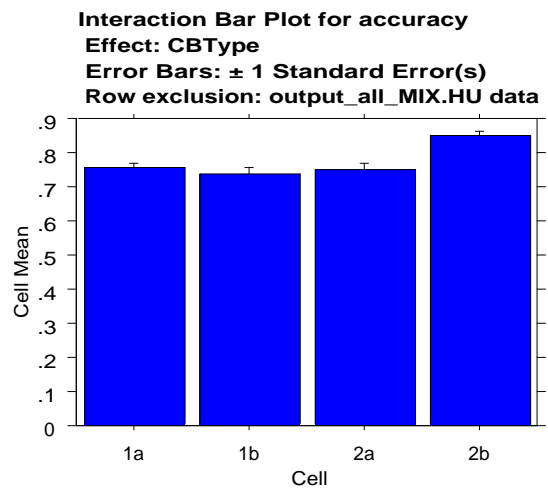
	Count	Mean	Std. Dev.	Std. Err.
Human, Female, standard	147	.381	.487	.040
Human, Female, control	147	.952	.214	.018
Human, Male, standard	234	.423	.495	.032
Human, Male, control	234	.974	.158	.010
LLM_MIX8X7B_T13, Female, standard	588	.621	.486	.020
LLM_MIX8X7B_T13, Female, control	588	.968	.177	.007
LLM_MIX8X7B_T13, Male, standard	936	.624	.485	.016
LLM_MIX8X7B_T13, Male, control	936	.972	.164	.005

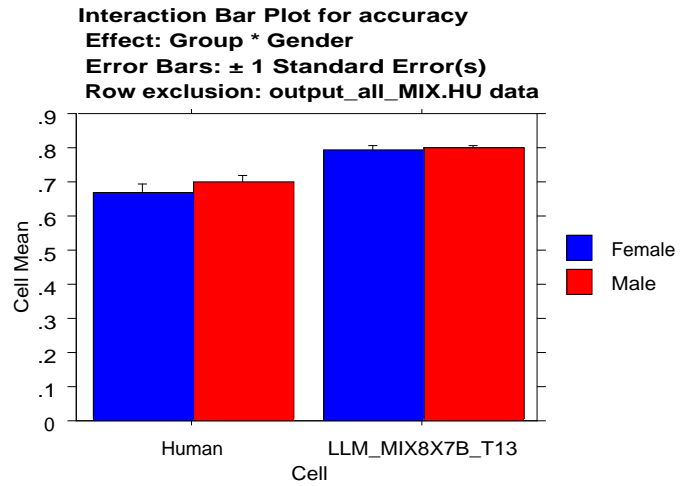
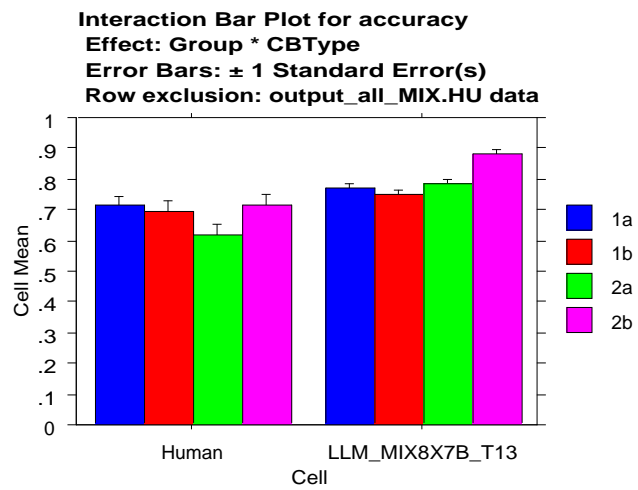
Means Table for accuracy
Effect: problem type * CBType * Gender
Row exclusion: output_all_MIX.HU data

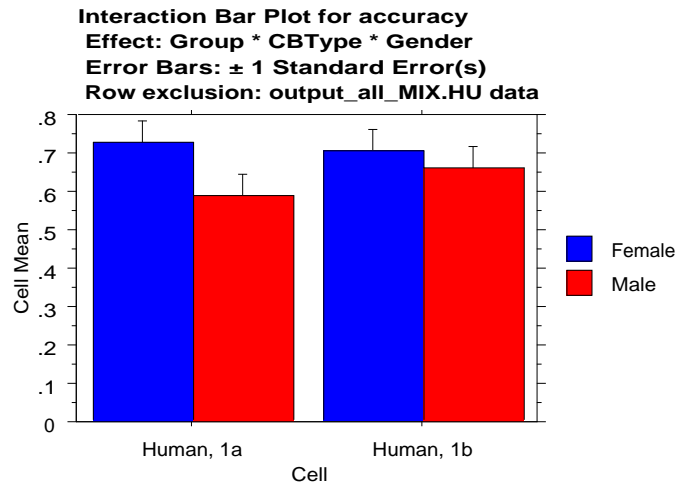
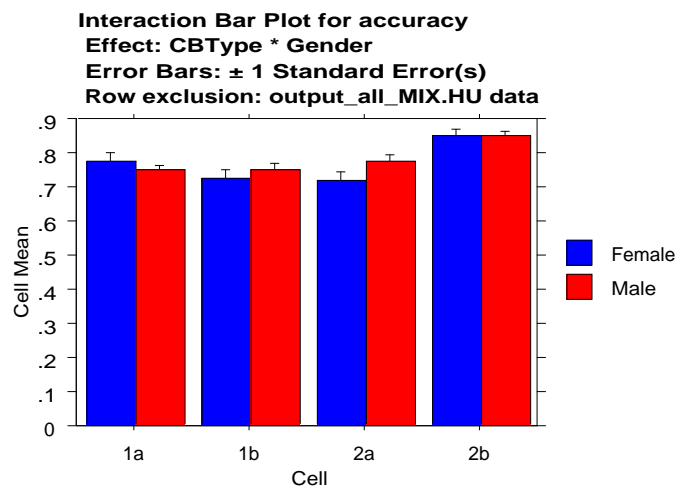
	Count	Mean	Std. Dev.	Std. Err.
1a, Female, standard	155	.574	.496	.040
1a, Female, control	155	.981	.138	.011
1a, Male, standard	340	.518	.500	.027
1a, Male, control	340	.979	.142	.008
1b, Female, standard	180	.478	.501	.037
1b, Female, control	180	.972	.165	.012
1b, Male, standard	300	.530	.500	.029
1b, Male, control	300	.970	.171	.010
2a, Female, standard	200	.525	.501	.035
2a, Female, control	200	.915	.280	.020
2a, Male, standard	265	.615	.487	.030
2a, Male, control	265	.940	.239	.015
2b, Female, standard	200	.705	.457	.032
2b, Female, control	200	.995	.071	.005
2b, Male, standard	265	.698	.460	.028
2b, Male, control	265	1.000	0.000	0.000

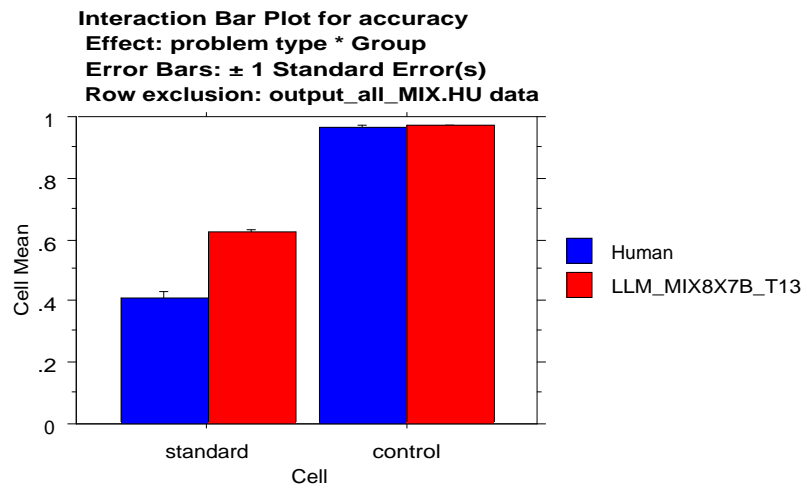
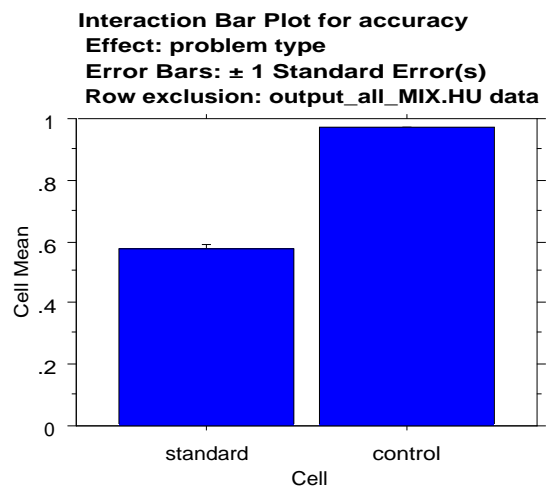
Means Table for accuracy**Effect: problem type * Group * CBType * Gender****Row exclusion: output_all_MIX.HU data**

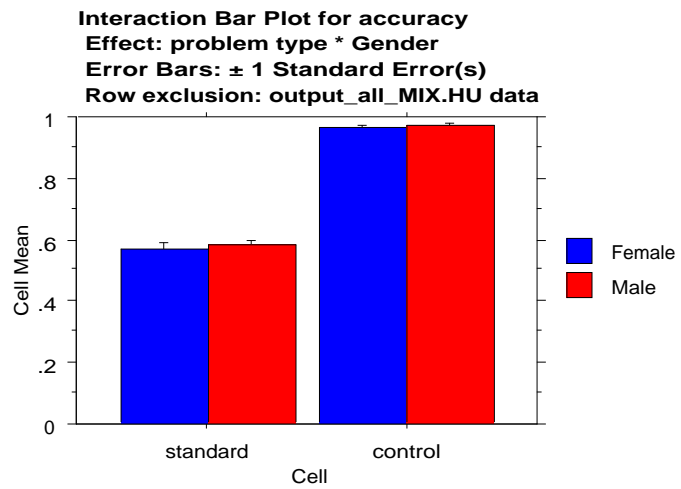
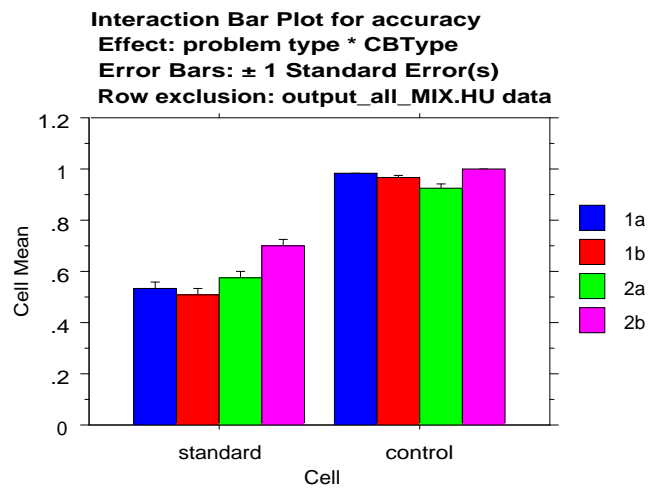
	Count	Mean	Std. Dev.	Std. Err.
Human, 1a, Female, standard	31	.484	.508	.091
Human, 1a, Female, control	31	.968	.180	.032
Human, 1a, Male, standard	68	.412	.496	.060
Human, 1a, Male, control	68	1.000	0.000	0.000
Human, 1b, Female, standard	36	.472	.506	.084
Human, 1b, Female, control	36	.944	.232	.039
Human, 1b, Male, standard	60	.433	.500	.065
Human, 1b, Male, control	60	.950	.220	.028
Human, 2a, Female, standard	40	.250	.439	.069
Human, 2a, Female, control	40	.925	.267	.042
Human, 2a, Male, standard	53	.340	.478	.066
Human, 2a, Male, control	53	.943	.233	.032
Human, 2b, Female, standard	40	.350	.483	.076
Human, 2b, Female, control	40	.975	.158	.025
Human, 2b, Male, standard	53	.509	.505	.069
Human, 2b, Male, control	53	1.000	0.000	0.000
LLM_MIX8X7B_T13, 1a, Female, standard	124	.597	.493	.044
LLM_MIX8X7B_T13, 1a, Female, control	124	.984	.126	.011
LLM_MIX8X7B_T13, 1a, Male, standard	272	.544	.499	.030
LLM_MIX8X7B_T13, 1a, Male, control	272	.974	.159	.010
LLM_MIX8X7B_T13, 1b, Female, standard	144	.479	.501	.042
LLM_MIX8X7B_T13, 1b, Female, control	144	.979	.143	.012
LLM_MIX8X7B_T13, 1b, Male, standard	240	.554	.498	.032
LLM_MIX8X7B_T13, 1b, Male, control	240	.975	.156	.010
LLM_MIX8X7B_T13, 2a, Female, standard	160	.594	.493	.039
LLM_MIX8X7B_T13, 2a, Female, control	160	.912	.283	.022
LLM_MIX8X7B_T13, 2a, Male, standard	212	.684	.466	.032
LLM_MIX8X7B_T13, 2a, Male, control	212	.939	.240	.017
LLM_MIX8X7B_T13, 2b, Female, standard	160	.794	.406	.032
LLM_MIX8X7B_T13, 2b, Female, control	160	1.000	0.000	0.000
LLM_MIX8X7B_T13, 2b, Male, standard	212	.745	.437	.030
LLM_MIX8X7B_T13, 2b, Male, control	212	1.000	0.000	0.000



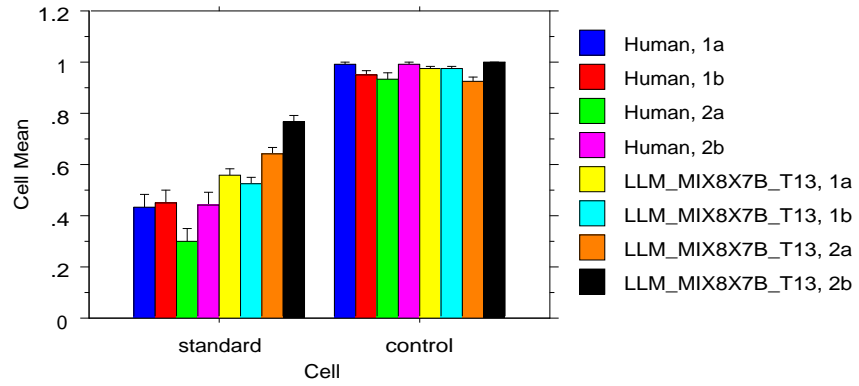




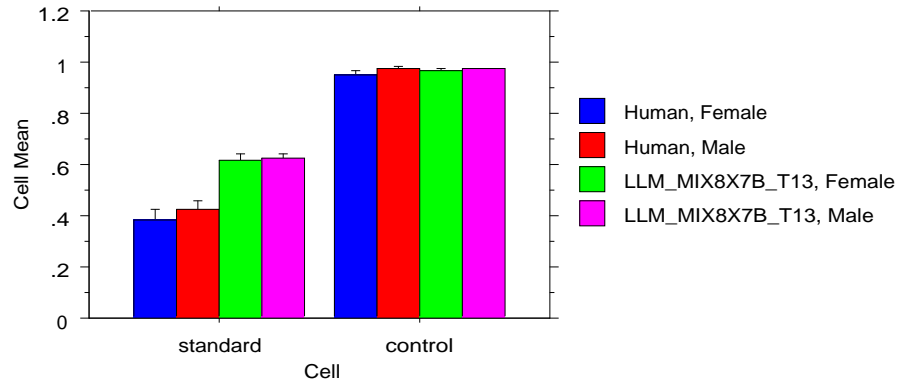


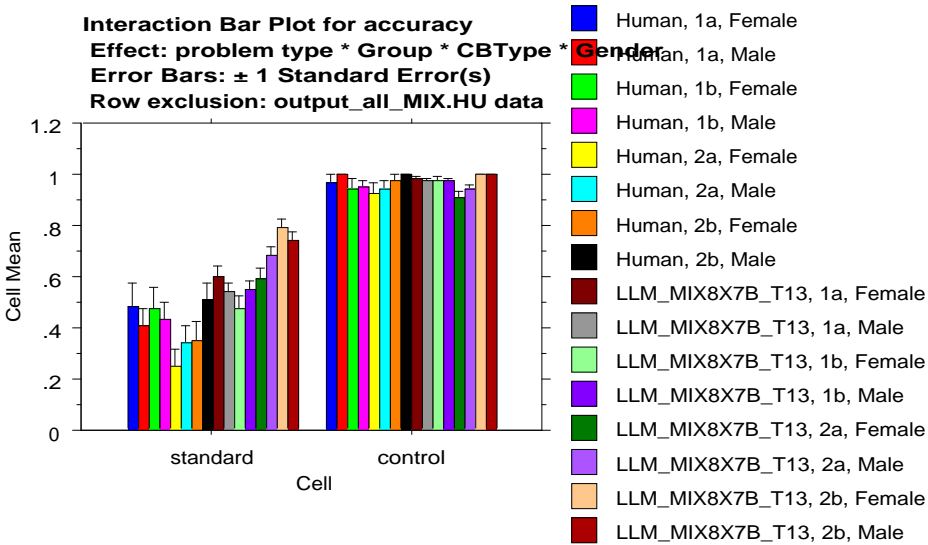
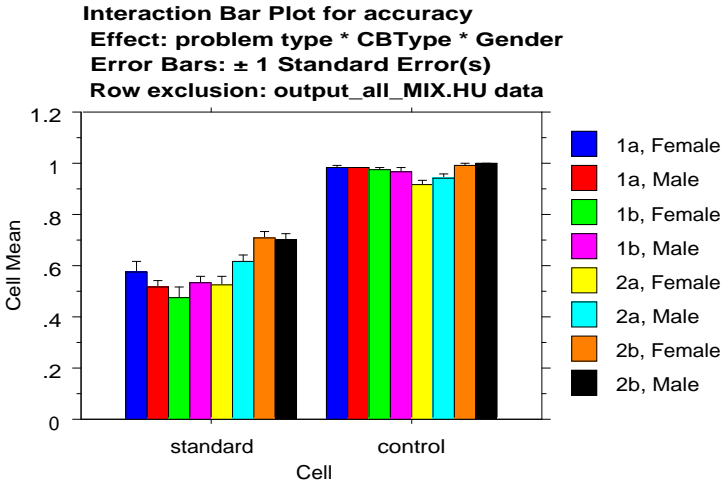


Interaction Bar Plot for accuracy
Effect: problem type * Group * CBType
Error Bars: ± 1 Standard Error(s)
Row exclusion: output_all_MIX.HU data



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





Fisher's PLSD for accuracy
Effect: Group
Significance Level: 5 %
Row exclusion: output_all_MIX.HU data

	Mean Diff.	Crit. Diff.	P-Value
Human, LLM_MIX8X7B_T13	-.110	.029	<.0001

S

Fisher's PLSD for accuracy
Effect: CBType
Significance Level: 5 %
Row exclusion: output_all_MIX.HU data

	Mean Diff.	Crit. Diff.	P-Value	
1a, 1b	.017	.032	.3035	
1a, 2a	.005	.033	.7685	
1a, 2b	-.092	.033	<.0001	S
1b, 2a	-.012	.033	.4710	
1b, 2b	-.109	.033	<.0001	S
2a, 2b	-.097	.033	<.0001	S

Fisher's PLSD for accuracy
Effect: Gender
Significance Level: 5 %
Row exclusion: output_all_MIX.HU data

	Mean Diff.	Crit. Diff.	P-Value
Female, Male	-.009	.024	.4327

Fisher's PLSD for accuracy
Effect: problem type
Significance Level: 5 %
Row exclusion: output_all_MIX.HU data

	Mean Diff.	Crit. Diff.	P-Value	
standard, control	-.390	.023	<.0001	S