A Datasets

Table 21: Mean confidence scores (μ) and standard deviation of confidence scores (σ) across each QT30 subcorpus.

Corpus Name	μ	σ
28May2020	0.76	0.16
4June2020	0.72	0.17
18June2020	0.76	0.16
30July2020	0.75	0.16
2September2020	0.78	0.15
22October2020	0.76	0.16
5November2020	0.77	0.17
19November2020	0.74	0.19
10December2020	0.77	0.16
14January2021	0.74	0.17
28January2021	0.70	0.17
18February2021	0.75	0.16
4March2021	0.76	0.16
18March2021	0.75	0.17
15April2021	0.75	0.15
29April2021	0.70	0.19
20May2021	0.76	0.17
27May2021	0.79	0.15
10June2021	0.75	0.17
24June2021	0.74	0.17
8July2021	0.72	0.17
22July2021	**0.44**	**0.28**
5August2021	0.76	0.17
19August2021	0.77	0.17
2September2021	0.78	0.16
16September2021	0.77	0.16
30September2021	**0.24**	**0.08**
14October2021	0.75	0.19
28October2021	0.75	0.17
11November2021	0.78	0.16
QT30	0.75	0.17

Table 22: Mean confidence scores (μ) and Standard Deviation of confidence scores (σ) across Moral Maze subcorpora.

Subcorpus	μ	σ
B	0.79	0.14
E	0.76	0.15
M	0.78	0.14
P	0.80	0.15
S	0.74	0.16
G	0.80	0.14
H	0.74	0.16
W	0.75	0.15

B Results

B.1 In-Domain

Table 23: In-domain results when testing different model architectures on the 4-class problem. F1: Macro-averaged F1, P: precision, R: recall. Highest scores in each column are shown in bold.

Fusion Methods		SCS			LCS			US		
Sequence	Multimodal	F1	P	R	F1	P	R	F1	P	R
Text Only										
Early	-	**.58**	**.58**	**.58**	**.59**	**.59**	**.59**	**.59**	**.59**	**.59**
Late	-	.36	.36	.35	.34	.36	.34	.35	.35	.35
Audio Only										
Early	-	.43	.48	.44	.41	.41	.42	.20	.31	.26
Late	-	.28	.29	.29	.29	.31	.29	.29	.29	.29
Multimodal										
Early	Concatenation	**.58**	**.58**	**.58**	.57	.57	.57	.58	.58	.57
	Product	.56	.56	.57	.57	.57	.57	.58	.58	**.59**
	CA Text	.57	.56	**.58**	.46	.46	.48	.57	.56	.57
	CA Audio	**.58**	**.58**	**.58**	.57	**.59**	.56	.57	.58	.57
Late	Concatenation	.36	.36	.35	.37	.36	.36	.35	.36	.35
Baselines										
Random		.22	.24	.23	.23	.25	.24	.24	.25	.26
Majority		.14	.09	.25	.14	.09	.25	.14	.09	.25

Table 24: In-domain results when testing different model architectures on the 3-class problem. F1: Macro-averaged F1, P: precision, R: recall. Highest scores in each column are shown in bold.

Fusion Methods		**SCS**			**LCS**			**US**		
Sequence	Multimodal	F1	P	R	F1	P	R	F1	P	R
Text Only										
Early	-	**.62**	**.63**	**.62**	.59	**.63**	.58	.61	.63	.60
Late	-	.43	.45	.43	.44	.45	.44	.43	.44	.44
Audio Only										
Early	-	.21	.16	.33	.54	.54	.55	.22	.16	.33
Late	-	.33	.33	.34	.34	.33	.34	.34	.33	.35
Multimodal										
Early	Concatenation	.61	.62	.60	.61	.61	.61	.60	.61	.59
	Product	.58	.60	.57	.61	.62	.61	.62	**.64**	.61
	CA Text	.53	.52	.54	.53	.52	.55	.52	.51	.53
	CA Audio	.61	.62	.60	**.62**	**.63**	**.62**	**.63**	.62	**.64**
Late	Concatenation	.43	.44	.43	.45	.47	.45	.44	.45	.44
Baselines										
Random		.28	.33	.33	.30	.34	.36	.24	.25	.26
Majority		.22	.16	.33	.22	.16	.33	.22	.16	.33

B.2 Cross-Domain

Table 25: Cross domain, 4-class macro-F1 results trained on SCS. Highest values in each column are shown in bold.

Fusion Methods		B	E	M	P	S	G	H	W	Mean
Sequence	Multimodal									
Text-Only										
Early	-	.44	.46	**.48**	.42	.43	.50	**.51**	.41	.46
Late	-	.31	.28	.32	.26	.30	.31	.33	.29	.30
Audio-Only										
Early	-	.34	.37	.39	.38	.33	.38	.40	.37	.37
Late	-	.20	.22	.24	.22	.22	.28	.25	.22	.23
Multimodal										
Early	Concatenation	.43	.43	.45	**.45**	.45	.49	.45	.43	.45
	Product	.41	.44	.42	.42	**.46**	.52	.46	.43	.45
	CA Text	.40	.44	.44	.40	.42	.49	.43	.44	.43
	CA Audio	**.45**	**.49**	.44	.42	.45	**.53**	.48	**.46**	**.47**
Late	Concatenation	.32	.28	.31	.27	.29	.31	.35	.30	.30
Baselines										
Random		.19	.23	.20	.19	.22	.25	.21	.24	.22
Majority		.16	.15	.15	.15	.15	.14	.15	.14	.15

Table 26: Cross domain, 4-class macro-F1 results trained on LCS. Highest values in each column are shown in bold.

Fusion Methods		B	E	M	P	S	G	H	W	Mean
Sequence	Multimodal									
Text-Only										
Early	-	**.41**	.46	**.46**	**.44**	.45	.52	.45	**.46**	**.46**
Late	-	.31	.28	.30	.33	.33	.33	.38	.32	.32
Audio-Only										
Early	-	.33	.37	.38	.36	.32	.36	.40	.36	.36
Late	-	.22	.22	.24	.24	.20	.27	.26	.23	.24
Multimodal										
Early	Concatenation	**.41**	.43	.41	.42	.39	**.53**	.45	.44	.43
	Product	.39	.44	.42	.41	**.47**	.51	.43	.44	.44
	CA Text	.25	.44	.33	.29	.27	.40	.34	.32	.32
	CA Audio	**.41**	**.49**	.44	.42	.46	.49	**.52**	**.46**	**.46**
Late	Concatenation	.29	.28	.27	.24	.28	.31	.34	.29	.29
Baselines										
Random		.19	.23	.20	.19	.22	.25	.21	.24	.22
Majority		.16	.15	.15	.15	.15	.14	.15	.14	.15

Table 27: Cross domain, 4-class macro-F1 results trained on US. Highest values in each column are shown in bold.

Fusion Methods		B	E	M	P	S	G	H	W	Mean
Sequence	Multimodal									
Text-Only										
Early	-	.43	.44	**.45**	.44	.45	**.54**	**.47**	.45	**.46**
Late	-	.29	.27	.33	.29	.34	.35	.31	.33	.31
Audio-Only										
Early	-	.23	.22	.23	.22	.23	.18	.22	.22	.22
Late	-	.29	.31	.27	.24	.28	.31	.34	.29	.29
Multimodal										
Early	Concatenation	.40	.43	.43	**.45**	.44	.51	.44	**.46**	.45
	Product	**.44**	**.50**	.42	.44	**.47**	.51	.46	**.46**	**.46**
	CA Text	.36	.45	.39	.38	.41	.48	.37	.42	.41
	CA Audio	.40	.40	.41	.41	.43	.50	.45	.43	.43
Late	Concatenation	.26	.24	.28	.28	.33	.33	.31	.33	.30
Baselines										
Random		.19	.23	.20	.19	.22	.25	.21	.24	.22
Majority		.16	.15	.15	.15	.15	.14	.15	.14	.15

Table 28: Cross domain, 3-class macro-F1 results trained on SCS. Highest values in each column are shown in bold.

Fusion Methods		B	E	M	P	S	G	H	W	Mean
Sequence	Multimodal									
Text-Only										
Early	-	.54	.47	.50	**.50**	.58	**.55**	.63	.51	**.54**
Late	-	.45	.38	.35	.36	.36	.37	.43	.34	.38
Audio-Only										
Early	-	.21	.21	.22	.21	.20	.21	.22	.21	.21
Late	-	.33	.32	.32	.33	.32	.34	.34	.32	.33
Multimodal										
Early	Concatenation	.58	**.49**	.47	**.50**	.57	.57	.59	.49	.53
	Product	.51	.45	.44	.47	.53	.54	**.64**	.47	.51
	CA Text	.43	.40	.43	.41	.44	.46	.47	.44	.44
	CA Audio	**.59**	.44	**.52**	.48	**.59**	.54	.60	**.52**	**.54**
Late	Concatenation	.37	.37	.34	.40	.43	.42	.43	.40	.40
Baselines										
Random		.27	.32	.30	.27	.30	.29	.29	.33	.30
Majority		.21	.21	.22	.21	.20	.21	.22	.20	.21

Table 29: Cross domain, 3-class macro-F1 results trained on LCS. Highest values in each column are shown in bold.

Fusion Methods		B	E	M	P	S	G	H	W	Mean
Sequence	Multimodal									
Text-Only										
Early	-	**.59**	.49	.48	.46	.51	.54	**.69**	**.52**	**.54**
Late	-	.38	.36	.36	.39	.44	.43	.41	.43	.40
Audio-Only										
Early	-	.51	.47	.45	.33	.47	.44	.49	.53	.49
Late	-	.22	.25	.24	.24	.20	.27	.26	.23	.24
Multimodal										
Early	Concatenation	.51	.47	.45	.48	.56	.53	.58	.49	.51
	Product	.53	.49	**.50**	**.52**	.56	.54	.59	.53	.53
	CA Text	.46	.40	.45	.42	.46	.45	.45	.51	.45
	CA Audio	.52	**.55**	.47	**.52**	.58	**.55**	.60	**.52**	**.54**
Late	Concatenation	.43	.37	.33	.36	.4	.42	.41	.39	.39
Baselines										
Random		.27	.32	.30	.27	.30	.29	.29	.33	.30
Majority		.21	.21	.22	.21	.20	.21	.22	.20	.21

Table 30: Cross domain, 3-class macro-F1 results trained on US. Highest values in each column are shown in bold.

Fusion Methods		B	E	M	P	S	G	H	W	Mean
Sequence	Multimodal									
Text-Only										
Early	-	.57	.48	.47	.50	**.62**	**.55**	.60	**.54**	.54
Late	-	.47	.35	.36	.39	.40	.40	.39	.40	.40
Audio-Only										
Early	-	.21	.21	.22	.21	.20	.21	.22	.21	.21
Late	-	.33	.34	.32	.33	.34	.34	.33	.33	.33
Multimodal										
Early	Concatenation	**.59**	.46	.47	.50	.55	**.55**	.60	.50	.53
	Product	.55	.49	**.50**	.52	.54	.52	.61	.50	.53
	CA Text	.42	.39	.42	.41	.42	.46	.49	.40	.43
	CA Audio	**.59**	**.52**	.49	**.53**	.59	.54	**.62**	.53	**.55**
Late	Concatenation	.38	.36	.36	.37	.36	.38	.43	.34	.37
Baselines										
Random		.27	.32	.30	.27	.30	.29	.29	.33	.30
Majority		.21	.21	.22	.21	.20	.21	.22	.20	.21