

Run ID:	Emory1
3-way accuracy:	0.546
2-way accuracy:	0.576
Average precision:	0.5994

		System Response			Total
		Entailment	Unknown	Contradiction	
Gold Standard	Entailment	397	103	0	500
	Unknown	201	149	0	350
	Contradiction	120	30	0	150
Total		718	282	0	1000

Table 1: Confusion matrix for the 3-way task for Emory1

Run ID:	Emory2
3-way accuracy:	0.547
2-way accuracy:	0.583
Average precision:	0.5954

		System Response			Total
		Entailment	Unknown	Contradiction	
Gold Standard	Entailment	388	111	1	500
	Unknown	191	158	1	350
	Contradiction	114	35	1	150
Total		693	304	3	1000

Table 2: Confusion matrix for the 3-way task for Emory2

Run ID:	Emory3
3-way accuracy:	0.543
2-way accuracy:	0.564
Average precision:	0.5975

		System Response			Total
		Entailment	Unknown	Contradiction	
Gold Standard	Entailment	432	68	0	500
	Unknown	239	111	0	350
	Contradiction	129	21	0	150
Total		800	200	0	1000

Table 3: Confusion matrix for the 3-way task for Emory3