Cohesion					Coherence				
Human judges preferred:					*Human judges preferred:*				
Our Method		Neutral	Comparison		Our Method		Neutral	Comparison	
$G_{\text{MLE+RL}}$	36.25%	26.62%	**37.13%**	G_{MLE}	$G_{\text{MLE+RL}}$	**39.25%**	23.12%	37.63%	G_{MLE}
$G_{\text{MLE+RL}}$	34.25%	23.63%	**42.12%**	Human	$G_{\text{MLE+RL}}$	35.63%	21.50%	**42.87%**	Human

Table 6: Results of **Human Evaluation** showing preferences (%) for our model $G_{\text{MLE+RL(coherence, cohesion)}}$ vis-a-vis the baseline G_{MLE} *before* adjustment for spamming. For simplicity, the 5-point Likert scale has been collapsed to a 3-point scale.

A Human evaluation un-adjusted scores

Crowd-sourced evaluation can be noisy because there may be human judges who do not take the task seriously, and rather randomly and/or deliberately choose options that prevent us from drawing accurate conclusions. Therefore, we removed crowd-sourced judges who chose $G_{\text{MLE+RL}}$ over the *Human* more than 40% of the time, which threshold value we considered appropriate to identify poor judges (probable spammers). In Table 6, we present the un-adjusted results *before* accounting for the poor judges.

B Sparse end-of-sequence rewards

Sequence-level rewards are available upon a completed generation, so they are sparse signals for the generator. In practice, sparse end-of-sequence rewards entail a noisy training, yet would want the learning generalize to the test data. We observed that, for our particular task, most noises were caused by exploration, and the learning generalized to the test data, as confirmed via both human and automatic evaluation results. Thus, reward shaping was unnecessary, unlike previous works (Li et al., 2017; Yang et al., 2018) that further provided signals for partially generated sequences.