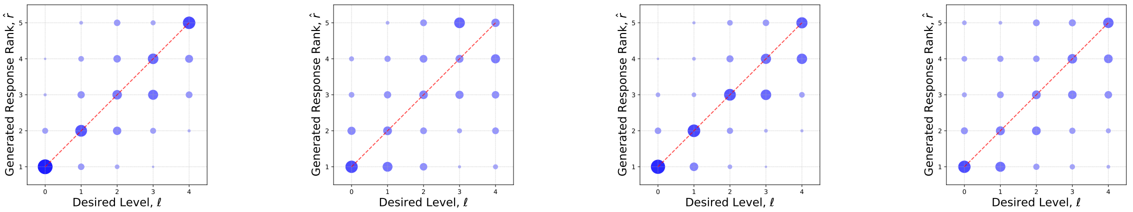

(a) Humor (b) Humor | Persuasiveness (c) Persuasiveness (d) Persuasiveness | Humor

Figure 2: Model-generated response rank of the target concept versus the desired level. Point size and density indicate the number of samples at each coordinate. Results shown for Llama-11B with the secondary concept level *randomly* sampled. For example, "Humor | Persuasiveness" denotes responses generated independently for each humor level (target concept) while persuasiveness is randomly set for each inference.

	Argument Generation			**Story Generation**			**Structured Text Generation**		
	Llama-11B	**Gemma-12B**	**Qwen-14B**	**Llama-11B**	**Gemma-12B**	**Qwen-14B**	**Llama-11B**	**Gemma-12B**	**Qwen-14B**
C_a (single)	$0.76{\pm}0.23$	$0.95{\pm}0.07$	$0.92{\pm}0.11$	$0.81{\pm}0.26$	$0.95{\pm}0.06$	$0.92{\pm}0.10$	$0.73{\pm}0.22$	$0.94{\pm}0.12$	$0.90{\pm}0.12$
$C_a \mid C_b$ fixed	$0.51{\pm}0.41$	$0.88{\pm}0.14$	$0.88{\pm}0.15$	$0.36{\pm}0.45$	$0.81{\pm}0.22$	$0.90{\pm}0.12$	$0.31{\pm}0.50$	$0.88{\pm}0.15$	$0.84{\pm}0.19$
$C_a \mid C_b$ rand	$0.54{\pm}0.35$	$0.83{\pm}0.20$	$0.88{\pm}0.16$	$0.33{\pm}0.49$	$0.74{\pm}0.25$	$0.88{\pm}0.15$	$0.17{\pm}0.48$	$0.79{\pm}0.21$	$0.81{\pm}0.21$
C_b (single)	$0.81{\pm}0.22$	$0.98{\pm}0.04$	$0.96{\pm}0.05$	$0.80{\pm}0.19$	$0.97{\pm}0.04$	$0.93{\pm}0.10$	$0.89{\pm}0.14$	$0.99{\pm}0.02$	$0.99{\pm}0.03$
$C_b \mid C_a$ fixed	$0.58{\pm}0.38$	$0.83{\pm}0.19$	$0.84{\pm}0.18$	$0.59{\pm}0.35$	$0.69{\pm}0.34$	$0.85{\pm}0.18$	$0.56{\pm}0.41$	$0.91{\pm}0.15$	$0.90{\pm}0.14$
$C_b \mid C_a$ rand	$0.52{\pm}0.40$	$0.76{\pm}0.21$	$0.81{\pm}0.21$	$0.58{\pm}0.34$	$0.70{\pm}0.31$	$0.83{\pm}0.20$	$0.51{\pm}0.39$	$0.79{\pm}0.19$	$0.83{\pm}0.19$

Table 1: **Humor–persuasiveness.** Spearman correlations for single-concept and dual-concept (fixed / random) across argument, story, and structured text generation.

	Argument Generation			**Story Generation**			**Structured Text Generation**		
	Llama-11B	**Gemma-12B**	**Qwen-14B**	**Llama-11B**	**Gemma-12B**	**Qwen-14B**	**Llama-11B**	**Gemma-12B**	**Qwen-14B**
C_a (single)	$-0.02{\pm}0.52$	$0.52{\pm}0.46$	$0.65{\pm}0.30$	$0.45{\pm}0.46$	$0.92{\pm}0.11$	$0.89{\pm}0.12$	$0.21{\pm}0.56$	$0.15{\pm}0.61$	$0.64{\pm}0.21$
$C_a \mid C_b$ fixed	$0.02{\pm}0.53$	$0.02{\pm}0.56$	$0.64{\pm}0.34$	$-0.01{\pm}0.53$	$0.35{\pm}0.43$	$0.74{\pm}0.26$	$0.02{\pm}0.45$	$-0.25{\pm}0.53$	$0.39{\pm}0.43$
$C_a \mid C_b$ rand	$-0.05{\pm}0.51$	$0.12{\pm}0.50$	$0.63{\pm}0.32$	$-0.07{\pm}0.50$	$0.29{\pm}0.47$	$0.64{\pm}0.29$	$0.08{\pm}0.40$	$-0.19{\pm}0.45$	$0.38{\pm}0.43$
C_b (single)	$0.76{\pm}0.25$	$0.95{\pm}0.07$	$0.93{\pm}0.10$	$0.84{\pm}0.21$	$0.98{\pm}0.03$	$0.96{\pm}0.07$	$0.73{\pm}0.28$	$0.97{\pm}0.03$	$0.93{\pm}0.09$
$C_b \mid C_a$ fixed	$0.76{\pm}0.25$	$0.83{\pm}0.19$	$0.88{\pm}0.14$	$0.71{\pm}0.30$	$0.86{\pm}0.17$	$0.95{\pm}0.08$	$0.45{\pm}0.42$	$0.79{\pm}0.31$	$0.79{\pm}0.26$
$C_b \mid C_a$ rand	$0.77{\pm}0.29$	$0.80{\pm}0.18$	$0.84{\pm}0.15$	$0.71{\pm}0.31$	$0.72{\pm}0.26$	$0.92{\pm}0.08$	$0.37{\pm}0.47$	$0.63{\pm}0.33$	$0.76{\pm}0.28$

Table 2: **Clarity–politeness.** Spearman correlations for single-concept and dual-concept (fixed / random) across argument, story, and structured text generation.

	Argument Generation			**Story Generation**			**Structured Text Generation**		
	Llama-11B	**Gemma-12B**	**Qwen-14B**	**Llama-11B**	**Gemma-12B**	**Qwen-14B**	**Llama-11B**	**Gemma-12B**	**Qwen-14B**
C_a (single)	$0.92{\pm}0.09$	$0.98{\pm}0.03$	$0.99{\pm}0.02$	$0.93{\pm}0.09$	$1.00{\pm}0.02$	$0.98{\pm}0.05$	$0.80{\pm}0.24$	$0.93{\pm}0.14$	$0.96{\pm}0.07$
$C_a \mid C_b$ fixed	$0.56{\pm}0.40$	$0.97{\pm}0.05$	$0.97{\pm}0.05$	$0.77{\pm}0.25$	$0.96{\pm}0.07$	$0.96{\pm}0.06$	$0.42{\pm}0.45$	$0.77{\pm}0.33$	$0.88{\pm}0.15$
$C_a \mid C_b$ rand	$0.41{\pm}0.43$	$0.92{\pm}0.10$	$0.94{\pm}0.08$	$0.77{\pm}0.23$	$0.96{\pm}0.06$	$0.96{\pm}0.05$	$0.22{\pm}0.48$	$0.71{\pm}0.33$	$0.86{\pm}0.17$
C_b (single)	$0.75{\pm}0.32$	$0.99{\pm}0.03$	$0.98{\pm}0.03$	$0.67{\pm}0.33$	$0.98{\pm}0.04$	$0.97{\pm}0.06$	$0.66{\pm}0.32$	$0.95{\pm}0.08$	$0.87{\pm}0.16$
$C_b \mid C_a$ fixed	$0.48{\pm}0.47$	$0.90{\pm}0.12$	$0.94{\pm}0.08$	$0.51{\pm}0.42$	$0.93{\pm}0.10$	$0.91{\pm}0.10$	$0.43{\pm}0.50$	$0.72{\pm}0.36$	$0.76{\pm}0.29$
$C_b \mid C_a$ rand	$0.45{\pm}0.44$	$0.85{\pm}0.15$	$0.93{\pm}0.07$	$0.41{\pm}0.46$	$0.89{\pm}0.12$	$0.89{\pm}0.12$	$0.40{\pm}0.51$	$0.72{\pm}0.26$	$0.75{\pm}0.24$

Table 3: **Formality–assertiveness.** Spearman correlations for single-concept and dual-concept (fixed / random) across argument, story, and structured text generation.

achieve significantly higher correlations. Politeness follows the standard pattern: high performance for a single concept, but a drop when clarity is introduced. Similarly, in formality–assertiveness, both concepts exhibit consistently high single-concept control (up to 1.00 for Gemma) but degrade under dual-control conditions.

General trends. Three broader insights emerge: (i) Qwen-14B and Gemma-12B consistently outperform Llama across all settings. This suggests that larger or more instruction-tuned models better preserve disentanglement between stylistic dimensions. (ii) Dual-concept interference remains a central limitation: even when single-concept control is strong, the introduction of a secondary dimension leads to drops in alignment (Figure 2), suggesting weak compositionality of stylistic control. (iii) Task context strongly modulates controllabil-