metric = Violated Request Count experiment = 0_easy 1.0 0.8 0.6 0.4 0.2 gemgem 8b 8b 170b 170b mis-mis mix-mix openopengwenqwe metric = Violated pattern experiment = 0_easy		metric = Violated Request Count experiment = 12_medium-less_turns 1.0 8.8 0.6 4.4 0.4 2.0 gemgem 8b 8b 70b 70b mismis mixmix openopengwengwen metric = Violated pattern experiment = 12_medium-less_turns		metric = Violated Request Count experiment = 1_medium 1.0 0.8 0.6 0.4 0.2	metric = Violated Request Count experiment = 2_difficult 1.0 0.8 0.6 0.4 0.2 gem 8b 8b 70b 70b mis-mis mix-mix openopenqwenqwen metric = Violated pattern experiment = 2_difficult metric metr	= Violated Request Count experiment = 4_medium-reprompt 1.0 0.8 0.6 0.4 0.2 m 8b 8b	metric = Violated Request Count experiment = 7_medium-themselves gemgem 18b18b 170b1170b mismis mixmix openopenqwenqwen metric = Violated pattern experiment = 7_medium-themselves
1.0 0.8 0.6 0.7 0.9 0.0 gemgem 8b 8b 70b 170b mismis mixmix openopengwenqwe metric = Violated token length experiment = 0_easy	1.0 0.8 0.6 0.4 0.2 0.0 0.0 0.2	1.0 88 0.8 0.6 4 0.4 2 0.2 0.2 0.2 0.2 0.2 0.	gemgem 8b 8b 70b 70b mismis mixmix openopenqwenqwen metric = Violated token length experiment = 13_medium-more_turns 0.04	gemgem 8b 8b 70b 70b mismis mixmix openopenqwenqwen gemmetric = Violated token length experiment = 1_medium 0.20		1.0 0.8 0.6 0.4 0.2 m 8b 8b 170b 70b mismis mixmix openopengwenqwen c = Violated token length experiment = 4_medium-reprompt 0.04	gemgem
0.02 -0.02 -0.04 gemgem 8b 8b 70b 70b mismis mixmix openopengwenqwe metric = Parsed Request Count experiment = 0_easy	0.02 0.00 -0.02 -0.04 gemgem 8b 8b 70b 70b mismis mixmix openopengwenqwen metric = Parsed Request Count experiment = 10_medium-same_gender 12 10	0.15 0.16 0.17 0.17 0.18 0.19 0.19 0.19 0.19 0.19 0.19 0.19 0.19	-0.02 -0.04	gemgem 8b 8b 70b 70b mismis mixmix openopenqwenqwen gemmetric = Parsed Request Count experiment = 1_medium 12 10	gem	0.02 -0.02 -0.04 -0.04 -0.04 -0.05 -0.04 -0.04 -0.05 -0.04 -0.04 -0.04 -0.04 -0.04	gemgem
gemgem II8bII8b II70bII70b mismis mixmix openopengwenqwemetric = Request Count experiment = 0_easy	gemgem 8b 8b 70b 70b mismis mixmix openopengwenqwen metric = Request Count experiment = 10_medium-same_gender	gemgem 8b 8b 70b 70b mismis mixmix openopengwenqwen metric = Request Count experiment = 12_medium-less_turns	gemgem 8b 8b 70b 70b mismis mixmix openopenqwenqwen metric = Request Count experiment = 13_medium-more_turns	gemgem 8b 8b 70b 70b mismis mixmix openopenqwenqwen gemmetric = Request Count experiment = 1_medium 12 10 8	gem	m II8bII8b II70bII70b mismis mixmix openoperqwenqwen etric = Request Count experiment = 4_medium-reprompt	gemgem
gemgem 8b 8b 70b 70b mismis mixmix openopengwenqwemetric = Request Success Ratio experiment = 0_easy	gemgem 8b 8b 70b 70b mismis mixmix openopenqwenqwen metric = Request Success Ratio experiment = 10_medium-same_gender 1.0 0.8 0.6 0.6	gemgem 8b 8b 70b 70b mismis mixmix openopengwenqwen metric = Request Success Ratio experiment = 12_medium-less_turns 1.0 8 0.8	gemgem 8b 8b 70b 70b mismis mixmix openopengwenqwen metric = Request Success Ratio experiment = 13_medium-more_turns 1.0 0.8 0.6	gemgem 8b 8b 70b 70b mismis mixmix openopenqwenqwen gemmetric = Request Success Ratio experiment = 1_medium 1.0 0.8 0.6	gem II8bII8b II70bII70b mismis mixmix openoperqwenqwen gemger metric = Request Success Ratio experiment = 2_difficult metric 1.0 0.8 0.8	m 8b 8b 70b 70b mismis mixmix openopenqwenqwen = Request Success experiment = 4_medium-reprompt 1.0 0.8 0.6	gemgem II8bII8b II70bII70b mismis mixmix openoperqwenqwen metric = Request Success Ratio experiment = 7_medium-themselves
0.4 0.2 gemgem 8b 8b 70b 70b mismis mixmix openopengwenqwe metric = Number of completed turns experiment = 0_easy 12 10 8 9 9 9 6 4	o.2 o.0 gemgem 8b 8b 70b 70b mismis mixmix openopengwenqwen metric = Number of completed turns experiment = 10_medium-same_gender 12 10 8 6 4	gemgem ll8bll8b ll70bll70b mismis mixmix openopenqwenqwen	gemgem II8bII8b II70bII70b mismis mixmix openopenqwenqwen	gemgem II8bII8b II70bII70b mismis mixmix openopenqwenqwen gem	gem II8bII8b II70bII70b mismis mixmix openopenqwenqwen gemge	m II8bII8b II70bII70b mismis mixmix openopenqwenqwen	gemgem II8bII8b II70bII70b mismis mixmix openoperqwenqwen metric = Number of completed turns experiment = 7_medium-themselves
2 0 gemgem 8b 8b 70b 70b mismis mixmix openopenqwenqwe metric = Number of Agreements experiment = 0_easy 3.0 2.5 2.0 0.5			gemgem				gemgem II8bII8b II70bII70b mismis mixmix openopenqwenqwen metric = Number of Agreements experiment = 7_medium-themselves
gemgem	metric = Agreement Efficiency experiment = 10_medium-same_gender 100 80 60 40 20 20	metric = Agreement Efficiency experiment = 12_medium-less_turns 100 80 80 60 60 20 20	gemgem			m	gemgem
gemgem		gemgem		gemgem	gem	mber of Mismatched Agreements experiment = 4_medium-repromptmets 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0 0.5 0.0	gemgem
0.02 -0.02 -0.04 gemgem 8b 8b 70b 70b mismis mixmix openopenqwenqwe metric = Number of Redundancy experiment = 0_easy 0.8		94 — 0.02 gemgem 118b118b 1170b1170b mismis mixmix openopenqwenqwen	-0.02 -0.04	gemgem	gem II8bII8b II70bII70b mismis mixmix openopenqwenqwen gemger	0.04 0.02 0.00 0.00 -0.02 -0.04 m 8b 8b 70b 70b mismis mixmix openopenqwenqwen = Number of Redundancy experiment = 4_medium-reprompt 0.4	gemgem
0.6 0.5 0.7 0.0 0.8 0.9 0.9 0.0 gemgem 8b 8b 70b 70b mismis mixmix openopengwenqwe metric = Friendzoned experiment = 0_easy 0.04 0.02	2.0 1.5 1.0 0.6 0.7 0.7 0.7 0.7 0.7 0.7 0	0.3 0.3 0.4 2. 0.1 gemgem 8b 8b 70b 70b mismis mixmix openopenqwenqwen metric = Friendzoned experiment = 12_medium-less_turns 0.20 0.20	gemgem 8b 8b 70b 70b mismis mixmix openopenqwenqwen metric = Friendzoned experiment = 13_medium-more_turns 0.20	gemgem 8b- 8b 70b 70b mismis mixmix openopenqwenqwen metric = Friendzoned experiment = 1_medium		m	gemgem
0.00 0.00	0.15 0.10 0.05 0.00	metric = Out of turns experiment = 12_medium-less_turns 0.04 8 0.02	0.15	gemgem	metric = Out of turns experiment = 2_difficult	0.15 0.10 0.05 m	gemgem
0.00 gemgem	0.05 en	2	metric = Out of retries experiment = 13_medium-more_turns 0.04	metric = Out of retries experiment = 1_medium 0.04 0.02	metric = Out of retries experiment = 2_difficult 1.0 0.8	0.05 m	gemgem
gemgem 8b 8b 70b 70b mismis mixmix openopengwenqwe	-0.04 en gemgem	gemgem 8b 8b 170b 70b mismis mixmix openopenqwenqwen metric = Turn penalty experiment = 12_medium-less_turns 14 12 12 10 10 11 12 12 10 11 12 12	gemgem	gemgem	gem II8bII8b II70bII70b mismis mixmix openopenqwenqwen metric = Turn penalty experiment = 2_difficult 14 12 10 8 6 4	m II8bII8b II70bII70b mismis mixmix openopengwenqwen netric = Turn penalty experiment = 4_medium-reprompt 7 6 5 4 3	gemgem II8bII8b II70bII70b mismis mixmix openoperqwenqwen metric = Turn penalty experiment = 7_medium-themselves
gemgem 8b 8b 70b 70b mismis mixmix openopenqwenqwe metric = Agreement penalty experiment = 0_easy 20 15 3 7	metric = Agreement penalty experiment = 10_medium-same_gender 20 15 10 5	gemgem 8b 8b 170b 170b mismis mixmix openopenqwenqwen metric = Agreement penalty experiment = 12_medium-less_turns	metric = Agreement penalty experiment = 13_medium-more_turns 20 15 10	metric = Agreement penalty experiment = 1_medium 20 15 10 5	metric = Agreement penalty experiment = 2_difficult 20 15 10 5	m	gemgem
metric = Inefficiency penalty experiment = 0_easy 8 7 6 9 9 9 9 9 9 9 9 9 9 118b118b 170b1170b mismis mixmix openopenqwenqwe metric = Error penalty experiment = 0_easy	metric = Inefficiency penalty experiment = 10_medium-same_gender 25 20 15 10 5	metric = Inefficiency penalty experiment = 12_medium-less_turns 2 0 8 4 2 1 1	metric = Inefficiency penalty experiment = 13_medium-more_turns 10 8 4 2	metric = Inefficiency penalty experiment = 1_medium 40 20 10	metric = Inefficiency penalty experiment = 2_difficult 5 4 3 2 gem 8b 8b 70b 70b mismis mixmix openopenqwenqwen gemgengen	ric = Inefficiency penalty experiment = 4_medium-reprompt 4 3 1	metric = Inefficiency penalty experiment = 7_medium-themselves gemgem II8bII8b II70bII70b mismis mixmix openopenqwenqwen metric = Error penalty experiment = 7_medium-themselves
0.04 0.02 -0.02 -0.04 gemgem 8b 8b 70b 70b mismis mixmix openopenqwenqwe metric = Location agreement experiment = 0_easy 1.0	0.04 0.02 0.00 -0.02 -0.02 -0.04 gemgem 8b 8b 70b 70b mismis mixmix openopenqwenqwen metric = Location agreement experiment = 10_medium-same_gender 1.0	0.02 00 0.00 02 04 -0.04	-0.02 -0.04	-0.02 -0.04	gem II8bII8b II70bII70b mismis mixmix openopenqwenqwen gemger	0.04 0.02 0.00 0.00 -0.02 -0.04 m 8b 8b 70b 70b mismis mixmix openopenqwenqwen ic = Location agreement experiment = 4_medium-reprompt 1.0	gemgem 8b 8b 70b 70b mismis mixmix openopenqwenqwen metric = Location agreement experiment = 7_medium-themselves
0.8 0.6 0.0 gemgem 8b 8b 70b 70b mismis mixmix openopengwenqwe metric = Action agreement experiment = 0_easy 1.0 0.8	0.8 0.6 0.7 0.8 0.8 0.8 0.8 0.9 0.9 0.9 0.9	0.8 0.8 0.8 0.6 0.6 0.6 0.7 0.7 0.7 0.7 0.7 0.7 0.7 0.7 0.7 0.7	gemgem	gemgem	gem	m 8b 8b 70b 70b mismis mixmix openopengwenqwen cric = Action agreement experiment = 4_medium-reprompt 1.0 0.8 0.6 0.4 0.2 0.0 0.1 0.8 0.	gemgem
0.4 0.2 0.0 gemgem 8b 8b 70b 70b mismis mixmix openopengwenqwe metric = Time agreement experiment = 0_easy 1.0 0.8 0.6 0.6 0.7 0.8	0.6 0.6 0.7 0.7 0.8 0.8 0.9 0.9 0.9 0.9 0.9 0.9	0.6 decomposition of the second of the seco	gemgem	gemgem 8b 8b 70b 70b mismis mixmix openoperqwenqwen gemmetric = Time agreement experiment = 1_medium 1.0 0.8 0.6	gem	0.6 0.4 0.2 m	gemgem II8bII8b II70bII70b mismis mixmix openopenqwenqwen metric = Time agreement experiment = 7_medium-themselves
0.4 0.2 0.0 gemgem 8b 8b 70b 70b mismis mixmix openopenqwenqwe metric = Aborted experiment = 0_easy 1.0 0.8 0.6 0.7 0.9	0.4 0.2 0.0 gemgem 8b 8b 70b 70b mismis mixmix openopengwenqwen metric = Aborted experiment = 10_medium-same_gender 1.0 0.8 0.6 0.6 0.6	0.4 2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0	gemgem 8b 8b 70b 70b mismis mixmix openopenqwenqwen metric = Aborted experiment = 13_medium-more_turns 1.0 0.8 0.6 0.4	gemgem 8b 8b 70b 70b mismis mixmix openopenqwenqwen gemmetric = Aborted experiment = 1_medium 1.0 0.8 0.6 0.6		0.4 0.3 0.2 0.1 0.0 m	gemgem II8bII8b II70bII70b mismis mixmix openopenqwenqwen metric = Aborted experiment = 7_medium-themselves
0.0 gemgem 8b 8b 70b 70b mismis mixmix openopenqwenqwe metric = Success experiment = 0_easy 1.0 0.8 0.6 0.7 0.4 0.2	0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.6 0.6 0.6 0.6 0.2	2 0.	gemgem 8b 8b 70b 70b mismis mixmix openopenqwenqwen metric = Success experiment = 13_medium-more_turns 1.0 0.8 0.6 0.4 0.2	gemgem 8b 8b 70b 70b mismis mixmix openoperqwenqwen gemmetric = Success experiment = 1_medium 1.0 0.8 0.6 0.4 0.2	gem 8b 8b 70b 70b mismis mixmix openopenqwenqwen gemgel	0.2 m	gemgem II8bII8b II70bII70b mismis mixmix openoperqwenqwen metric = Success experiment = 7_medium-themselves
0.0 gemgem 8b 8b 170b 70b mismis mixmix openopengwenqwe metric = Lose experiment = 0_easy	0.40 0.35 0.04 0.35 0.02 0.25 0.25 0.02 0.02	0.0 gemgem 8b 8b 70b 70b mismis mixmix openopenqwenqwen metric = Lose experiment = 12_medium-less_turns 0.20 0.15 0.16 0.20 0.17 0.18 0.20 0.19 0.20 0.20 0.20 0.30	gemgem	gemgem 8b 8b 70b 70b mismis mixmix openopenqwenqwen metric = Lose experiment = 1_medium 0.20 0.15 0.10 0.05	gem 8b 8b 70b 70b mismis mixmix openopenqwenqwen metric = Lose experiment = 2_difficult 0.20	0.0 m	gemgem
gemgem 8b 8b 70b 70b mismis mixmix openopenqwenqwe metric = Main Score experiment = 0_easy 100	gemgem 8b 8b 70b 70b mismis mixmix openopengwenqwen metric = Main Score experiment = 10_medium-same_gender 100	gemgem 8b 8b 70b 70b mismis mixmix openopenqwenqwen metric = Main Score experiment = 12_medium-less_turns 100 80 80 60 60 60 60 60	gemgem	gemgem	gem	m II8bII8b II70bII70b mismis mixmix openopengwenqwen metric = Main Score experiment = 4_medium-reprompt 100 40 20 m II8bII8b II70bII70b mismis mixmix openopengwenqwen metric = Played experiment = 4_medium-reprompt	gemgem
netric = Played experiment = 0_easy 1.0 0.8 0.6 0.4 0.2 gemgem 8b 8b 70b 70b mis-mis mix-mix openopenqwenqwe model	1.0 0.8 0.6 0.4 0.2 gemgem 8b 8b 70b 70b mismis mixmix openopenqwenqwen model	metric = Played experiment = 12_medium-less_turns 1.0 88 0.6 4 0.4 2 gemgem 8b 8b 70b 70b mismis mixmix openopenqwenqwen model	metric = Played experiment = 13_medium-more_turns 1.0 0.8 0.6 0.4 0.2 gemgem 8b 8b 70b 70b mismis mixmix openopenqwenqwen model	metric = Played experiment = 1_medium 1.0 0.8 0.6 0.4 0.2 gemgem 8b 8b 70b 70b mismis mixmix openopenqwenqwen model	1.0	1.0 0.8 0.6 0.4 0.2 m	gemgem II8bII8b II70bII70b mismis model