A. Appendices

A.1. Additional Simulation Results

Table 7 shows the required labelling effort for worker capabilities sampled from $Unif(a, b)$ such that $a = 0.5$ and $b = 1$. In the evaluation, we analysed 3500 request samples for $\mu = 0.25$, 8500 for $\mu = 0.125$ and 30000 for $\mu = 0.0625$. Thus with the identical originally sampled request, we additionally evaluate 3500 samples for $\mu = 0.125$ and 15000 samples for $\mu = 0.0625$ to the samples presented in Table 2, since we have increased the variety of worker capabilities and thus require more requests for harder comparisons to reach a confident decision in each iteration.

$\mu = 0.25$		
Method	**Avg.**	**99% CI**
7 Workers	1051	1021-1083
5 Workers	897	870-922
Max 3 Workers	604	584-625
Fixed Worker	**550**	522-576
One Worker	**508**	490-526
$\mu = 0.125$		
Method	**Avg.**	**99% CI**
7 Workers	4468	4320-4618
5 Workers	3885	3764-3995
Max 3 Workers	2560	2477-2641
Fixed Worker	2129	2035-2217
One Worker	**1913**	1846-1975
$\mu = 0.0625$		
Method	**Avg.**	**99% CI**
7 Workers	15542	15143-15933
5 Workers	12750	12412-13064
Max 3 Workers	8374	8145-8624
Fixed Worker	**7025**	6660-7405
One Worker	**6563**	6331-6811

Table 7: Labelling effort for each labelling strategy averaged over 1000 iterations for three difficulty distributions. A decision is made with $1 - \delta = 0.999$ probability. Worker capabilities are sampled from $Unif(0.5, 1.0)$. The confidence intervals are computed with bootstrap resampling with 99% confidence.

A.2. Controlled Text Generation

Automatic evaluation Table 8 shows attribute matching accuracy for all generated sentences by each evaluated model. Note that the rightmost column indicates an unexpected low accuracy for the person number attribute compared to the results reported by Russo et al. (2020). Still, low person number accuracy has little or no impact on the configuration of defined comparison settings.

Generated sentences Tables 9, 10, and 11 summarise examples of generated sentences for all evaluated models according to supported attribute combinations.

Model	Sentiment	Tense	Person
V1	65.60%	39.48%	41.03%
V2	95.93%	96.53%	56.53%
CGA	98.68%	98.08%	56.02%

Table 8: Attribute matching accuracy (in %) of 6K generated sentences for each evaluated model.

A.3. Computing Infrastructure

The simulation framework is implemented with Python 3.6.12 and all experiments were executed on a Intel(R) Core(TM) i5-7360U CPU @ 2.30GHz CPU with 12 GB memory. All NLG models were trained on a single Titan XP GPU with 12 GB memory.

Sentence	Attributes
There are closed.	Present / Positive / Plural
I am always packed.	Present / Positive / Singular
The first time was packed.	Past / Positive / Plural
Oh and the food.	Past / Positive / Singular
Nothing is awesome.	Present / Negative / Plural
But i am going to.	Present / Negative / Singular
There were incredibly cold.	Past / Negative / Plural
Money at this place.	Past / Negative / Singular

Table 9: Examples of generated sentences from model: L_{ADV} + standard WD (V1), according to three input attributes (tense, sentiment, and pronoun).

Sentence	Attributes
The rooms are clean and nicely appointed.	Present / Positive / Plural
Everything else is great.	Present / Positive / Singular
All of the steaks were great.	Past / Positive / Plural
He also was very good.	Past / Positive / Singular
They are better than you.	Present / Negative / Plural
Do not waste your time here.	Present / Negative / Singular
The people that used to be the other reviews.	Past / Negative / Plural
I just went to the drive-thru and the service.	Past / Negative / Singular

Table 10: Examples of generated sentences from model: L_{ADV} + standard WD (V2), according to three input attributes (tense, sentiment, and pronoun).

Sentence	Attributes
They have a great selection of beers and they are always friendly.	Present / Positive / Plural
The food here is always good.	Present / Positive / Singular
This was my favorite restaurants.	Past / Positive / Plural
The best i had in phoenix.	Past / Positive / Singular
Worst wings i have ever had.	Present / Negative / Plural
This is a very expensive hotel.	Present / Negative / Singular
We were not happy with the food.	Past / Negative / Plural
The waiter did not know what i wanted to pay for a drink.	Past / Negative / Singular

Table 11: Examples of generated sentences from model: L_{CTX} + cyclical WD (CGA), according to three input attributes (tense, sentiment, and pronoun).