The effects of distance on NPI illusive effects in BERT

So Young Lee
Miami University
soyoung.lee@miamioh.edu

Mai Ha Vu
University of Toronto Mississauga
maiha.vu@utoronto.ca

Abstract

Previous studies have examined the syntactic capabilities of large pre-trained language models, such as BERT, by using stimuli from psycholinguistic studies. Studying well-known processing errors, such as Negative Polarity Item (NPI) illusive effects can reveal whether a model prioritizes linear or hierarchical information when processing language. Recent experiments have found that BERT is mildly susceptible to NPI illusion effects (Shin et al., 2023; Vu and Lee, 2022). We expand on these results by examining the effect of distance on the illusive effect, using and modifying stimuli from Parker and Phillips (2016). We also further tease apart whether the model is more affected by hierarchical distance or linear distance. We find that BERT is highly sensitive to syntactic hierarchical information: added hierarchical layers affected its processing capabilities compared to added linear distance.

1 Introduction

The recent proliferation of large language models, such as BERT (Devlin et al., 2019), have inspired investigations into these models' linguistic behavior. BERT, a ubiquitous baseline model in NLP experiments, provides context-based representation of text data. Numerous studies have attempted to reveal how accurately these language models simulate human behavior. One of the significant challenges in sentence processing is the resolution of long-distance syntactic dependencies between words or phrases.

Here we study the processing of long-distance dependencies in BERT to better understand the limitations and strengths of transformer-based language models. We focus on the illusion effect in Negative Polarity Item (NPI) licensing. We evaluate BERT on a psycholinguistic dataset that examines the effect of distance on NPI illusive effects (Parker and Phillips, 2016), and additionally investigate the nature of the distance that affects NPI illusive effects. We find that BERT is more affected by hierarchical than linear information.

2 Related work

2.1 Related psycholinguistic work

Syntactic illusive effects are a type of psycholinguistic phenomenon where humans are tricked to accept a syntactically ungrammatical sentence due to an interferer. Most notably, illusive effects have been shown in subject-verb agreement and in NPI licensing (Xiang et al., 2009; Parker and Phillips, 2016; Orth et al., 2021). Here we focus on NPI licensing effects.

English NPIs, such as *any* and *ever*, must be licensed by a c-commanding licensor, for example, negation (1) (Ladusaw, 1980). In other words, to be licensed, the NPI has to be in a particular syntactic relation with its licensor.

(1) a. No restaurant has ever gone out of business.
b. *Some restaurant has ever gone out of business.

At the same time, Xiang et al. (2009) have shown with EEG measurements that speakers can process unlicensed NPIs, such as *ever*, as if they were licensed, when there is an *intrusive* licensor (i.e., *no*) that precedes, but not structurally licenses the NPI, as in (2b). This suggests that linear word order can override syntactic information for humans.

(2) a. No restaurants [that the local newspapers have recommended in their dining reviews] have *ever* gone out of business
b. *The restaurants [that *no* local newspapers have recommended in their dining reviews] have *ever* gone out of business.

Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, pages 9443–9457

Illusive effects are affected by distance. Parker and Phillips (2016) have shown that the illusive effect is cancelled when the intrusive licensor and the NPI are more distant from each other (3a) compared to when they are closer (3b): that is, speakers correctly judge (3a) as ungrammatical, but not (3b).

(3) a. *[The journalists [that *no* editors recommended for the assignment] thought [that the readers would *ever* understand the complicated situation.]]

b. *[The journalists [that *no* editors recommended for the assignment] *ever* thought [that the readers would understand the complicated situation.]]

However, the material in Parker and Phillips (2016) conflates hierarchical and linear distance. The NPI in (3a) is linearly more distant from *no* than in (3b), measured by the number of lexical items between the NPI and *no*. It is also hierarchically more distant, as the NPI is one clause boundaries away from *no* (clause boundaries are shown with square brackets). In contrast, in (3b), the NPI is in the same clause as *no*.

2.2 Related NLP work

We build on previous work that has applied psycholinguistic tests to probe the syntactic capacity of pre-trained LMs. The earliest studies tested pre-trained, self-supervised LSTMs for their capability to detect syntactic dependencies (Linzen et al., 2016; Gulordava et al., 2018; Marvin and Linzen, 2018; Wilcox et al., 2018; Jumelet and Hupkes, 2018; Chowdhury and Zamparelli, 2018; Futrell et al., 2018, 2019), including NPI licensing (Marvin and Linzen, 2018; Jumelet and Hupkes, 2018; Futrell et al., 2018). Overall, these studies found that while LSTMs can detect syntactic dependencies remarkably well for most phenomena, they perform only at chance level when having to discriminate between the real and intrusive licensor for NPIs (Marvin and Linzen, 2018). Language model performance improved if it received structural supervision (Wilcox et al., 2019). These results together indicate that LSTMs learn a linearly based licensing rule for NPIs rather than a structural one, and so are consistent with human illusive effects in NPI-licensing.

Similarly, experiments on BERT found high performance for recognizing most syntactic dependencies (Goldberg, 2019), but have mixed results for NPI licensing (Warstadt et al., 2019; Warstadt and Bowman, 2020). In particular, Warstadt et al. (2019) found that BERT's performance greatly depended on licensing environment and evaluation method. Warstadt and Bowman (2020) furthermore found that when fine-tuned on a classifier to discriminate between sentences with licensed and unlicensed NPIs, BERT learned spurious rules that did not have to do with either hierarchical or linear generalization. Neither of these studies tested BERT explicitly for illusive effect, and did not use stimuli similar to Xiang et al. (2009) or Parker and Phillips (2016).

Most closely relevant to our work, Shin and Song (2021), Shin et al. (2023) and Vu and Lee (2022) tested the materials in Xiang et al. (2009) on pre-trained BERT. They found that BERT displayed no illusive effect when surprisal score for the licensor was measured, but it did have a tendency for the illusive effect when looking at its surprisal scores for the NPIs, as surprisal for 'ever' in illusive sentences was lower than in ungrammatical sentences and higher than in grammatical sentences. In comparison, Xiang et al. (2009) found that human subjects displayed a stronger illusive effect, as they had the same average ERP measures for 'ever' in the illusive and grammatical contexts.

The current study further contributes to these findings by also examining whether distance affects the syntactic capabilities of BERT, the same way it does for human processors, and further aims to tease apart whether hierarchical or linear distance matters more.

3 General methodology

3.1 Model

In all experiments, we test the capacities of a pre-trained $\text{BERT}_{\text{base}}$ model (uncased). BERT is a small (110 million parameter), bi-directional transformer model that has been pre-trained on masked token prediction and next sentence prediction tasks, on a corpus of English language Wikipedia and English language books (Devlin et al., 2019). We choose BERT specifically because it is a well-studied and open-source model. Additionally BERT's pre-training dataset is publically available on HuggingFace[1], which makes it possible to closely study the potential effects of the training dataset on model behavior in

[1] https://huggingface.co/datasets/legacy-datasets/wikipedia, https://huggingface.co/datasets/bookcorpus/bookcorpus

the future. We download the PyTorch implementation of BERT from https://github.com/huggingface/transformers and use code from Shin and Song (2021) to run our experiments. All experiments together took at most 1 GPU hour.

3.2 Materials and methods

For Experiment 1, we used the stimuli in Parker and Phillips (2016). For other experiments, we handcrafted the sentences by modifying the original stimuli in Experiment 1. We further describe these modifications in the relevant sections. We provide all our stimuli in the appendix.

In all stimuli, the NPI is replaced with a [MASK] token, as in (4). We also append the [CLS] and [SEP] tokens at the beginning and the end of the sentence, respectively, to mimic the pre-training conditions of BERT.

(4) [CLS] No journalists said that the author thought that the readers would [MASK] understand the complicated situation [SEP].

In all experiments, we measure the surprisal score for *ever* following the methodology in Shin and Song (2021) and Shin et al. (2023). Namely, we calculate the negative log probability of 'ever' in place of the [MASK] token given its context in the softmax layer. High surprisal scores in language models have been argued to correlate with processing effort in humans (Levy, 2008; Michaelov and Bergen, 2022). Beyond the mean surprisal scores, we report on *accuracy* following the cloze test in Warstadt et al. (2019). This is the percentage of times when BERT outputs lower surprisal scores to 'ever' in pairwise comparisons of minimal pairs belonging to different conditions.

By reporting on both mean surprisal scores and accuracy based on pairwise comparisons, we aimed to provide a comprehensive analysis that captures both global trends and specific contrasts. This dual approach made us ensure that we do not overlook any potential effects that could be critical for understanding the processing mechanisms at play. Mean surprisal scores give us an overall sense of the processing difficulty associated with different conditions, which is important for establishing baseline differences. However, the statistical tests with mean surprisal might fail to capture subtler distinctions between specific sentence pairs within the same condition, especially if the effect sizes are small, and given the small size of our dataset. Pairwise comparisons allow us to delve deeper into specific contrasts, uncovering finer distinctions that might be relevant for theoretical and practical reasons. For example, in some of our experiments we find that while there is no statistical difference in mean surprisal score, the surprisal scores for sentence in one condition are lower than the scores for another condition at almost 100% of the time.

Shin et al. (2023), following Warstadt et al. (2019), measure surprisal scores for both the NPI and the licensor positions. Since Xiang et al. (2009) measured human EEG reaction at *ever* only, we expect surprisal scores at the NPI position to be comparable to human results. This is also born out in previous results: BERT shows the same tendency for an 'illusive effect' for surprisal scores at the NPI position, but not at the licensor positions (Vu and Lee, 2022; Shin et al., 2023). Consequently, we only measure surprisal scores at the NPI position in this paper.

In all three experiments, we study the effects of various types of distances on the NPI illusive effect. To this end, we compare the surprisal scores for 'ever' in grammatical (gr), illusive (ill), and ungrammatical (ungr) sentences.[2] Across the board, we expect the surprisal scores for 'ever' to be lowest in grammatical sentences. In comparing grammatical vs. ungrammatical sentences, as well as grammatical vs. illusive sentences, we expect near 100% accuracy – that BERT would almost always assign a lower surprisal score to grammatical sentences compared to either of the other conditions. If BERT is not affected by the illusive effect, we expect the surprisal scores in the illusive and the ungrammatical conditions to not differ significantly. In that case, the comparison of illusive and ungrammatical sentences would yield about 50% accuracy, that

[2]Our analysis follows established methodologies in psycholinguistics research on NPI illusions, as demonstrated by (Parker and Phillips, 2016). They separated data into short and long conditions, allowing for clearer identification of main effects. Similarly, we divided our data by target-attractor distance (long vs. short) and conducted separate statistical analyses for each subset. This allowed us to identify main effects more effectively. The linear mixed-effects model for each distance type was specified as: *lm_L1 <- lmer(surprisal ~ licensor * position + (1 | item), data = L1R)*. For post hoc comparisons, we used: *emmeans(lm_L1, pairwise ~ licensor + position, adjust = "tukey")*. To explore interactions between distance and sentence type, we used a combined model: *model <- lmer(surprisal ~ grammaticality * distance + (1 | item), data = combined_data)*. This analysis showed significant main effects for grammaticality and distance, as well as a significant interaction.

is BERT assigns lower surprisal scores to 'ever' in illusive sentences compared to ungrammatical sentences at chance level. On the other hand, if BERT is affected by the illusive effect, we expect lower surprisal scores for the illusive sentences compared to the ungrammatical sentences at a more than chance level. In this case, accuracy should be greater than 50%.

4 Experiment 1

4.1 Stimuli

To set a baseline for the effect of distance on NPI illusive effects, we used the stimuli from the psycholinguistics study by Parker and Phillips (2016). One example from the data set is in Table 1. To test the distance effect, the data set consisted of 6 conditions with two factors crossed: the licensing of the NPI (Grammatical vs Illusive vs Ungrammatical) and the distance type (Short vs. Long). In each condition, there were 36 different sentences. In the short distance condition, the average number of words between negation and the NPI was 8.17 for grammatical sentences and 5.12 for illusive sentences. In the long distance condition, it was 13.22 words for grammatical sentences, and 10.22 for illusive sentences.

If BERT is to behave similarly to human subjects, we expect it to output the same surprisal scores for illusive and ungrammatical sentences in the long distance condition, but not in the short distance condition.

4.2 Results

Figures 1 and 2 show the average surprisal scores for each condition. Overall, the ungrammatical conditions yielded significantly higher scores (M=12.25) than the grammatical conditions (M=*4.62*) regardless of the distance between the NPI and the licensor (Linear mixed effect regression model (lmer): $p < .0001$). There was no significant difference between the illusive condition (M=11.25) and the ungrammatical conditions (M=11.465) when the NPI and the licensor were at a long distance from each other (Tukey post hoc test: $p = 0.9272$). In the short distance condition, the surprisal score for NPIs in illusive sentences (M=9.98) was lower compared to an ungrammatical sentence (13.03))(Tukey post hoc test:$p <$ .0001).

Accuracy percentages are summarized in Table 2. As expected, BERT assigned a lower surprisal

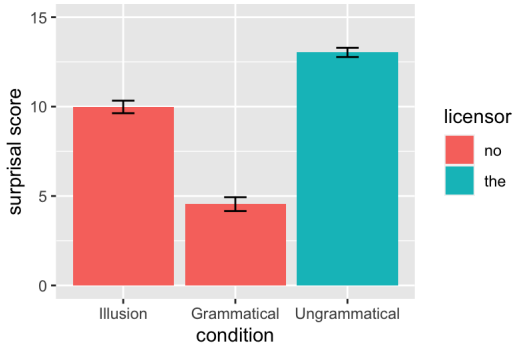

Figure 1: The average suprisal scores in the short distance conditions in Experiment 1

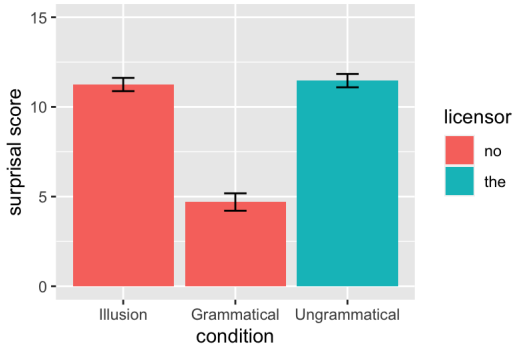

Figure 2: The average suprisal scores in the long-distance conditions in Experiment 1

score to *ever* in grammatical sentences compared to the others in both distance conditions at nearly 100% of the time. BERT showed consistently a lower surprisal score for illusive sentences compared to ungrammatical sentences, especially in the short distance condition. This implies that even though the differences in surprisal scores of illusive conditions and ungrammatical conditions might be subtle, BERT generally preferred the existence of a potential licensor. Interestingly, this preference became weaker in long-distance conditions. It confirms that BERT is susceptible to the illusive effect and a longer distance between negation and NPI weakens this effect for BERT.

These results are similar, but not identical to the results reported in the psycholinguistics study by Parker and Phillips (2016), who found that in the short distance conditions, the illusive sentences pattern with the grammatical sentences, whereas in the long distance conditions, they patterned together with the ungrammatical sentences. As in previous studies (Shin et al., 2023; Vu and Lee, 2022), BERT did not display a full illusive effect in the short distance condition: its surprisal scores for the

Distance	NPI licensing	Example
Short	grammatical	No journalists [that the editors recommended for the assignment [MASK]] thought [that the readers would understand the complicated situation].
Short	illusive	The journalists that [no editors recommended for the assignment [MASK]] thought [that the readers would understand the complicated situation].
Short	ungrammatical	The journalists [that the editors recommended for the assignment [MASK]] thought [that the readers would understand the complicated situation].
Long	grammatical	No journalists [that the editors recommended for the assignment] thought [that the readers would [MASK] understand the complicated situation].
Long	illusive	The journalists [that no editors recommended for the assignment] thought [that the readers would [MASK] understand the complicated situation].
Long	ungrammatical	The journalists [that the editors recommended for the assignment] thought [that the readers would [MASK] understand the complicated situation].

Table 1: Example sentences for each condition. We indicate the negative licensor in red, and clause boundaries with square brackets.

	Short	Long
grammatical < ungrammatical	100%	100%
grammatical < illusive	97.2%	100%
illusive < ungrammatical	97.2%	72.2%

Table 2: Pairwise comparison of surprisal scores in Experiment 1

illusive sentences were lower than for ungrammatical sentences, but were not the same as the scores for grammatical sentences. In the long distance condition, however, the illusive effect completely disappeared, as BERT output the same surprisal scores for illusive and ungrammatical sentences. Thus, BERT's outputs matched human results in the long distance condition.

Since the NPI is not only linearly farther located but also hierarchically deeper in the long distance than the short distance condition, it is unclear whether the trigger of the illusive effect hinges on hierarchical distance or linear distance information. In the next experiments, we tease apart BERT's sensitivity to linear information compared to hierarchical information.

5 Experiment 2

5.1 Stimuli

In Experiment 2, we created a new data set by modifying the data set in Experiment 1, where the distance between the negation and NPI was increased either due to added hierarchical depth and linear distance (hierarchical condition) or due to added linear distance alone (linear condition) (Table 3). We accomplished this by adding adjunctive modifiers in the linear distance condition. For example, we inserted two-word modifiers such as "American Broadcast" and "advanced younger" before nouns in the relative clause subject position and the embedded complement clause subject position, bolded in Table 3. For the hierarchical distance condition, instead of adding modifiers, we added another layer of embedded clause as shown bolded in Table 3. Since in both distance conditions we always added four words, the linear distance between negation and NPI was the same across hierarchical and linear distance conditions: an average of 17.28 words in the grammatical condition, and 14.28 words in the illusive condition.

Parker and Phillips (2016) theorized that the illusive effect switches off in long-distance conditions due to increased time: as humans have more time to process the sentence, the less likely they are to be subject to illusion. If this is true for BERT also, then there should be no difference in its surprisal scores between the linear and hierarchical conditions.

5.2 Results

The results suggest that BERT is more affected by hierarchical distance than linear distance (Figure

Distance type	NPI licensing	Example
Linear	grammatical	No journalists [that the **American broadcast** editors recommended for the assignment] thought [that the **advanced younger** readers would [MASK] understand the complicated situation].
Linear	illusive	The journalists [that no **American broadcast** editors recommended for the assignment] thought [that the **advanced younger** readers would [MASK] understand the complicated situation].
Linear	ungrammatical	The journalists [that the **American broadcast** editors recommended for the assignment] thought [that the **advanced younger** readers would [MASK] understand the complicated situation].
Hierarchical	grammatical	No journalists [that the editors recommended for the assignment] said [**that the author thought** [that the readers would [MASK] understand the complicated situation]]. (17.28)
Hierarchical	illusive	The journalists [that no editors recommended for the assignment] said [**that the author thought** [that the readers would [MASK] understand the complicated situation]].
Hierarchical	ungrammatical	The journalists [that the editors recommended for the assignment] said [**that the author thought** [that the readers would [MASK] understand the complicated situation]].

Table 3: Example sentences for each condition in Experiment 2. We indicate the words we have added in **bold**, the licensor in red, and clause boundaries with square brackets.

3 and Figure 4). In particular, the surprisal score for the grammatical sentences surged in the hierarchical distance condition (M=8.97) (Figure 3) compared to the linear distance condition (M=4.76, p <.0001) (Figure 4).

On the other hand, there was no significant difference in the average surprisal scores between the illusive and ungrammatical sentences in either distance conditions. In the linear distance condition, surprisal scores for illusive sentences (M=10.15) were not significantly different from those for ungrammatical sentences (M=10.61) (lmer Tukey post hoc test: p = 0.49). The same was true in the hierarchical distance condition (M=10.878 for illusive sentences, M=10.91 for ungrammatical sentences, p = 0.99). This would suggest a lack of illusive effect in both conditions.

However, pairwise comparision of surprisal scores in illusive sentences to those in ungrammatical sentences gives a more nuanced picture. In the linear distance condition, the illusive sentences were preferred 72% of the time compared to the ungrammatical sentences, while this preference completely disappeared in the hierarchical condition at 44.4%. This implies that the illusive effect in BERT is completely eroded by increasing hierarchical distance to two embedded clauses, but not by increasing only linear distance.

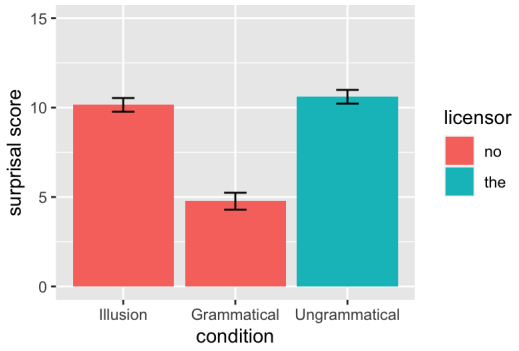

Figure 3: The average suprisal scores in the linear distance conditions in Experiment 2

6 Experiment 3

In Experiment 2, we found that hierarchical distance affected BERT's capacity to distinguish between grammatical and ungrammatical sentences, more so than linear distance when looking at accuracy, but not when comparing average surprisal scores. We suspect that the size of the illusive effect in the linear distance condition was almost undetectable due to the hierarchical distance between

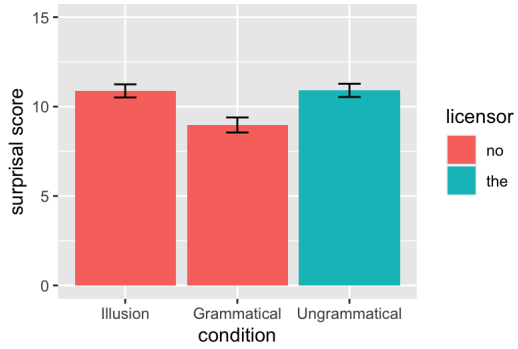

Figure 4: The average suprisal scores in the hierarchical distance conditions in Experiment 2

	Linear	Hierarchical
gr < ungr	100%	100%
gr < ill	100%	97.2%
ill < ungr	72.2%	44.4%

Table 4: Pairwise comparison of surprisal scores in Experiment 2

negation and NPI being too long in both conditions. To address this problem, we tested BERT's performance with reduced hierarchical and linear distances.

6.1 Stimuli

We modified the stimuli in Experiment 2 to shorten the sentences across the board, both linearly and hierarchically. Specifically, we deleted one layer of complement clause from all conditions, and added two modifiers into the relative clause in the linear conditions so that the distance between negation and NPI would stay constant between linear and hierarchical conditions. Compared to Experiment 2, 4 words on average were reduced in Experiment 3 (Table 5), resulting in 13.17 words between negation and the NPI in grammatical conditions, and 10.17 words in illusive conditions.

Based on the results in Experiment 2, we expect to see a stronger illusive effect in the linear distance condition than in the hierarchical distance condition.

6.2 Results

As expected, BERT shows a stronger illusive effect in the linear than the hierarchical condition. In the linear condition, the mean surprisal score (M=8.67) for the illusive sentences is significantly lower compared to the mean surprisal score (M=10.97) in the ungrammatical condition (lmer Tukey post hoc test: p <.0001) (Figure 5). In comparison, there is no significant difference between the illusive sentences (M=10.79) and the ungrammatical sentences (M=11.235) (lmer Tukey post hoc test: p = 0.32) in the hierarchical condition (Figure 6). This implies that the added hierarchical layer reduces the illusive effect.

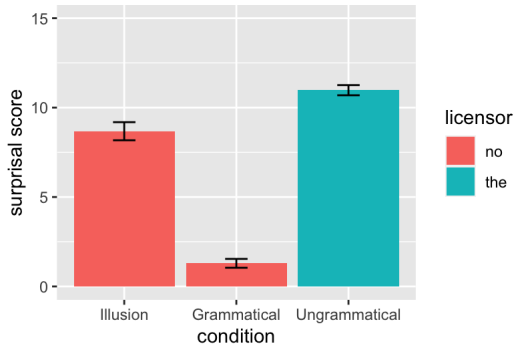

Figure 5: The average suprisal scores in the linear distance conditions in Experiment 3

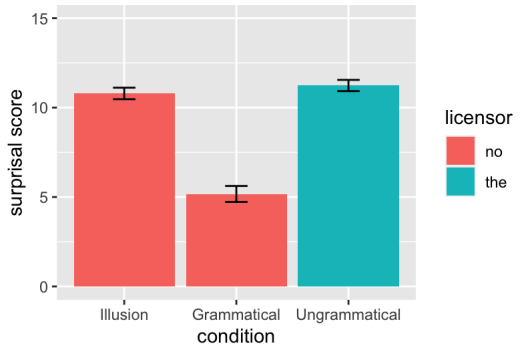

Figure 6: The average suprisal scores in the hierarchical distance conditions in Experiment 3

When looking at accuracy scores, BERT assigns lower surprisal scores to the illusive sentences in 91.6% of the cases compared to the ungrammatical sentences in the linear distance condition, but this preference is weakened in the hierarchical distance condition to 72.2%. These results confirm that switching off the illusive effects are closely related to the hierarchical rather than linear distance of NPI dependents.

7 Discussion

Our study shows three main results. First, we have replicated the results in Parker and Phillips (2016) that have shown that NPI illusive effects are modulated by the distance between the licensor and the NPI.

Distance	NPI licensing	Example
Linear	grammatical	*No* journalist that the **American broadcast** editor **sincerely** recommended for the **interview** assignment would [MASK] understand the complicated situation.
Linear	illusive	The journalist that *no* **American broadcast** editor **sincerely** recommended for the **interview** assignment would [MASK] understand the complicated situation.
Linear	ungrammatical	The journalist that the **American broadcast** editor **sincerely** recommended for the **interview** assignment would [MASK] understand the complicated situation.
Hierarchical	grammatical	*No* journalist that the editor recommended for the assignment **thought that the readers** would [MASK] understand the complicated situation.
Hierarchical	illusive	The journalist that *no* editor recommended for the assignment **thought that the readers** would [MASK] understand the complicated situation.
Hierarchical	ungrammatical	The journalist that the editor recommended for the assignment **thought that the readers** would [MASK] understand the complicated situation.

Table 5: Example sentences for each condition in Experiment 3. We indicate the words that differ between the linear and hierarchical conditions in **bold**, the licensor in red, and clause boundaries with square brackets.

	Linear	Hierarchical
gr < ungr	100%	100%
gr < ill	100%	100%
ill < ungr	91.6%	72.2%

Table 6: Pairwise comparison of surprisal scores in Experiment 3

Second, we have teased apart linear and hierarchical distance and found that BERT's surprisal score to licensed NPIs worsens with increased hierarchical distance, but not with increased linear distance. Table 7 shows that surprisal scores in grammatical and illusive sentences were the lowest when there were no additional embedded clauses between negation and NPI. With one embedded clause in-between negation and NPI, surprisal scores were the same, regardless of the number of words between the two. In comparison, ungrammatical sentences yielded the same surprisal score in all conditions.

Finally, the NPI illusive effect was sharper with fewer embedded complement clauses but not with fewer words, further confirming BERT's sensitivity to the hierarchical distance over linear distance. Our results thus add to our knowledge about pretrained BERT's sensitivity to hierarchical versus linear information.

We have reported both mean surprisal scores for each condition and pairwise comparison between the conditions. We found that each type of measurement gave a slightly different picture of BERT's syntactic capabilities. In particular, in Experiment 2, mean surprisal scores showed no illusive effect in either the linear or hierarchical distance condition. At the same time, pairwise comparison between the illusive and ungrammatical sentences revealed that in fact, BERT assigned lower surprisal scores to illusive sentences compared to ungrammatical sentences. This result suggests that there was some tendency for illusive effect even in sentences with one complement clause, but it was undetectable when comparing mean surprisal scores. This highlights the necessity for using multiple diagnostics when studying language model capability, as noted by Warstadt et al. (2019).

Our results overall are mixed about BERT's capabilities for learning syntactic structure. On the one hand, the fact that BERT was susceptible to illusive effects suggests that at least for NPI licensing, the model has relied to some extent on

Distance between licensor and NPI		Mean surprisal scores			
# of clauses	# of words Gr/Ill	Grammatical	Illusive	Ungrammatical	ill < ungr.
0	13.17/10.17	1.29	8.68	10.97	91.6%
1	13.17/10.17	5.17	10.79	11.24	72.2%
1	17.28/14.28	4.77	10.15	10.61	72.2%
2	17.28/14.28	8.97	10.88	10.91	44.4%

Table 7: Summary results of Experiments 2 and 3.

a linear generalization rather than on the correct structural generalization. At the same time, previous experiments on BERT have already suggested that NPI licensing is exceptional, as BERT was able to make the correct structural generalization for subject-verb agreement (Goldberg, 2019; Cho, 2023). This is, in particular, surprising since illusive effects also apply to subject-verb agreement in humans (Wagers et al., 2009). Further research on illusive effects in BERT is needed to understand the asymmetry between NPI-licensing and other long-distance dependencies.

At the same time, BERT was more affected by added hierarchical structure than by added linear information. Added hierarchical distance increased the surprisal score for NPIs even when they were licensed and minimized illusive effects as well. This result suggests that BERT is at least sensitive to hierarchical distance in the form of embedded complement clauses when evaluating long-distance dependencies. It could be interesting to see if other types of added hierarchical information, such as nested relative clauses would have the same effect.

At a more general level, our study raises interesting questions about the desired target behavior of language models: should they capture human behavior, including grammatical illusions, or should they capture pure grammar, devoid of any performance errors? The desired outcome will be dependent on application, and practicioners should be aware ahead of time about the differences.

It also remains an open question as to *why* BERT has such a tendency to human-like performance errors, even though its bidirectional transformer architecture is not a plausible model of human linguistic processing (as opposed to autoregressive models). There are multiple possible reasons. The simplest one is that the errors are baked in the pretraining data. A thorough corpus examination of the pre-training data is necessary to rule this out. Another possibility is that learning English grammar involves generalizing to illusive errors, and BERT in fact is generalizing to the correct grammar. This option, however, is implausible, as there are numerous linguistic phenomena where BERT does not reflect human processing behavior (Ettinger, 2020; Cho, 2023). A third remaining possibility is that the architecture or pre-training objectives of BERT result in an output that converges to the output of human language processing behavior when it comes to specifically NPI licensing effects. This possibility also requires further linguistically informed investigations, some of which we have outlined above.

8 Conclusion

In this paper, we have conducted experiments inspired by psycholinguistic studies to examine the sensitivity of the pre-trained BERT model to hierarchical information. In particular, we studied the effect of distance on NPI licensing illusions in pre-trained BERT and designed our own stimuli to tease apart whether it is the hierarchical or linear distance that mattered. We found that BERT in fact displays some illusive effects, meaning that it did not perfectly learn the correct structural generalization for NPI-licensing, but at the same time remained sensitive to hierarchical distance and not linear distance.

9 Limitations

Because we adopted materials from a psycholinguistic experiment, we tested a very small number of sentences on BERT: only 36 sentences for each condition in all three experiments. Given that similar studies on BERT usually use thousands of sentences in their stimuli, our results are limited in comparison. We also only tested English stimuli, and our results might not hold for other languages.

References

Ye-eun Cho. 2023. Grammatical illusions in BERT: Attraction effects of subject-verb agreement and

reflexive-antecedent dependencies. *Linguistic Research*, 40(2):317–352.

Shammur Absar Chowdhury and Roberto Zamparelli. 2018. RNN Simulations of Grammaticality Judgments on Long-distance Dependencies. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 133–144, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Allyson Ettinger. 2020. What BERT Is Not: Lessons from a New Suite of Psycholinguistic Diagnostics for Language Models. *Transactions of the Association for Computational Linguistics*, 8:34–48.

Richard Futrell, Ethan Wilcox, Takashi Morita, and Roger Levy. 2018. RNNs as psycholinguistic subjects: Syntactic state and grammatical dependency. *Preprint*, arxiv:1809.01329.

Richard Futrell, Ethan Wilcox, Takashi Morita, Peng Qian, Miguel Ballesteros, and Roger Levy. 2019. Neural language models as psycholinguistic subjects: Representations of syntactic state. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 32–42, Minneapolis, Minnesota. Association for Computational Linguistics.

Yoav Goldberg. 2019. Assessing BERT's Syntactic Abilities. *arXiv:1901.05287 [cs]*.

Kristina Gulordava, Piotr Bojanowski, Edouard Grave, Tal Linzen, and Marco Baroni. 2018. Colorless green recurrent networks dream hierarchically. *Preprint*, arxiv:1803.11138.

Jaap Jumelet and Dieuwke Hupkes. 2018. Do Language Models Understand Anything? On the Ability of LSTMs to Understand Negative Polarity Items. *arXiv:1808.10627 [cs]*.

William Ladusaw. 1980. *Polarity Sensitivity as Inherent Scope Relations*. Ph.D. thesis, University of Iowa.

Roger Levy. 2008. Expectation-based syntactic comprehension. *Cognition*, 106(3):1126–1177.

Tal Linzen, Emmanuel Dupoux, and Yoav Goldberg. 2016. Assessing the Ability of LSTMs to Learn Syntax-Sensitive Dependencies. *Transactions of the Association for Computational Linguistics*, 4:521–535.

Rebecca Marvin and Tal Linzen. 2018. Targeted Syntactic Evaluation of Language Models. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1192–1202, Brussels, Belgium. Association for Computational Linguistics.

James Michaelov and Ben Bergen. 2022. The more human-like the language model, the more surprisal is the best predictor of N400 amplitude.

Wesley Orth, Masaya Yoshida, and Shayne Sloggett. 2021. Negative polarity item (NPI) illusion is a quantification phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 47(6):906–947.

Dan Parker and Colin Phillips. 2016. Negative polarity illusions and the format of hierarchical encodings in memory. *Cognition*, 157:321–339.

Unsub Shin and Sanghoun Song. 2021. BERT, a deep-learning language model, learns NPI licensing but does not suffer from NPI illusion.

Unsub Shin, Eunkyung Yi, and Sanghoun Song. 2023. Investigating a neural language model's replicability of psycholinguistic experiments: A case study of NPI licensing. *Frontiers in Psychology*, 14.

Mai Ha Vu and So Young Lee. 2022. Comparing neural-network based language models to human sentence processing: Choice of task matters.

Matthew W. Wagers, Ellen F. Lau, and Colin Phillips. 2009. Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, 61(2):206–237.

Alex Warstadt and Samuel R. Bowman. 2020. Can neural networks acquire a structural bias from raw linguistic data? *arXiv:2007.06761 [cs]*.

Alex Warstadt, Yu Cao, Ioana Grosu, Wei Peng, Hagen Blix, Yining Nie, Anna Alsop, Shikha Bordia, Haokun Liu, Alicia Parrish, Sheng-Fu Wang, Jason Phang, Anhad Mohananey, Phu Mon Htut, Paloma Jeretič, and Samuel R. Bowman. 2019. Investigating BERT's Knowledge of Language: Five Analysis Methods with NPIs. *arXiv:1909.02597 [cs]*.

Ethan Wilcox, Roger Levy, Takashi Morita, and Richard Futrell. 2018. What do RNN Language Models Learn about Filler–Gap Dependencies? In *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pages 211–221, Brussels, Belgium. Association for Computational Linguistics.

Ethan Wilcox, Peng Qian, Richard Futrell, Miguel Ballesteros, and Roger Levy. 2019. Structural Supervision Improves Learning of Non-Local Grammatical Dependencies. *arXiv:1903.00943 [cs]*.

Ming Xiang, Brian Dillon, and Colin Phillips. 2009. Illusory licensing effects across dependency types: ERP evidence. *Brain and Language*, 108(1):40–55.

A Experiment 1

A.1 Parker - long

1. No/The journalists that no/the editors recommended for the assignment thought that the readers would [MASK] understand the complicated situation.
2. No/The investors that no/the businessmen informed about the recession predicted that the stock would [MASK] drop below the initial offering price.
3. No/The ambassadors that no/the diplomats consulted about the treaty thought that the journalists would [MASK] reveal the truth about election.
4. No/The professors that no/the students trusted at the college thought that the administrators would [MASK] increase the yearly tuition.
5. No/The customers that no/the salesmen assisted in the showroom thought that the manager would [MASK] consider their lowest offer.
6. No/The protestors that no/the journalists interviewed at the rally implied that the legislators could [MASK] pass the necessary laws.
7. No/The senators that no/the corporations supported with campaign donations thought that the lobbyists would [MASK] accept the sly bribe.
8. No/The lawyers that no/the policemen respected after the trial anticipated that the judge would [MASK] deliver such a harsh sentence.
9. No/The students that no/the teachers punished for bad behavior expected that the principal would [MASK] hear about the incident.
10. No/The accountants that no/the inspectors audited in the past year thought that the IRS would [MASK] find out about the scandal.
11. No/The actors that no/the fans recognized at the after-party believed that the paparazzi would [MASK] find out about the affair.
12. No/The teachers that no/the parents recommended for the award expected that the faculty would [MASK] receive a huge pay raise.
13. No/The students that no/the librarians could help in the afternoon expected that the teacher would [MASK] extend the approaching deadline.
14. No/The children that no/the bullies picked on at recess thought that the teacher would [MASK] give such a harsh punishment.
15. No/The criminals that no/the policemen could catch in the raid expected that the judge would [MASK] accept a plea bargain.
16. No/The employees that no/the managers recommended for the promotion anticipated that the boss would [MASK] ask such difficult questions.
17. No/The investors that no/the managers trusted with the money thought that the stock prices would [MASK] increase drastically overnight.
18. No/The candidates that no/the voters supported during the election believed that the mayor would [MASK] be re-elected for a second term.
19. No/The doctors that no/the nurses assisted during the operation assumed that the insurance company would [MASK] cover the hospital bill.
20. No/The criminals that no/the witnesses could identify in the courtroom suspected that the jury would [MASK] find out about the evidence.
21. No/The actresses that no/the critics liked in the movie expected that the director would [MASK] win a prestigious award.
22. No/The legislators that no/the congressmen consulted about the proposal suggested that the government should [MASK] increase military spending for the war.
23. No/The politicians that no/the journalists endorsed in the newspaper thought that the election would [MASK] cause such a huge scandal.
24. No/The teenagers that no/the parents trusted with a car believed that an accident could [MASK] happen in sunny weather.
25. No/The survivors that no/the medics could treat with a first-aid kit expected that a full recovery would [MASK] be possible in one month.
26. No/The athletes that no/the coaches recruited for the team anticipated that the scandal would [MASK] receive so much media coverage.
27. No/The congressmen that no/the citizens supported during the crisis assumed that the treasury would [MASK] lower the national debt.
28. No/The professors that no/the students visited during office hours anticipated that the exam would [MASK] be so difficult for the class.
29. No/The actors that no/the judges nominated for an award expected that the movie would [MASK] be such a blockbuster hit.
30. No/The actresses that no/the directors auditioned for the role thought that the movie would [MASK] cause so much controversy.
31. No/The champions that no/the competitors defeated in the race expected that that the coach would [MASK] receive a life-time achievement award.
32. No/The artists that no/the collectors regarded very highly suggested that the gallery should [MASK] buy cheap frames for the expensive paintings.
33. No/The scientists that no/the reporters cited in the story believed that the public would [MASK] care about the new discovery.
34. No/The teenagers that no/the teachers motivated before the test claimed that the parents should [MASK] offer more help on assignments.
35. No/The students that no/the professors could tutor on the weekend thought that the assignments should [MASK] be more than seven pages.
36. No/The protestors that no/the reporters interviewed on live television expected that the mayor would [MASK] give in to the numerous demands.

A.2 Parker - short

1. No/The journalists that no/the editors recommended for the assignment [MASK] thought that the readers would understand the complicated situation.
2. No/The investors that no/the businessmen informed about the recession [MASK] predicted that the stock would drop below the initial offering price.
3. No/The ambassadors that no/the diplomats consulted about the treaty [MASK] thought that the journalists would reveal the truth about election.
4. No/The professors that no/the students trusted at the college [MASK] thought that the administrators would increase the yearly tuition.
5. No/The customers that no/the salesmen assisted in the showroom [MASK] thought that the manager would consider their lowest offer.
6. No/The protestors that no/the journalists interviewed at the rally [MASK] implied that the legislators could pass the necessary laws.
7. No/The senators that no/the corporations supported with campaign donations [MASK] thought that the lobbyists would accept the sly bribe.
8. No/The lawyers that no/the policemen respected after the trial [MASK] anticipated that the judge would deliver such a harsh sentence.
9. No/The students that no/the teachers punished for bad behavior [MASK] expected that the principal would hear about the incident.
10. No/The accountants that no/the inspectors audited in the past year [MASK] thought that the IRS would find out about the scandal.
11. No/The actors that no/the fans recognized at the after-party [MASK] believed that the paparazzi would find out about the affair.
12. No/The teachers that no/the parents recommended for the award [MASK] expected that the faculty would receive a huge pay raise.
13. No/The students that no/the librarians could help in the afternoon [MASK] expected that the teacher would extend the approaching deadline.
14. No/The children that no/the bullies picked on at recess [MASK] thought that the teacher would give such a harsh punishment.

15. No/The criminals that no/the policemen could catch in the raid [MASK] expected that the judge would accept a plea bargain.

16. No/The employees that no/the managers recommended for the promotion [MASK] anticipated that the boss would ask such difficult questions.

17. No/The investors that no/the managers trusted with the money [MASK] thought that the stock prices would increase drastically overnight.

18. No/The candidates that no/the voters supported during the election [MASK] believed that the mayor would be re-elected for a second term.

19. No/The doctors that no/the nurses assisted during the operation [MASK] assumed that the insurance company would cover the hospital bill.

20. No/The criminals that no/the witnesses could identify in the courtroom [MASK] suspected that the jury would find out about the evidence.

21. No/The actresses that no/the critics liked in the movie [MASK] expected that the director would win a prestigious award.

22. No/The legislators that no/the congressmen consulted about the proposal [MASK] suggested that the government should increase military spending for the war.

23. No/The politicians that no/the journalists endorsed in the newspaper [MASK] thought that the election would cause such a huge scandal.

24. No/The teenagers that no/the parents trusted with a car [MASK] believed that an accident could happen in sunny weather.

25. No/The survivors that no/the medics could treat with a first-aid kit [MASK] expected that a full recovery would be possible in one month.

26. No/The athletes that no/the coaches recruited for the team [MASK] anticipated that the scandal would receive so much media coverage.

27. No/The congressmen that no/the citizens supported during the crisis [MASK] assumed that the treasury would lower the national debt.

28. No/The professors that no/the students visited during office hours [MASK] anticipated that the exam would be so difficult for the class.

29. No/The actors that no/the judges nominated for an award [MASK] expected that the movie would be such a blockbuster hit.

30. No/The actresses that no/the directors auditioned for the role [MASK] thought that the movie would cause so much controversy.

31. No/The champions that no/the competitors defeated in the race [MASK] expected that that the coach would receive a life-time achievement award.

32. No/The artists that no/the collectors regarded very highly [MASK] suggested that the gallery should buy cheap frames for the expensive paintings.

33. No/The scientists that no/the reporters cited in the story [MASK] believed that the public would care about the new discovery.

34. No/The teenagers that no/the teachers motivated before the test [MASK] claimed that the parents should offer more help on assignments.

35. No/The students that no/the professors could tutor on the weekend [MASK] thought that the assignments should be more than seven pages.

36. No/The protestors that no/the reporters interviewed on live television [MASK] expected that the mayor would give in to the numerous demands.

B Experiment 2

B.1 Hierarchical distance

1. No/The journalist that no/the editor recommended for the assignment said that the author thought that the readers would [MASK] understand the complicated situation.

2. No/The investor that no/the businessman informed about the recession said that the expert predicted that the stock would [MASK] drop below the initial offering price.

3. No/The ambassador that no/the diplomat consulted about the treaty said that the government thought that the journalists would [MASK] reveal the truth about election.

4. No/The professor that no/the student trusted at the college said that the dean thought that the administrators would [MASK] increase the yearly tuition.

5. No/The customer that no/the salesman assisted in the showroom said that the headhunter thought that the manager would [MASK] consider their lowest offer.

6. No/The protestor that no/the journalist interviewed at the rally said that the senator implied that the legislators could [MASK] pass the necessary laws.

7. No/The senator that no/the corporation supported with campaign donations suggested that the businessman thought that the lobbyists would [MASK] accept the sly bribe.

8. No/The lawyer that no/the policeman respected after the trial thought that the attorney anticipated that the judge would [MASK] deliver such a harsh sentence.

9. No/The student that no/the teacher punished for bad behavior thought that the parents expected that the principal would [MASK] hear about the incident.

10. No/The accountant that no/the inspector audited in the past year expected that the boss thought that the agent would [MASK] find out about the scandal.

11. No/The actor that no/the fan recognized at the after-party thought that the manager believed that the paparazzi would [MASK] find out about the affair.

12. No/The teacher that no/the parent recommended for the award thought that the student expected that the faculty would [MASK] receive a huge pay raise.

13. No/The student that no/the librarian could help in the afternoon claimed that the classmate expected that the teacher would [MASK] extend the approaching deadline.

14. No/The child that no/the bully picked on at recess claimed that the teacher thought that the teacher would [MASK] give such a harsh punishment.

15. No/The criminal that no/the policeman could catch in the raid claimed that the lawyer expected that the judge would [MASK] accept a plea bargain.

16. No/The employee that no/the manager recommended for the promotion claimed that the mentors anticipated that the boss would [MASK] ask such difficult questions.

17. No/The investor that no/the manager trusted with the money claimed that the CEO thought that the stock prices would [MASK] increase drastically overnight.

18. No/The candidate that no/the voter supported during the election claimed that the media believed that the mayor would [MASK] be re-elected for a second term.

19. No/The doctor that no/the nurse assisted during the operation suggested that the government assumed that the insurance company would [MASK] cover the hospital bill.

20. No/The criminal that no/the witness could identify in the courtroom suggested that the police suspected that the jury would [MASK] find out about the evidence.

21. No/The actress that no/the critic liked in the movie suggested that the producer expected that the director would [MASK] win a prestigious award.

22. No/The legislator that no/the congressman consulted about the proposal thought that the president suggested that the government should [MASK] increase military spending for the war.

23. No/The politician that no/the journalist endorsed in the newspaper suggested that the voters thought that the election would [MASK] cause such a huge scandal.

24. No/The teenager that no/the parent trusted with a car suggested that the paramedics believed that an accident could [MASK] happen in sunny weather.

25. No/The survivor that no/the medic could treat with a first-aid kit thought that the doctor expected that a full recovery would [MASK] be possible in one month

26. No/The athlete that no/the coach recruited for the team expected that the sponsors anticipated that the scandal would [MASK] receive so much media coverage.

27. No/The congressman that no/the citizen supported during the crisis expected that the senate assumed that the treasury would [MASK] lower the national debt.

28. No/The professor that no/the student visited during office hours expected that the dean anticipated that the exam would [MASK] be so difficult for the class.

29. No/The actor that no/the judge nominated for an award believed that the fans expected that the movie would [MASK] be such a blockbuster hit.

30. No/The actress that no/the director auditioned for the role expected that the critics thought that the movie would [MASK] cause so much controversy.

31. No/The champion that no/the competitor defeated in the race believed that the comittee expected that that the coach would [MASK] receive a life-time achievement award.

32. No/The artist that no/the collector regarded very highly believed that the curator suggested that the gallery should [MASK] buy cheap frames for the expensive paintings.

33. No/The scientist that no/the reporter cited in the story expected that the researchers believed that the public would [MASK] care about the new discovery.

34. No/The teenager that no/the teacher motivated before the test believed that the principal claimed that the parents should [MASK] offer more help on assignments.

35. No/The student that no/the professor could tutor on the weekend believed that the teacher thought that the assignments should [MASK] be more than seven pages.

36. No/The protestor that no/the reporter interviewed on live television believed that the council expected that the mayor would [MASK] give in to the numerous demands.

B.2 Linear distance

1. No/The journalist that no/the American broadcast editor recommended for the assignment thought that the advanced younger readers would [MASK] understand the complicated situation.

2. No/The investor that no/the famous British businessman informed about the recession predicted that the free market stock would [MASK] drop below the initial offering price.

3. No/The ambassador that no/the black American diplomat consulted about the treaty thought that the Russian CNBC journalists would [MASK] reveal the truth about election.

4. No/The professor that no/the female linguistics student trusted at the college thought that the leading university administrators would [MASK] increase the yearly tuition.

5. No/The customer that no/the arrogant Chinese salesman assisted in the showroom thought that the white snobish manager would [MASK] consider their lowest offer.

6. No/The protestor that no/the young female journalist interviewed at the rally implied that the Texas state legislators could [MASK] pass the necessary laws.

7. No/The senator that no/the corrupt non-profit corporation supported with campaign donations thought that the newly registered lobbyists would [MASK] accept the sly bribe.

8. No/The lawyer that no/the tired head policeman respected after the trial anticipated that the federal court judge would [MASK] deliver such a harsh sentence.

9. No/The student that no/the English language teacher punished for bad behavior expected that the private school principal would [MASK] hear about the incident.

10. No/The accountant that no/the certified public inspector audited in the past year thought that the non-profit organization agent would [MASK] find out about the scandal.

11. No/The actor that no/the british film fan recognized at the after-party believed that the ingreasingly aggressive paparazzi would [MASK] find out about the affair.

12. No/The teacher that no/the enthusiastic novice parent recommended for the award expected that the research active faculty would [MASK] receive a huge pay raise.

13. No/The student that no/the new medical librarian could help in the afternoon expected that the very lenient teacher would [MASK] extend the approaching deadline.

14. No/The child that no/the extremely wild bully picked on at recess thought that the martial arts teacher would [MASK] give such a harsh punishment.

15. No/The criminal that no/the college campus policeman could catch in the raid expected that the well known judge would [MASK] accept a plea bargain.

16. No/The employee that no/the hard working manager recommended for the promotion anticipated that the genuinely kind boss would [MASK] ask such difficult questions.

17. No/The investor that no/the famous billionaire manager trusted with the money thought that the IT related stock prices would [MASK] increase drastically overnight.

18. No/The candidate that no/the actively concerned voter supported during the election believed that the notoriously arrogant mayor would [MASK] be re-elected for a second term.

19. No/The doctor that no/the responsible medical nurse assisted during the operation assumed that the large health insurance company would [MASK] cover the hospital bill.

20. No/The criminal that no/the careless chatty witness could identify in the courtroom suspected that the randomly assembled jury would [MASK] find out about the evidence.

21. No/The actress that no/the universally acclaimed critic liked in the movie expected that the new film director would [MASK] win a prestigious award.

22. No/The legislator that no/the fairly elected congressman consulted about the proposal suggested that the current federal government should [MASK] increase military spending for the war.

23. No/The politician that no/the popular opposition journalist endorsed in the newspaper thought that the next presidential election would [MASK] cause such a huge scandal.

24. No/The teenager that no/the responsible American parent trusted with a car believed that a fatal car accident could [MASK] happen in sunny weather.

25. No/The survivor that no/the trained emergency medic could treat with a first-aid kit expected that an unexpectedly speedy full recovery would [MASK] be possible in one month

26. No/The athlete that no/the female professional coach recruited for the team anticipated that the small local scandal would [MASK] receive so much media coverage.

27. No/The congressman that no/the politically involved citizen supported during the crisis assumed that the current American treasury would [MASK] lower the national debt.

28. No/The professor that no/the reasonable college student visited during office hours anticipated that the final written exam would [MASK] be so difficult for the class.

29. No/The actor that no/the new theater judge nominated for an award expected that the independent horror movie would [MASK] be such a blockbuster hit.

30. No/The actress that no/the old prominent director auditioned for the role thought that the blockbuster action movie would [MASK] cause so much controversy.

31. No/The champion that no/the gold medal competitor defeated in the race expected that that the abusive athletic coach would [MASK] receive a life-time achievement award.

32. No/The artist that no/the talented fine collector regarded very highly suggested that the modern art gallery should [MASK] buy cheap frames for the expensive paintings.

33. No/The scientist that no/the distinguished climate reporter cited in the story believed that the wider general public would [MASK] care about the new discovery.

34. No/The teenager that no/the typical American teacher motivated before the test claimed that the strict immigrant parents should [MASK] offer more help on assignments.

35. No/The student that no/the poorly motivated professor could tutor on the weekend thought that the final math assignments should [MASK] be more than seven pages.

36. No/The protestor that no/the elderly angry reporter interviewed on live television expected that the current governing mayor would [MASK] give in to the numerous demands.

C Experiment 3

C.1 Hierarchical distance

1. No/The journalist that no/the editor recommended for the assignment thought that the readers would [MASK] understand the complicated situation.

2. No/The investor that no/the businessman informed about the recession predicted that the manager would [MASK] offer below the initial dollar price.

3. No/The ambassador that no/the diplomat consulted about the treaty thought that the journalists would [MASK] reveal the truth about election.

4. No/The professor that no/the student trusted at the college thought that the administrators would [MASK] increase the yearly tuition.

5. No/The customer that no/the salesman assisted in the showroom thought that the manager would [MASK] consider their lowest offer.

6. No/The protestor that no/the journalist interviewed at the rally implied that the legislators could [MASK] endorse the necessary laws.

7. No/The senator that no/the corporation supported with campaign donations thought that the lobbyists would [MASK] accept the sly bribe.

8. No/The lawyer that no/the policeman respected after the trial anticipated that the judge would [MASK] want such a harsh sentence.

9. No/The student that no/the teacher punished for bad behavior expected that the principal would [MASK] talk about the incident.

10. No/The accountant that no/the inspector audited in the past year thought that the agent would [MASK] find out about the scandal.

11. No/The actor that no/the fan recognized at the after-party believed that the paparazzi would [MASK] find out about the affair.

12. No/The teacher that no/the parent recommended for the award expected that the faculty would [MASK] receive a huge pay raise.

13. No/The student that no/the librarian could help in the afternoon expected that the teacher would [MASK] meet the approaching deadline.

14. No/The child that no/the bully picked on at recess thought that the teacher would [MASK] ignore such a strict punishment.

15. No/The criminal that no/the policeman could catch in the raid expected that the judge would [MASK] accept a plea bargain.

16. No/The employee that no/the manager recommended for the promotion anticipated that the boss would [MASK] ask such difficult questions.

17. No/The investor that no/the manager trusted with the money thought that the thieves would [MASK] be arrested overnight.

18. No/The candidate that no/the voter supported during the election believed that the mayor would [MASK] be re-elected for a second term.

19. No/The doctor that no/the nurse assisted during the operation assumed that the insurance would [MASK] review the hospital bill.

20. No/The criminal that no/the witness could identify in the courtroom suspected that the jury would [MASK] find out about the evidence.

21. No/The actress that no/the critic liked in the movie expected that the director would [MASK] win a prestigious award.

22. No/The legislator that no/the congressman consulted about the proposal suggested that the government should [MASK] increase military spending for the war.

23. No/The politician that no/the journalist endorsed in the newspaper thought that the election would [MASK] cause such a huge scandal.

24. No/The teenager that no/the parent trusted with a car believed that a toddler could [MASK] behave well at school.

25. No/The survivor that no/the medic could treat with a first-aid kit expected that a shooter would [MASK] be interviewed so soon.

26. No/The athlete that no/the coach recruited for the team anticipated that the scandal would [MASK] receive so much media coverage.

27. No/The congressman that no/the citizen supported during the crisis assumed that the treasury would [MASK] lower the national debt.

28. No/The professor that no/the student visited during office hours anticipated that the teacher would [MASK] be so exhausted in the class.

29. No/The actor that no/the judge nominated for an award expected that the director would [MASK] become a famous celebrity.

30. No/The actress that no/the director auditioned for the role thought that the movie would [MASK] cause so much controversy.

31. No/The champion that no/the competitor defeated in the race expected that the coach would [MASK] receive a life-time achievement award.

32. No/The artist that no/the collector regarded very highly suggested that the gallery should [MASK] buy cheap frames for the expensive paintings.

33. No/The scientist that no/the reporter cited in the story believed that the public would [MASK] care about the new discovery.

34. No/The teenager that no/the teacher motivated before the test claimed that the parents should [MASK] care about the next exam.

35. No/The student that no/the professor could tutor on the weekend thought that the professor should [MASK] be ready for the exam.

36. No/The protestor that no/the reporter interviewed on live television expected that the mayor would [MASK] agree with the numerous demands happily.

C.2 Linear distance

1. No/The journalist that no/the American broadcast editor sincerely recommended for the interview assignment would [MASK] understand the complicated situation.

2. No/The investor that no/the famous British businessman regrettably informed about the recent recession would [MASK] offer below the initial dollar price.

3. No/The ambassador that no/the black American diplomat confidentially consulted about the international treaty would [MASK] reveal the truth about election.

4. No/The professor that no/the female linguistics student fully trusted at the small college would [MASK] increase the yearly tuition.

5. No/The customer that no/the arrogant Chinese salesman regularly assisted in the fancy showroom would [MASK] consider their lowest offer.

6. No/The protestor that no/the young female journalist secretly interviewed at the political rally could [MASK] endorse the necessary laws.

7. No/The senator that no/the corrupt non-profit organization fully supported with the campaign donations would [MASK] accept the sly bribe.

8. No/The lawyer that no/the tired head policeman sincerely respected after the criminal trial would [MASK] want such a harsh sentence.

9. No/The student that no/the English language teacher cruelly punished for the bad behavior would [MASK] talk about the incident.

10. No/The accountant that no/the certified public inspector carefully audited in the past few years would [MASK] find out about the scandal.

11. No/The actor that no/the british film fan happily recognized at the wild after-party would [MASK] find out about the affair.

12. No/The teacher that no/the enthusiastic novice parent highly recommended for the prestigious award would [MASK] receive a huge pay raise.

13. No/The student that no/the new medical librarian could willingly help in the late afternoon would [MASK] meet the approaching deadline.

14. No/The child that no/the extremely wild bully regularly picked on at the recess would [MASK] ignore such a strict punishment.

15. No/The criminal that no/the college campus policeman could successfully catch in the successful raid would [MASK] accept a plea bargain.

16. No/The employee that no/the hard working manager sincerely recommended for the new promotion would [MASK] ask such difficult questions.

17. No/The investor that no/the famous billionaire manager completely trusted with the investment money would [MASK] be arrested overnight.

18. No/The candidate that no/the actively concerned voter proudly supported during the senator election would [MASK] be re-elected for a second term.

19. No/The doctor that no/the responsible medical nurse carefully assisted during the long operation would [MASK] review the hospital bill.

20. No/The criminal that no/the careless chatty witness could confidently identify in the quiet courtroom would [MASK] find out about the evidence.

21. No/The actress that no/the universally acclaimed critic really liked in the new movie would [MASK] win a prestigious award.

22. No/The legislator that no/the fairly elected congressman confidentially consulted about the legislative proposal should [MASK] increase military spending for the war.

23. No/The politician that no/the popular opposition journalist fully endorsed in the local newspaper would [MASK] cause such a huge scandal.

24. No/The teenager that no/the responsible American parent completely trusted with an electric car could [MASK] behave well at school.

25. No/The survivor that no/the trained emergency medic could successfully treat with a prepared first-aid kit would [MASK] be interviewed so soon.

26. No/The athlete that no/the female professional coach confidently recruited for the soccer team would [MASK] receive so much media coverage.

27. No/The congressman that no/the politically involved citizen happily supported during the recent crisis would [MASK] lower the national debt.

28. No/The professor that no/the reasonable college student regularly visited during the office hours would [MASK] be so exhausted in the class.

29. No/The actor that no/the new theater judge proudly nominated for a movie award would [MASK] become a famous celebrity.

30. No/The actress that no/the old prominent director willingly auditioned for the lead role would [MASK] cause so much controversy.

31. No/The champion that no/the gold medal competitor brutally defeated in the motorbike race would [MASK] receive a life-time achievement award.

32. No/The artist that no/the talented fine collector regarded very highly and often should [MASK] buy cheap frames for the expensive paintings.

33. No/The scientist that no/the distinguished climate reporter intentionally cited in the fake story would [MASK] care about the new discovery.

34. No/The teenager that no/the typical American teacher tirelessly motivated before the current test should [MASK] care about the next exam.

35. No/The student that no/the poorly motivated professor could secrely tutor on the final weekend should [MASK] be ready for the exam.

36. No/The protestor that no/the elderly angry reporter extensively interviewed on the live television would [MASK] agree with the numerous demands happily.