On semantic illusions and the limits of attention: Fallibility in linguistic processing

Haley Hsu & Elsi Kaiser, <haleyhsu@usc.edu>, University of Southern California

Language processing output is fallible and nonveridical, as illustrated by Moses Illusions: When asked "how many animals of each kind did Moses bring on the Ark?", many respond 'two' despite knowing the biblical story is about *Noah*, not Moses (e.g. [1,6,7,8]). Susceptibility to such semantic illusions is often attributed to top-down processing and a failure to create detailed representations bottom-up from linguistic input (e.g. [2]), but many questions remain open (e.g. [8]). To further explore this, we conducted two studies on English to examine three hypotheses on whether presence of a second (*non-semantic* error) impacts *detection of a semantic illusion.*

Hyp1: No effect? Given that linguistic input often contains errors/distortions/noise and humans are good at fixing/correcting errors (e.g. [3,4,5]), errors that do not require forming full semantic representations to detect, such as typos/deletions, may be easy enough that they do not detract from detecting semantic illusions/errors.

Hyp2: Facilitate? If the bottom-up processing needed for detecting a typo/deletion can facilitate bottom-up processing more generally, detecting a non-semantic error could *boost* detection of semantic errors.

Hyp3: Hinder? Since attentional resources are limited (e.g. [2]), we may find an inhibitory effect: Anything that requires attentional resources, e.g. detecting another error of any kind, might render people less likely to detect the semantic error.

We ran two studies: U.S.-born native English speakers (Exp1 n=58, Exp2 n=55) checked potential quiz questions for errors (Fig.1), corrected any errors or answered the question if there were no errors. Instructions and practice made clear questions could contain multiple errors.

Exp1 manipulated whether the questions (see ex.1-2) contained (i) no error (baseline), (ii) a missing / deleted word, (iii) a typo (e.g. *holidat, objevt, otehr)*, (iv) an illusion error, (v) an illusion error *and* a deleted word, or (vi) an illusion error *and* a typo. Exp1 consisted of 24 targets (Latin-Square design), and 28 fillers without errors. **Exp2** was the same except the Typo condition was replaced with Typo+deletion (ex.2). Thus, Exp2 is a replication check, and including a two-error condition without an illusion error allows us to see how often people notice two errors in general.

Results. In both studies, in no-error conditions, people overwhelmingly report knowing the answer and say no correction is needed (Figs.2a-b): they are attending to the task and have the relevant knowledge. Further analyses confirm the actual answers people give are >98% correct.

Now let's look at how often people say (some kind of) correction is needed (red bars, Fig.2a-b). In both studies, error detection rates are equally high in the Typo, Illusion+deletion and Illusion+typo conditions (*emmeans*, pairwise Bonferroni-corrected, p's>0.8 in both exp). In Exp1, the Deletion condition does not differ from other error conditions (|z|'s<2.6, p's>0.15); in Exp2, Deletion+typo has higher error-detection than Deletion-only (|z|=3.7, p<0.005). Strikingly, in both Exp1-2, the **illusion condition** yields *lower rates* of error detection than the Typo, Illusion+deletion, Illusion+typo and Typo+deletion conditions (|z|'s>3.6, p's<0.005 in Exp1-2).

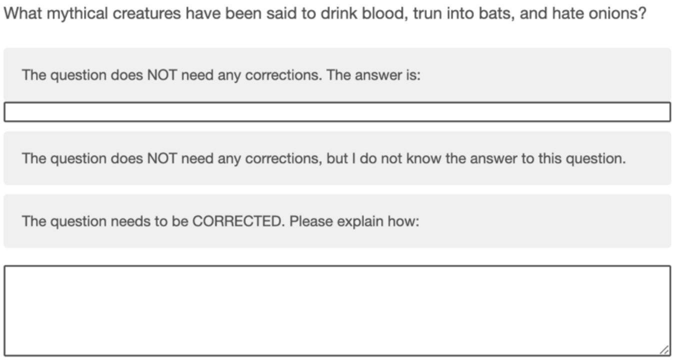
But when we look at ***what kind of error*** people detect, it becomes clear that ***presence (and detection) of typo and deletion errors decreases the detection of illusion errors.*** This is shown in Fig.3. In both Exp1-2, illusion errors are detected at chance rates in Illusion conditions (p's>.3) but at *below-chance rates* in Illusion+deletion and Illusion+typo conditions (|z|'s>2.2, p's<0.03) – conditions where participants detect deletions and typos at high rates.

Further analyses of the two-error conditions show that in the *non-illusion* typo+del condition, both errors are reported almost twice as often as in the Illusion+del and illusion+typo conditions. Thus, presence of a second error *lowers the likelihood of detecting* ***illusion errors*** *specifically.*

In sum, our results support Hyp.3, pointing to a general attentional cost, and suggest that seemingly unimportant factors (typos, deletions) can impact the ability to notice semantic inaccuracies in linguistic input. We find no evidence that noticing an error could trigger a shift to deeper processing, suggesting that shallow processing may be a strong cognitive default.

(1) *Exp1 (topics varied; deletions vary in part-of-speech; del & typo locations counterbalanced)*
baseline:What mythical creatures have been said to drink blood, turn into bats, and hate garlic?
del: What mythical creatures been said to drink blood, turn into bats, and hate garlic?
typo: What mythical creatures have been said to drink blood, trun into bats, and hate garlic?
illusion:What mythical creatures have been said to drink blood, turn into bats, and hate onions?
ill+del: What mythical creatures been said to drink blood, turn into bats, and hate onions?
ill+typo:What mythical creatures have been said to drink blood, trun into bats, and hate onions?

(2) *Exp2: same design as Exp1, but Typo condition was replaced by Typo + deletion condition:*
typo + del: What mythical creatures been said to drink blood, trun into bats, and hate garlic?



What mythical creatures have been said to drink blood, trun into bats, and hate onions?

The question does NOT need any corrections. The answer is:

The question does NOT need any corrections, but I do not know the answer to this question.

The question needs to be CORRECTED. Please explain how:

Fig.1. (left) Sample item (instructions and practice items explained questions could involve *multiple errors*)

Fig.2a. Exp1: Proportion of responses

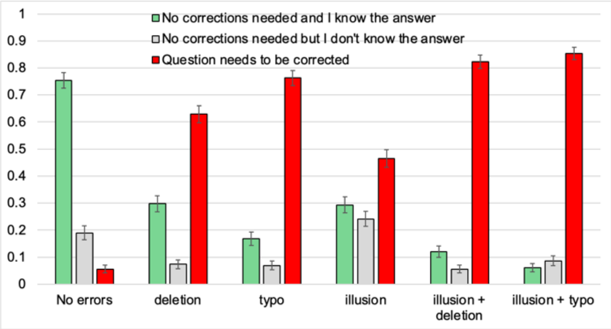
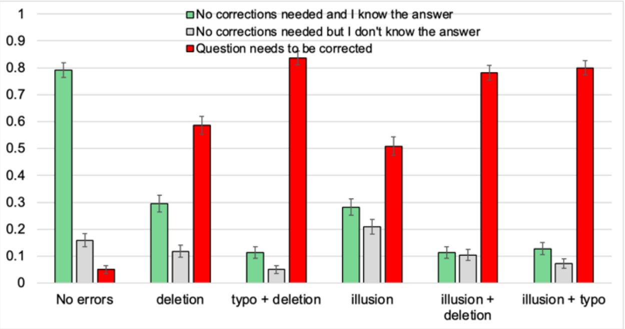


Fig.2b. Exp2: Proportion of responses



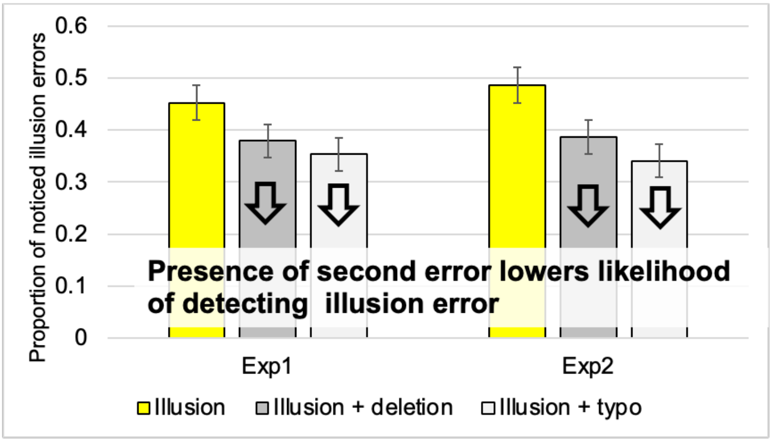


Fig.3. (left) Exp1 and Exp2: In the conditions where the sentence contains an illusion error, how often did participants notice the illusion error?

References: [1] Erickson & Matteson 1981. From words to meanings. *JVLVB* [2] Ferreira & Patson 2007. *LLC* The 'good enough' approach. [3] Gibson et al 2013. Rational integration of noisy evidence and prior semantic expectations. *PNAS* [4] Jurafsky & Martin 2024. Speech and Language Processing. [5] Park & Levy 2011. Automated whole sentence grammar correction using a noisy channel model. *ACL* [6] Sanford & Sturt 2002. Depth of processing. *TICS* [7] Shafto & MacKay 2000. Moses, mega-Moses, and Armstrong illusions. *Psych Sci* [8] Speckmann & Unkelbach 2022. The Moses illusion. *Cog Illusions*