The Accidental Ambiguity of Inversion Illusions

Ellen O’Connor, USC

Overview: (1) is a commonly cited example of semantic processing going off the rails: comprehenders almost universally (and very persistently) entertain the meaning in (2a), which is at odds with its actual grammar-based meaning in (2b), as described in greater detail below. This grammatical illusion, unlike others, has a well-formed (if implausible) LF that is apparently ‘inverted’ in perception. In this paper I identify three possible accounts of the phenomenon which variously attribute it to incomplete or “shallow” logical semantics, to incomplete or “shallow” lexical semantics, to “deep” processing in conjunction with perceived grammatical ambiguity, and provide some preliminary evidence for the latter.

(1) No head injury is too trivial to ignore.

(2) a. Perceived: No head injury should be ignored, even the most trivial one.
b. Actual: # All head injuries should be ignored, even the most trivial one.

The meaning contributed by the grammar: The distribution of *too* elsewhere in English would lead us to expect a perception approximating (3) (see Meier 2003, von Stechow et al 2004 on the meaning of *too*).

(3) $\neg\exists$ x: head injury(x) & max{d: trivial(x) $\geq$ d)} >

max{d: $\exists$ w’ $\in$ Acc(w) & x is ignored in w’ & trivial(x) $\geq$ d in w’}

Note that this LF is doubly implausible: first, *too trivial to ignore* is pragmatically odd in the same way as *too thirsty to drink water:* world knowledge tells us that people who are more thirsty are *more* likely to drink water, and head injuries that are more trivial are *more* likely to be ignored. The maximal triviality threshold introduced by *too trivial* requires the opposite to be true, yielding the meaning that injuries can only be shrugged off if they are *lower* on the scale of triviality, i.e. if they are more serious. I will call this the illusion’s INTERNAL ANOMALY. Moreover, because *no head injury* falls above that maximal threshold, it follows that one should be able to shrug off *any* head injury – what I term its EXTERNAL ANOMALY. The overwhelming perception, however, is that (3) is not the meaning of (1).

Three hypotheses: Although this illusion is often speculatively associated with its multiple negative elements (*no, too, trivial, ignore*), no research has systematically unpacked and empirically investigated whether and why negation affects the illusion. I divide existing accounts into three broad approaches. The *Channel Capacity Hypothesis* (Wason & Reich 1979, Sanford & Sturt 2002) attributes the illusion to negation-associated cognitive overload: because the logical semantics of (1) is computationally intractable, interpretations are driven exclusively by grammar-external plausibility heuristics (i.e., world knowledge). The *Change Blindness Hypothesis* (Pickering & Garrod 2007) focuses on the role of the sentence onset *no head injury is too trivial* in generating strong predictions about the likely sentence meaning (~ ‘head injuries should always be treated’); incomplete lexical processing of *ignore* might allow comprehenders to proceed with this analysis even when it is no longer available. Finally, the *Hypernegation Hypothesis* (Horn 2009) connects the illusion to phenomena observed elsewhere in the grammar, such as negative concord, paratactic negation, and negative polarity licensing, with the illusion possibly a result of overapplication of these existing operations.

Experiments 1-2: To establish basic facts about the illusion we crossed degree quantifier polarity (*too* vs. *enough*) with verbal polarity (*neg* vs *pos*) in Exp 1, and degree quantifier polarity with determiner polarity (*no* vs *all*) in Exp 2 (see below); all target stimuli were internally anomalous (e.g., # *wasteful enough to support*). Participants were asked to ‘tutor’ cartoon aliens about *too/enough* sentences by indicating whether each sentence made sense, and offering a paraphrase for sensical sentences or a correction for nonsensical ones. Norming studies also elicited for each item and condition (a) plausibility scores for paraphrases (e.g. *No/all social program can/should be opposed/supported*), (b) measures of how predictable participants found the final verb to be. Accuracy rates are shown below.

Experiment 1	Experiment 2
No social program is *too wasteful* to *oppose*. (29%)	*No* social program is *too wasteful* to oppose. (27%)
No social program is *efficient enough* to *oppose*. (79%)	*No* social program is *efficient enough* to oppose. (74%)
No social program is *too efficient* to *support*. (63%)	*All* social programs are *too wasteful* to oppose. (81%)

No social program is *wasteful enough* to *support*. (79%) *All* social programs are *efficient enough* to oppose. (87%)

Data were analyzed using mixed effects linear regression models to analyze (a) accuracy rates (did participants detect the internal anomaly?) and (b) inversion rates (if the internal anomaly was undetected, did participants invert the meaning from *all* NP*s can/should be* V*ed* to *no* NP*s can/should be* V*ed*?).

Effects of negation: As expected, accuracy was indeed impaired by the presence of negation, though not additively. Experiment 1 elicited a main effect of degree quantifier polarity (p < .001) and an interaction with verbal polarity (p < .001), with verbal polarity affecting *too*-sentences much more strongly. Experiment 2 elicited a main effect of determiner type (p < .01) qualified by a significant interaction with degree quantifier polarity (p < .05), again showing *no-too* to be the critical condition. In sum, results imply an explosion of difficulty associated with anomaly detection in the *no-too* condition.

Effects of plausibility: Plausibility ratings were used to predict accuracy and illusory percept in the *no-too* data (collapsed across experiments). In line with the Channel Capacity Hypothesis participants were more likely to notice internal anomalies when the grammar-based meaning was plausible (p = .001), although accuracy never rose above chance even in the most plausible cases. Contrary to the Channel Capacity Hypothesis, participants who failed to notice the internal anomaly did not interpret the sentence in the most plausible way – rather, in such cases they overwhelmingly inverted the external meaning, regardless of the resulting plausibility (80.3% of the time in Exp 1, and 90.9% of the time in Exp 2; no effect of plausibility). In other words, perceived meaning in the most difficult condition was neither random nor driven by plausibility heuristics; rather, participants either interpreted *no-too* grammatically, or they systematically interpreted it as though it were a *no-enough* sentence; they appeared to use pragmatic cues mainly to select between these two possible analyses.

Effects of predictability/confidence: Participants were shown the target stimuli with the final verb missing; they were asked to provide a verb completion and indicate their confidence in their choice. Contrary to the Change Blindness Hypothesis, confidence measures (which were negatively correlated with cloze probability) did not significantly modulate accuracy rates: items with a highly predictable completion were no easier to process correctly than those that were more open-ended and less predictable.

Experiment 3: We next focused on whether *no* and *too* might be interpreted in a negative concord or NPI-like configuration, with two instances of logical negation conflated into one. This could explain why *no-too* sentences are perceived as *no-enough* sentences: the logical negation implicit in the semantics of *too* is neutralized in particular environments. Factive verbs disrupt NPI dependencies because they are strongly veridical, presupposing the truth of their complement clause, and negative concord deos not persist across clausal boundaries. Thus, the Hypernegation Hypothesis would predict the illusion to be mitigated when *no* and *too* are separated by a factive verb and clausal boundary. To test this, participants were asked to identify stimuli where “one or more words [had] been swapped for similar, but incorrect, alternatives – rendering the sentence false or nonsensical”; 10 target items were combined with 60 fillers containing other types of substitutions. Verb factivity strongly affected illusion rates, with participants more readily flagging sentences as nonsensical when *no* and *too* were separated by a factive verb and clausal boundary, as in (9a) (p < .001). Thus, the illusion was much easier to detect when the structural and semantic environment precluded this type of dependency formation.

(9) a. **No** politician *knew* that the social program was **too** wasteful to oppose. (64% detected)
b. The politician *knew* that **no** social program was **too** wasteful to oppose. (33% detected)

Conclusion: Our results verify that negation facilitates inversion illusions, and point to a likely reason why: comprehenders treat *no-too* sentences as if they were ambiguous between a *no-too* reading and a *no-enough* reading in environments that license NPIs/negative concord. This seems to implicate a special type of processing error leading to the misapplication of a grammatical operation that is in principle available within UG, but not in standard English – i.e., an illusion driven by accidental ambiguity.

Selected References: Horn, L. (2009). Hypernegation, hyponegation and parole violations. *Proceedings of BLS.* Sanford, A. & P. Sturt (2002). Depth of processing in language comprehension: not noticing the evidence. *Trends in Cognitive Science*, *6* (9), 382-386. Wason, P. & S. Reich (1979). A verbal illusion. *Quarterly Journal of Experimental Psychology* 31(4): 591-597.