DECONSTRUCTING A COMPARATIVE ILLUSION

ALEXIS WELLWOOD, ROUMYANA PANCHEVA, VALENTINE HACQUARD, COLIN PHILLIPS

ABSTRACT. Comparative illusions like *More Americans have been to Russia than I have* are well-known in the linguistics community, but surprisingly little formal, experimental work has been conducted to understand the nature of the phenomenon. Is the illusion robust? Under what conditions does it obtain? Certainly, the classic example seems highly acceptable to most native speakers of English, yet even such speakers will (eventually) readily admit that it has no coherent interpretation. What are the linguistic properties of the classic example? To what extent does the illusory effect depend on these properties? Putting comparative illusions under the microscope, we consider whether the source of the effect is due to a 'shallow' parser, some type of repair by ellipsis, an incorrectly-resolved lexical ambiguity, or to the semantics of comparison more generally. We report the results of four acceptability judgment studies and one sentence recall study which support only the semantics of comparison explanation. Concluding the paper, we consider these results in terms of what they reveal about the nature of grammar and language processing.

1. Comparative illusions

The sentence in (1), while often judged to be a perfectly acceptable sentence of English, seems to have no stable, meaningful interpretation.[1]

(1)   More people have been to Russia than I have.

Such sentences have come to be referred to as 'comparative illusions': comparative structures that give rise to an illusion of syntactic and semantic well-formedness. The illusions are interesting in that they seem to challenge some of our most basic assumptions about language architecture—that we perceive sentences veridically, that we interpret them fully, and that sentence form and meaning are tightly coupled. In this paper, we investigate whether the illusions are as acceptable as has been thought, and which properties are essential for the illusion effect.

---

[1]This sentence was first reported by Montalbetti 1984 as "the most amazing */? sentence I've ever heard", attributed to Hermann Schultze.

The source of the ungrammaticality of the illusion is the pronominal subject of the *than*-clause. The non-illusory (2) is superficially very similar to (1), but (2) succeeds in being a meaningful, not just acceptable, sentence of English.

(2) More people have been to Russia than **elephants** have.

(2) states a comparison between two quantities, the number of people that have been to Russia and the number of elephants that have, where the first quantity is greater than the second. No such understanding is possible for (1): the first person pronoun *I* in the *than*-clause fails to provide a countable quantity of individuals.

Acceptability judgments reflect a composite of factors, rather than being a direct measure of grammaticaility. Apart from grammatical status, other factors both linguistic (e.g. lexical frequency, parsing preferences) and non-linguistic (e.g. memory capacity, structural complexity) affect how acceptable a given string is. To explain discrepancies between the predictions of grammatical theory on the one hand, and acceptability ratings on the other, these factors are differentially invoked depending on the case at hand. One of the most prominent examples of appeal to non-linguistic factors in explaining a mismatch between grammaticality and acceptability is multiple center-embedding. Grammatically, (3c) is formed by the same rule as (3a,b).

(3) a. The man jumped.
   b. The man the woman knows jumped.
   c. The man the woman the child kissed knows jumped.

For most speakers, no amount of repetition or attempts with different prosodic patterns will make (3c) seem acceptable, despite its grammaticality. A widely held view of such paradigms blames the parser: for reasons that are not altogether clear, it is unable to process more than two iterations of this rule (see Lewis 1996).

A different sort of case entirely is when an ungrammatical string is perceived to be acceptable. The example in (4a) initially strikes most English speakers as acceptable, despite its ill-formedness. In contrast, (4b) and (4c) are not even momentarily considered acceptable.

(4) a. The key to the cabinets are on the table.
   b. Key the the to cabinets the on are table.
   c. The key are on the table.

The only difference between (4c) and (4a) is that the prepositional phrase *to the cabinets* has been removed, which makes the mismatch in number between the subject head noun and the verb more apparent. A notable feature of the illusion of grammaticality with sentences like (4a) is that it rarely persists once the error has been consciously attended to.

With neither the center-embedding nor agreement illusions cases is there any obvious need to appeal to anything other than syntax and syntactic processing to understand what has gone wrong. Other well-known examples, presented in (5), are different, featuring fully-grammatical sentences that are perceived as meaning something starkly different from their literal meaning (Wason & Reich 1979).

(5) a. Can a man marry his widow's sister?

b.   No head injury is too trivial to ignore.

If a man has a widow, that man is dead, and no dead man can marry. Anyone, upon reflection, will agree that the opposite is suggested by the question in (5a), yet 30% of respondents answered "yes" when presented with it (Sanford & Sturt 2002). Similarly, (5b) is literally equivalent to *Every head injury is trivial enough to ignore*, a suggestion to let all such injuries pass by unattended. Yet, speakers typically construct a meaning for (5b) that is equivalent to *Every head injury is too important to ignore* (Wason & Reich 1979). To understand what happens when speakers are presented with such sentences, a more nuanced account of grammatical processing is required.

The comparative illusions appear to be more like the latter sort of case, yet they are different from both: unlike the experience of (4a), further consideration does not lead those not trained in syntax and semantics to notice that the sentence is ill-formed, and unlike the experience of (5a) and (5b), one never arrives at the "correct" or literal meaning. Indeed, the illusions appear not to *have* a correct or literal meaning to appreciate. This disconnect, between apparent syntactic well-formedness and shifting-sands semantic interpretation, suggests that the study of such examples could reveal something important about the mental architecture of language.

In this paper, we test the robustness of naive speakers' acceptability of (1), and examine four possible sources for the illusion. The account described in §2.1 considers the illusion of acceptability due to general properties of the language processor, viewed as a syntactic-template matching procedure. We take this account to rely on the presence of the quantifier *more* as opposed to its dual, *fewer*. The account in §2.2 takes ellipsis to be essential: ellipsis has been shown to rescue an otherwise clearly unacceptable string in other constructions, perhaps here as well. The account in §2.3 posits that the lexical ambiguity between comparative and additive *more* is responsible, and so requires that the *than*-clause subject potentially be included in the group denoted by the matrix subject. Finally, the account in §2.4 posits that the effect is due to a general semantic property of comparatives, namely a requirement for plurality, taking as crucial the potential for the VP's action to be repeated for a given agent.

In §3, we detail the five experiments we designed and conducted to put these accounts to the test. In §3.1-3.3, we discuss the results of four formal acceptability judgment studies designed to test which features of the illusion-type sentences have an impact on their acceptability. These experiments effectively implement tests of the kinds of semantic intuitions speakers mention when they encounter the illusions. In §3.4, we present the results of a novel sentence recall experiment. This experiment was designed to investigate what speakers actually *do* when presented with the illusions. §4 concludes with a discussion of how our results inform understanding of grammar and linguistic processing.

First, a brief digression: the question of whether examples like (1) are grammatical may be a sticking point for some, so in §1.1 we briefly consider arguments for and against this conclusion.

1.1. Are illusion sentences grammatical?

Perceiving the classic comparative illusion in (1) gives rise to conflicting hypotheses about its nature. Some would argue that the string is perfectly *un*grammatical compared to (2), while others may contend that both are perfectly grammatical. The only surface difference between the two strings is in the properties of the *than*-clause subject. In this short section, we'd like to present what we take to be the two salient views on this issue, while staking our own claim.

We take the illusion sentences to be ungrammatical, in accord with constraints on the syntactic and semantic composition of nominal comparatives. Part of the representation of nominal comparatives is an expression in the *than*-clause, *x-many* NP*s*, a correlate of *more* in the matrix clause (Bresnan 1973). A critical aspect of the construction of the *than*-clause is merger of a degree operator with the subject NP, which is impossible with a pronoun (and equally impossible with a full DP).[2] From a semantic perspective, nominal comparatives require that the expression whose denotation is measured be either plural or mass, never singular (Hackl 2001, Nakanishi 2007, Wellwood et al. 2012; Wellwood to appear). The interest for us, then, is how an ungrammatical sentence can so robustly *seem* to be both acceptable and meaningful.

Those that consider the illusions to be *grammatical* take a less rich view of what is involved in the construction of sentences like this one. On such a view, there are no syntactic differences between (1) and (2), as expressions of the same syntactic category appear in the same positions in both: despite the apparent formal and semantic differences (i.e., person and number) between the *than*-clause subjects, both such expressions are *just* instantiations of NP. If the appearance of one such NP appears in a legitimate syntactic representation (e.g. (2)), what could reasonably prevent the other from being legitimate also? The interest for the researcher of this persuasion, then, is why a grammatical sentence can yet elude a stable, meaningful interpretation.

Regardless of the reader's position, we think that the investigations reported below will provide new insight and challenges in the ultimate explanation of the contrast in (1-2). In the next section, we sketch four plausible sources for the illusion phenomenon. Following that, we present novel experimental data that test these hypotheses and, at the same time, rule out some properties of the classic illusion as inessential.

2. Plausible sources

In this section, we consider four plausible hypotheses, each designed to provide at least part of an explanation for what it is about illusion sentences that gives rise to the perception of acceptability, and in many cases, the perception of meaningfulness. §2.1 considers the illusion due to the operation of a 'shallow' language processor; §2.2 to some type of repair by ellipsis; §2.3 to an incorrectly-resolved (English-particular)

---

[2]The degree operator in the *than*-clause has the covert form *how many*, which is overt in some languages, e.g. Bulgarian, see Pancheva 2009. It is not possible to construct, e.g. *[[how many I] have been to Russia] or *[[how many the boy] has been to Russia]. See Chomsky 1977 and Heim 2000 for the claim that the *x* in the text is a *wh*-operator.

lexical ambiguity; and §2.4 to the semantics of plurals and comparatives. For each hypothesis, we specify the particular predictions that flow from them. These predictions are tested in §3.

2.1. Syntactic template matching

A quite general way of thinking about the problem exploits the superficial similarity between (1) and (2), suggesting that a syntactic template matching procedure along the lines of Townsend & Bever (2001) may explain the phenomenon.[3]

Under such a view, judgments of acceptability reflect fairly directly how the parser operates, namely, independently of grammatical analysis. A string of words 'passes' the parser if subparts of the string match familiar sentence templates. If the subparts don't fit a template, the string is rejected; if it does, only then is it sent to the grammar for further analysis.

To see how such an account would work in its details, consider the two sentences in (6). Parsing (1) means matching its matrix clause to that in a sentence like (6a), and its *than*-clause to that in (6b).

(6) a. **More people have been to Russia** than I would have thought.

b. People have been to Russia more **than I have**.

(6a-b) are perfectly interpretable: in (6a), *more* compares the cardinalities of individuals that have actually been to Russia and of those individuals I would have thought have been to Russia;[4] in (6b), *more* compares the number of *events* of some-person-or-other going to Russia with the number of events of me going to Russia. From these sentences, a licit matrix and *than*-clause template may be extracted, comparison with which lead to a sentence like (1) being judged acceptable.

Thus (1) passes the parser as surely as (2) does, and both are judged acceptable before any more complete analysis has been conducted. The template matching approach suggests that while the illusions may fail at deeper levels of analysis, their success at shallower levels accounts for their apparent acceptability. We may state this hypothesis as in (7).

(7) **Syntactic template matching hypothesis**
The parser's implementation of a (surface) syntactic template matching procedure is the source of the illusion.

While this style of account is quite general in the details, we might think that the string identity between *more* the determiner and *more* the adverbial is crucial. Otherwise, it is unclear why (6b) would be relevant to the template matching procedure.

If access to sentences like (7b) are required for this procedure, and if it is the presence of *more* that makes its relevance plain, then this account makes a clear prediction. In particular, the adverbial equivalent to nominal *fewer* is *less*, so presenting

---

[3]We have focused on an account in the spirit of Townsend and Bever, as they have directly addressed comparative illusions and make specific predictions in their direction. One may be able to construct a similar account in the style of "good enough processing" (e.g. Christianson et al. 2001, Ferreira et al. 2002) or shallow parsing (Sanford & Sturt 2002), but the predictions of these accounts are not clear for the case at hand.

[4]These sets need not be identical, of course.

subjects with sentences just like (1) except with *more* replaced by *fewer* should result in degraded acceptability. With *fewer*, the equivalent of (6b) is grammatically unavailable.

(8)   Fewer people have been to Russia than I have.

    a.   Fewer **people have been to Russia** than I would have thought.

    b.   * People have been to Russia **fewer than I have**.

Indeed, the adverbial *People have been to Russia fewer than I have* is ungrammatical. If participants rate such sentences as (8) as highly as their counterparts with *more*, the syntactic template-matching account would have to be made considerably more abstract in order to explain the illusion.

2.2. Ellipsis resolution

A different kind of account links the acceptability of an otherwise-ungrammatical sentence with a process of repair by ellipsis. This account differs from the preceding one in that it posits a role for the grammar in facilitating the illusion, by folding the phenomenon into the literature that reveals ellipsis ameliorating various sorts of grammatical problems.

    Both the theoretical (e.g. Ross 1969, Lasnik 2001, Merchant 2001, Kennedy 2003) and experimental literature (Frazier & Clifton 2011) confirms that sluicing saves sentences which would otherwise present robust island violations. Consider the sentence in (9).

(9)   They want to hire someone who speaks a Balkan language, but I don't remember which.

(9) is robustly acceptable (and interpretable), with speakers understanding the *but*-clause to implicitly contain the information *which Balkan language it is such that they want to hire someone who speaks that language.* However, undoing the ellipsis, as in (10), reveals a complex-NP island effect:

(10)   * They want to hire someone who speaks a Balkan language, but I don't remember which **Balkan language they want to hire someone who speaks.**

'Undoing' the ellipsis in such cases results in robust measures of unacceptability.

    Investigating the possibility that it was the application of ellipsis that underpins the illusion effect, Fults and Phillips (2004) tested illusion-type sentences with and without ellipsis in the *than*-clause, and found a significant degradation in acceptability when ellipsis had not been applied. (Numbers indicate mean ratings out of 5.)[5]

(11)   a.   More people have been to Russia than I have.   *3.58*

    b.   More people have been to Russia than I have been to Russia.   *2.90*

---

[5]The illusions were rated significantly lower than their grammatical counterparts, by about .8/5, yet higher than ungrammatical fillers, which were rated at a mean of 2.87/5.

These results were taken to suggest that, whatever the particular nature of the illusion effect, such cases are relevantly like other familiar examples from the literature in requiring ellipsis for robust acceptability.

Although the reasons or mechanisms aren't well understood, successful resolution of ellipsis seems to blind comprehenders to problems elsewhere, like the cases of increased acceptability in center-embedded sentences with missing verbs (see e.g. Gibson & Thomas 1996), or Norvin Richards' (Richards (1997)) "subjacency tax" cases.

The hypothesis that the application of ellipsis is critical to the acceptability of comparative illusions is sketched in (12).

(12) **Repair by ellipsis hypothesis**
Successful resolution of ellipsis in the *than*-clause masks the unacceptability of illusion-type sentences.

If an account of the illusions in terms of the processing mechanisms underlying ellipsis resolution is on the right track, then we predict that illusion-type sentences should be better whenever the VP has been elided in the *than*-clause.

However, there is an alternative explanation for the degradation reported in (11), which may have nothing to do with detection of the comparative error. Sentences like that in (13), where a superficially different VP is pronounced in the *than*-clause, seem to us to be less degraded than (11b) when compared to (11a).

(13)   More people have been to Russia than **I have been to Canada**.

Indeed, the repeated predicate in (11b) may itself be responsible for the reduced ratings, because of general constraints the grammar imposes on comparative constructions: in such constructions, identical material **must** be deleted.[6] If this alternative is correct, then contra the hypothesis in (12), there should be no difference in acceptability between the classic example in (1) and counterparts like (13).

2.3. more *ambiguity*

One thing is certain about the classic illusion: applying the normal interpretive rules to such constructions yield uninterpretability. Yet, perhaps part of the effect is due to speakers temporarily constructing an alternative interpretation that is coherent, and this accounts for heightened judgments of acceptability. The approach considered here differs from the preceding two in that it explictly invokes semantic analysis in facilitating this process.

A coherent interpretation of the illusion is that interpretive processes translate it into an assertion with the meaning roughly of either of the two paraphrases in (14):

(14) a.   **I'm not the only person** that has been to Russia.
     b.   More people have been to Russia than **just me**.

On such a view, the interpretation of (1) is not comparative, but *additive*, so that the sentence is true if there are any individuals who have been to Russia *in addition*

---

[6]See Bresnan 1973 for 'Comparative deletion', Lechner 2004 for 'Comparative ellipsis', among many others since.

to the speaker.[7] Indeed, Fults & Phillips (2004) Expt. 1 found that sentences like (14b) were rated as highly as sentences like (1), with means of 3.45/5 and 3.58/5 respectively.

Such an account crucially depends on a different string ambiguity of *more* in English, namely that it may describe a comparative or an additive semantics (Greenberg 2009, Thomas 2010). To see the distinction, consider (15).

(15)   (Al worked 10 hours so far on the project.) Now he has to work on it **more**.

    a.   **Comparative**: ...at least another 10 in addition.

    b.   **Additive**: ...some quantity extra, possibly less than 10 hours.

The comparative interpretation, paraphrased in (15a), specifies that the desired quantity is greater compared to some previous quantity. The additive interpretation, paraphrased in (15b), specifies simply that the desired quantity is *in addition to* (but not necessarily greater than) the previous quantity.

One of the properties of additive *more* that distinguishes it from the comparative is that it never appears with a *than*-clause; as such, the illusion-type sentences could not literally support this interpretation. However, if speakers were grammatically unfaithful at some earlier level of processing, they could entertain such a possibility and this could account at least in part for the perception of well-formedness. This hypothesis is given in (16).

(16) **Additive 'more' hypothesis**
    Speakers mis-analyze comparative *more* as additive *more*.

If an account of the illusions in terms of lexical ambiguity is correct, then we predict that the illusion is facilitated just when the semantics of additivity are supported.

Regardless of the type of mechanism that would get a speaker from the classic illusion to the meanings in (14), the predictions of this account are clear. An additive-type interpretation is only supported when the subject of the *than*-clause is plausibly interpreted as being a member of the set denoted by the matrix subject. In (17) below, an additive interpretation is not supported, as the *than*-clause subject may not be included in the denotation of the matrix subject.

(17) More **people** have been to Russia than **that elephant** has.

The classic illusion, as in (1), features a first person subject of the *than*-clause, which in many contexts may be interpreted as part of the matrix subject set. The hypothesis in (16) thus predicts a decline in acceptability for sentences like (17), which do not support this inclusion relationship.[8]

---

[7]Indeed, we find some plausibility for this: anecdotally, the first author's sister, when the first author asked her to consider the sentence *More Americans have been to Russia than I have*, exclaimed, "but you're not American!".

[8]This account makes another prediction, namely that comparative illusions will only be possible in languages where the comparative and additive morpheme are identical morphophonologically. Greenberg 2010 asserts that not all languages are like English in this respect.

2.4. Event comparison

The final plausible source for the illusion we consider finds its motivation in the processes underlying the semantics of comparison. It shares with the preceding hypothesis the positing of a critical role for semantic processing in generating the illusion effect.

Naive informants often offer the impression that the illusions have an interpretation as a comparison over numbers of events, rather than numbers of individuals. This meaning is rather like one gets with an adverbial comparative like that in (18).

(18) **People** have been to Russia **more than I have**.

In such examples, *more* overtly calls for comparison of counts of events, with the subject referents implicated only indirectly.

A reanalysis of comparative illusion sentences as comparison of events can be achieved in at least two ways. Speakers may reanalyze the syntactic structure they have built, "moving" the determiner *more* from its nominal position to an adverbial position, thus creating a structure like that in (18). Or the reanalysis could be semantic in nature, accomplished without constructing a syntax where *more* is an adverbial.

We think the latter approach–semantic reanalysis–is on the right track. In particular, the formal semantics literature considers there to be two ways of counting individuals, which is what nominal *more* calls for: count them directly *qua* individuals, or count them indirectly, *qua* participants in distinct events. To see how this works in non-comparative contexts, consider (19). Here, the numeral *4000* is naturally interpreted either as a count of unique ships, or as a count of unique ship-passings.

(19) 4000 ships passed through the lock last week.

(19) can be true even if a number of individual ships passed through the lock several times each, with only the total ship-passings counted to 4000 (Krifka 1990, Schein to-appear, Barker 1999).

Crucially, the availability of the two types of readings for numerals are only available when there is, for any given individual, a plurality of events in which they participated. With a 'once-only' predicate, the two readings collapse.

(20) 4000 ships **were destroyed** in the battle last week.

Destruction is something that, for a given object, can happen only once; so regardless of whether the ships are counted *qua* ships or *qua* destroyed things, the count is 4000.

We can see this requirement for 'repeatability' with *more*. Regardless of whether it is used to count individuals indirectly (determiner *more*) or events directly (adverbial *more*), only plural expressions are acceptable, relevantly singular expressions are ruled out (see Hackl 2001, Nakanishi 2007 and Wellwood et al. 2012; Wellwood to appear for discussion). Consider the paradigm in (21) with nominal *more*.[9]

(21) a. More **students** than **professors** came to the party.
      b. # More **student** than **professor** came to the party.

---

[9]Note, we are not discussing the compatibility with mass nominal or atelic verbal expressions, though we return to this in §3.4.

Plural *students/professors* are perfectly felicitous, whereas count singular *student* is ruled out. The same pattern obtains with verbal *more*.

(22)   a.   Mary **jumped** more than John did.
         b.   # Mary **died** more than John did.

The "plural" *jump* is perfectly acceptable, but singular *die* is ruled out.[10]

   An account in this spirit says that speakers interpret illusion-type as *indirect* comparisons of individuals, i.e., as comparisons of *participations*.

(23)   **Event comparison hypothesis**
         Speakers analyze illusion-type sentences as an indirect count of individuals *qua* event participants.

If a semantic account in terms of event comparison is supported, then we expect the illusion to be facilitated just in case plurality in the verbal domain is supported.

   To count a plurality of participants, there must be a plurality of events, which requires a repeatable VP. In the classic illusion, the predicate *go to Russia* appears. Indeed, this predicate is repeatable. Thus, an account of the illusions in terms of the semantics of plurality required by comparison predicts that introducing singularity into the VP should result in degraded acceptability for illusion-type sentences.

2.5. Summary

We have considered four accounts for the contrast in acceptability between the classic illusion and their fully acceptable/interpretable counterparts. Each account was seen to make different claims about which properties of the illusion are necessary. From these accounts we were able to generate testable predictions regarding the quantifier (*more* versus *fewer*), ellipsis (presence versus absence), *than*-clause subject type (inclusion versus non-inclusion with the matrix subject), and event plurality (singular or plural VP), summarized in Table 1.

TABLE 1.   Predictions by hypothesis and factor. "-" indicates no prediction, and > indicates the prediction of greater acceptability for the factor on the left than on the right.

Hypothesis	Quantifier *more—fewer*	Ellipsis yes—no	Subject inclusion yes—no	Repeatable VP yes—no
Template matching	>	-	-	-
Repair by ellipsis	-	>	-	-
Additive *more*	-	-	>	-
Event comparison	-	-	-	>

   In the next section, we test these predictions using two methodologies: in §3.1 and §3.2, we report the results of four formal acceptability rating studies; in §3.4,

---

[10]English marks plural in the nominal domain by suffixing *-s*, but fails to morphologically mark semantic plurality in the verbal domain. Some languages have overt markers which signal verbal singularity (e.g., perfective aspect morphology in Romance and Slavic languages), and others have overt pluractional markers (e.g. Greenlandic, see van Geenhoven 2005; see Cusic 1981 for a wider range of languages) that signal verbal plurality.

the results of a sentence recall experiment. The acceptability judgment results will let us to closely control the conditions under which participants rate different factor combinations as acceptable, and the recall study will allow us to collect evidence of what speakers actually do when they understand such sentences.

3. Anatomy of an illusion

How robust are the illusions? Is their acceptability reliable, or is it more malleable? Which properties of the classic example are essential? In light of the above hypotheses, we conducted 4 acceptability judgment experiments with 88 (unique) participants and 164 trials, allowing us to evaluate these questions.

3.1. Experiments 1a-c: Acceptability

Experiments 1a-c were identical in their basic design, execution, and participant population. Each study was an offline acceptability judgment task. Responses were recorded on a 7 point scale where 1 is *unacceptable* and 7 is *acceptable*. Participants were University of Maryland undergraduates who were all native speakers of American English as determined in a pre-test questionnaire, and received either course credit or $10 for participation.

3.1.1. Design

The accounts discussed in the previous section predict that different factors should affect the acceptability of illusion-type sentences (defined as those with a non-bare plural NP in the *than*-clause) but not controls (trials with a bare plural NP in *than*-clause). Each of the >'s in Table 1 should be understood as indicating an interaction between control versus illusion on the one hand, and one of these factors. At issue is whether acceptability of the illusions but not controls declines as a function of QUANTIFIER (*more*, *fewer*), ELLIPSIS (yes, no), SUBJECT INCLUSION (possible, not possible), and PREDICATE TYPE (repeatable, non-repeatable). We tested the factors QUANTIFIER and ELLIPSIS within items, and counterbalanced (and anti-correlated) SUBJECT INCLUSION and PREDICATE TYPE across items.

The possibility of an individual's participating in an event type multiple times is critical to the event comparison hypothesis. On illusions trials (e.g. *the boy*), a non-repeatable VP should result in significant degradation of acceptability compared to a repeatable VP, but should have no corresponding effect on controls (e.g. *boys*).

(24) PREDICATE TYPE
    a.   More girls **ate pizza** than **the boy/boys** did.   *repeatable*
    b.   More girls **graduated high school** than **the boy/boys** did. *non*

Given that this manipulation results in very different sentences being tested, we present sample items for each factor below as a pair of first a sentence with a repeatable predicate, and then a non-repeatable predicate.

The choice of quantifier directly tests the syntactic template matching hypothesis. If the availability of templates is critical to the illusion effect, then sentences with *fewer* (which may not be reanalyzed as an adverbial comparative) should be significantly degraded.[11]

(25) QUANTIFIER

    a. **More**/**fewer** girls ate pizza than than the boy did.

    b. **More**/**fewer** girls graduated high school than the boy did.

The application/non-application of ellipsis directly tests the repair-by-ellipsis hypothesis. If applying ellipsis is critical to the illusion effect, then sentences without such application should be significantly degraded.

(26) ELLIPSIS

    a. More girls ate pizza than the boy **did**/**ate yogurt**.

    b. More girls graduated high school than the boy **did**/**failed out**.

The possibility of the *than*-clause subject's being included in the extension of the matrix subject is critical to the *more*-ambiguity hypothesis. In cases where this inclusion is *not* supported, acceptability should significantly decline. Above, we presented examples where inclusion is not supported, in (27), we present first a repeatable and then a non-repeatable example where inclusion is supported.

(27) SUBJECT INCLUSION

    a. More **customers** call to complain than **she** does.   *repeat*

    b. More **poets** published their first book than **he** did.   *non-repeat*

Two sample items are given schematically in Figure 1, the top diagram representing an item with a repeatable VP, and the bottom a non-repeatable VP. Neither sample item supports subject inclusion. Starting at the left with choice of QUANTIFIER, through the predicate, on to either illusion (singular subject) or control (plural subject), to ELLIPSIS, one follows any path through the state diagram to generate a single trial for that item. Each diagram thus represents 8 trials.

Experiments 1a-1c differ only in which properties we assigned to the *than*-clause subject in illusion conditions. In Expt. 1a, we used only singular 3rd person pronouns and definite descriptions. In Expt. 1b, for reasons we discuss below, we used only singular first person pronouns (and so SUBJECT INCLUSION was always possible). In Expt. 1c, we counterbalanced 3rd person pronouns and definite descriptions with 1st person pronouns.

3.1.2. Results

Unless otherwise stated, all reported statistics are the result of by-subjects ANOVAs, which were consistent with the by-items analyses.

---

[11]In addition, this can be seen as a test of the syntactic version of the event comparison hypothesis. If speakers are interpreting the illusions as a comparison of events via syntactic reanalysis of the determiner *more* as an adverbial, this should be more difficult with *fewer*.
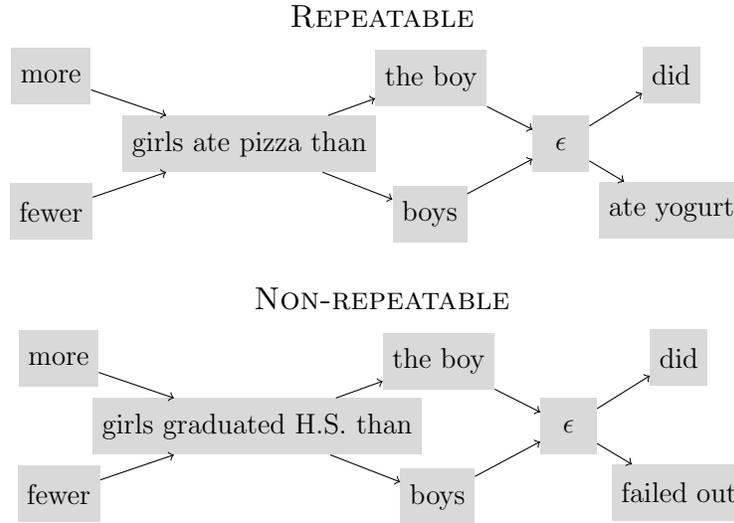
FIGURE 1. Schemata for repeatable and non-repeatable items in Expts. 1a-c, representing 16 unique sentences. Factors represented are PREDICATE TYPE (between items—repeatable, non-repeatable), QUANTIFIER TYPE (*more*, *fewer*), illusions (*the boy*) versus controls (*boys*), ELLIPSIS (presence, absence).

Experiment 1a. Illusions trials contained only singular 3rd person pronouns and definite descriptions, for maximal similarity with control trials as in Figure 1; $n = 16$.

We found a highly reliable, highly significant main effect of illusions versus controls (illusions 3.08, controls 5.86; $F = 95.5, p < 0.001$) and PREDICATE TYPE (repeatable 4.69, non-repeatable 4.25; $F = 26.6, p < 0.001$), as well as a significant interaction between illusions and controls with PREDICATE TYPE ($F = 9.3, p < 0.05$). This effect was due to a large difference in means within the illusions (repeatable 3.44, non-repeatable 2.73) but not within the controls (repeatable 5.95, non-repeatable 5.77). These results support the event comparison hypothesis.

There was a main effect of QUANTIFIER (*more* 4.59, *fewer* 4.35; $F = 11.2, p < 0.05$), but no interactions between this factor and illusions versus controls ($F = 0.2, p = 0.704$). An effect of quantifier was predicted by the template matching hypothesis, but not in the manner we found. Under that hypothesis, *more* should facilitate the illusion whereas *fewer* hinder it. If this account were supported, we would have seen an interaction between this factor and illusions versus controls, but this was not the case ($F = 0.1, p > 0.05$). There were no other effects.

We failed to find evidence for any account other than event comparison, but it is not yet clear that such evidence can't be found. In general, the average rating for illusion sentences were much lower compared to controls, which could mean that any effect of other factors was washed out.

Experiment 1b. Expt. 1a featured 3rd person pronouns and definite descriptions in the illusion conditions, but the classic examples feature singular 1st person pronouns. In this experiment, we use only singular 1st person pronouns; $n = 24$.

---

[11]3 trials were excluded because of lack of participant response.

Again we found a highly significant main effect of illusions versus controls (illusions 3.22, controls 5.75; $F = 97.4, p < 0.001$) and of PREDICATE TYPE (repeatable 4.78, non-repeatable 4.19; $F = 47.8, p < 0.001$), and an interaction effect between these ($F = 25.5, p < 0.001$). As in Expt. 1a, this interaction was caused by an asymmetry in the means for illusions (repeatable 3.73, non-repeatable 2.72), but not in the controls (repeatable 5.82, non-repeatable 5.67). These results support the event comparison hypothesis.

Unlike in the preceding experiment, we found a slight main effect of ELLIPSIS (elision 4.63, no elision 4.35; $F = 9.8, p < 0.05$), and an interaction between this factor and QUANTIFIER ($F = 6.82, p < 0.05$). In particular, non-application of ellipsis resulted in greater degradation for *more* (elision 4.83, no elision 4.32) than *fewer* (elision 4.42, no elision 4.37). There was also a significant 3-way interaction between illusions versus controls with ELLIPSIS and PREDICATE TYPE ($F = 11.67, p = 0.001$), driven by the fact that ellipsis boosted ratings for illusions with non-repeatable predicates (elision 2.95, no elision 2.49) but not with repeatable predicates (elision 3.81, no elision 3.66), while failing to do so within controls. There were no other effects.

That ELLIPSIS did not have the predicted effect of distinguishing ratings between illusions and controls failed to provide support for the repair by ellipsis hypothesis. Rather, application of ellipsis seems to have the effect of boosting ratings across the board. As in the preceding experiment, we found fairly low ratings for illusion-type sentences, which may suggest that the choice of 1st person pronouns over 3rd person pronouns and descriptions does not distinguish illusory versus non-illusory comparatives. Lower ratings may be due to subjects 'catching on' to these oddball examples, which would be facilitated by a repetition in the form of the *than*-clause in illusions trials.

Experiment 1c. Fults & Phillips (2004) reported fairly high ratings for illusion-type sentences. Inspection of their materials reveals that they allowed for plural subjects on illusions trials, which our Expts. 1a,b did not. In this experiment, we counterbalanced *than*-clause subjects of illusions trials along the person/pronoun/number dimensions; $n = 24$.

Again we found a main effects of illusions versus controls (illusions 3.89, controls 5.82; $F = 181.1, p < 0.001$) and of PREDICATE TYPE (repeatable 5.09, non-repeatable 4.62; $F = 20.1, p < 0.001$), as well as a significant interaction between these ($F = 5.8, p < 0.05$). This interaction, as in Expts. 1a,b, was due to higher ratings for repeatable than non-repeatable predicates within the illusion conditions (repeatable 4.26, non-repeatable 3.53) but not within the controls (repeatable 5.93, non-repeatable 5.70). These results support the event comparison hypothesis.

Unlike in Expt. 1b, and as in Expt. 1a, we found no effect of ELLIPSIS (elision 4.95, no elision 4.76; $F = 3.25, p = 0.085$), and no significant interactions. Like Expt. 1a, but unlike Expt. 1b, we found a main effect of QUANTIFIER (*more* 5.03, *fewer* 4.67; $F = 13.7, p < 0.05$), yet there were no interactions of either ELLIPSIS or QUANTIFIER with illusions versus controls. It appears that, in general, participants prefer *more* and application of ellipsis over *fewer* and lack of such application.

Unlike in Expts. 1a,b, we found an effect of SUBJECT INCLUSION, but it was in the opposite direction to that predicted by the additive *more* hypothesis (inclusion

TABLE 2.   Means of **illusions trials** by factor, and significant inter-actions across Expts. 1a-1c. - $p > 0.05$, * $p < 0.05$, ** $p < 0.001$.

Expt.	QUANTIFIER more	fewer	p	ELLIPSIS yes	no	p	SUBJECT INCLUSION yes	no	p	PREDICATE TYPE rep.	non-rep	p
1a	3.18	2.98	-	3.14	3.03	-	2.91	3.23	-	3.44	2.73	*
1b	3.31	3.14	-	3.38	3.07	-	n/a	n/a	n/a	3.73	2.72	**
1c	4.07	3.71	-	4.00	3.78	-	3.46	4.18	**	4.26	3.53	**

possible 3.46, not possible 4.18; $p < 0.001$ Welch two sample). There were no other effects.

Summary. The interest of Expts. 1a-c was in testing four factors that may have an effect on the acceptability of illusion-type sentences while not affecting controls. The results of our manipulations on the illusions trials are summarized in Table 2.

   We find that the only consistent effect is due to PREDICATE TYPE: that illusion-type sentences with non-repeatable VPs are rated significantly worse than those with repeatable VPs.

3.1.3. Discussion

The results of our first set of experiments suggest that *more* can be replaced with its dual, *fewer*, without affecting acceptability, and undoing the ellipsis in the *than*-clause has no effect, or no consistent effect on acceptability. A potential inclusion relation between the *than*-clause subject with the matrix clause subject had, in just one experiment, the opposite of the effect predicted by the additive *more* hypothesis (higher ratings were given when no inclusion was possible). Changing the repeatability of the VP, in contrast, consistently degraded acceptability. The significance of these results for the hypotheses under consideration is given schematically in Table 3.

TABLE 3.   Predictions by hypothesis and factor, with indications that they were met (✓) versus not (×).

Hypothesis	Quantifier *more—fewer*	Ellipsis yes—no	Subject inclusion yes—no	Repeatable VP yes—no
Template matching	> ×	-	-	-
Repair by ellipsis	-	> ×	-	-
Additive *more*	-	-	> ×	-
Event comparison	-	-	-	> ✓

   Informally, we saw that substituting 1st for 3rd person pronouns (e.g. *she*) and descriptions (e.g. *the girl*) didn't affect the overall acceptability of the illusions (Expts. 1a,b, means of 3.08 and 3.22 respectively), but adding in some instances of plural *than*-clause subjects (e.g. *the girls*) seemed to raise the overall ratings quite a bit (Expt. 1c, mean of 3.89). This finding may support the event comparison hypothesis, in that plural subjects allow for a plurality of events even with non-repeatable predicates (e.g. the graduations described in *than the girls graduated from college*). As we noted above, this points to the second route of satisfying the semantic requirements of *more*: reanalyzing the subject of the *than*-clause as plural.

In the next study, we explicitly test the effect of varying different properties of the *than*-clause subject.

3.2. Experiment 2: Acceptability

We observed that repeatability of the VP had a highly significant effect on acceptability within illusion conditions. However, there is more than one way to the plurality of individuals required by *more* in the *than*-clause: where a plural VP allows for the indirect method of counting individuals, namely through counting events, a plural subject allows for the direct method of counting (whether or not this is grammatically licit).

Given the strength of the repeatability manipulation, the event comparison hypothesis can be refined. If the crux of this hypothesis is just that a plurality of participations be, in principle, possible, then changing person (1st, 3rd) or NP-type (pronoun, description) features of the *than*-clause subject will not affect acceptability, but changing number (singular, plural) will. In this experiment, we systematically vary these properties in a within-subjects design. If number boosts the ratings for illusion-type sentences, we take this as further support for the event comparison hypothesis.

We tested 24 University of Maryland undergraduates, all native speakers of American English as determined by a pre-test questionnaire. Participants were given course credit or $10 for participating.

3.2.1. Design

We manipulated PLURALITY (singular, plural), PERSON (1st, 3rd person) and NP-TYPE (pronoun, description). As some combinations of these factors are in principle impossible (i.e., 1st person descriptions), the design is not a Latin Square. A guide to the conditions is given in Table 4. As in Expts. 1a-c, a bare plural subject of the *than*-clause is used as a control.[12]

TABLE 4. Sample items (repeatable/non-repeatable) by condition in Expt. 2.

	Condition	Sentence
I	SG-1-PRO	More girls ate pizza/graduated high school than **I** did
II	PL-1-PRO	More girls ate pizza/graduated high school than **we** did
III	SG-3-DEF	More girls ate pizza/graduated high school than **the boy** did
IV	PL-3-DEF	More girls ate pizza/graduated high school than **the boys** did
V	SG-3-PRO	More girls ate pizza/graduated high school than **he** did
VI	CONTROL	More girls ate pizza/graduated high school than **boys** did

In order to replicate the result from Expts. 1a-c, half of our items featured repeatable VPs, and the other half non-repeatable VPs (PREDICATE TYPE manipulation).

---

[12]We did not test the combination that would deliver the subject *they* (3rd person plural pronoun), as we find that, intuitively, all such examples sound contradictory and should thereby independently give rise to low acceptability. That is, *they* seems to resume the matrix bare plural; consider *More girls ate pizza than they did.*

Beyond simple replication, we reasoned that Expt. 2 had the potential to test for additive effects of subject or predicate plurality. SUBJECT INCLUSION was counterbalanced across items in SG-3-PRO, but was always in principle possible with SG-1-PRO and PL-1-PRO. The CONTROL condition, as before, features a bare plural *than*-clause subject. We created 36 items which, when combined with 108 filler sentences, were distributed across 6 questionnaires.

3.2.2. Results

Comparing conditions III and V which are both singular and 3rd person (SG-3-DEF and SG-3-PRO), we found no difference in acceptability for NP-TYPE (pronoun 4.22, definite 4.06; $p = 0.472$, Welch two sample t-test). Similarly, comparing conditions I and V which are both pronominal and singular (SG-1-PRO and SG-3-PRO), we found no effect of PERSON (1st person 4.58, 3rd person 4.22; $p = 0.110$, Welch two sample t-test). These results confirm our suspicion that these properties are non-essential to the illusions' acceptability.

Rather, we expected PLURALITY and PREDICATE TYPE to make a difference in acceptability. Thus, we conducted ANOVAs within the pairs of conditions I and II (SG-1-PRO, PL-1-PRO) and III and IV (SG-3-DEF, PL-3-DEF). With respect to the comparison of I and II, we observed a significant main effect of PREDICATE TYPE (repeatable 5.19, non-repeatable 4.38; $p < 0.001$), and no other effects. With respect to comparison of III and IV, we observed a significant main effect of PLURALITY (plural 5.25, singular 4.06; $p < 0.001$) and PREDICATE TYPE (repeatable 4.88, non-repeatable 4.43; $p = 0.004$). There were no interactions.

Finally, we tested for an effect of PREDICATE TYPE in conditions V (SG-3-PRO) and VI (CONTROL). There was a significant effect within V ($p = 0.002$, Welch two sample t-test). However, there was *no* effect within CONTROLs ($p = 0.248$, Welch two sample t-test). As observed in the previous experiments, the PREDICATE TYPE manipulation has no effect when the *than*-clause subject is a bare plural.

We found no significant differences for SUBJECT INCLUSION for SG-3-PRO.

The mean ratings are presented split by singular/plural subjects, condition, and PREDICATE TYPE are plotted in Figure 2.

3.3. Discussion

The illusion-type sentences reach nearly the acceptability level of their non-illusory counterparts in cases where (a) the elided VP is repeatable, and (b) the *than*-clause subject is plural. These results provide support for the conclusion that the phenomenon of comparative illusions has a non-trivial connection to the interpretation of comparative quantification.

That we do not find evidence that string identity of the comparative quantifier (determiner versus adverbial) matters suggests that an account of the illusions in terms of superficial rule applications during syntactic parsing is unlikely to bear fruit. In particular, the sentence template-matching approach that we considered in §2.1 is not supported. Correspondingly, the interpretation of the event comparison hypothesis that posits syntactic repair of *more* from a determiner to an adverbial position
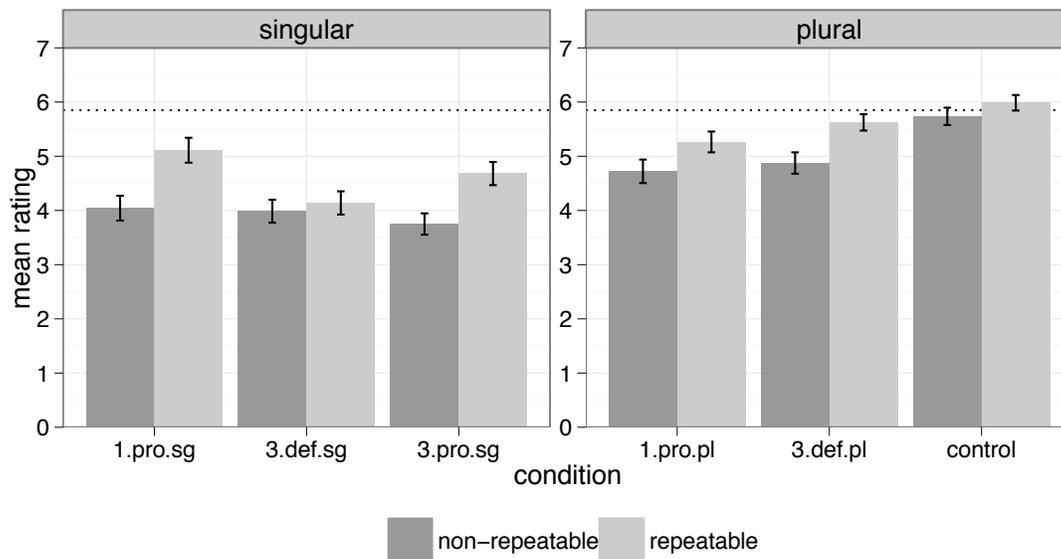
FIGURE 2. Plot of mean ratings by condition in Expt. 2, plotted by repeatability and split by singular and plural conditions.

was not supported. Similarly, while ellipsis can sometimes repair *syntactic* anomaly, it is not clear that ellipsis is playing that kind of role suggested in §2.2 here. Lastly, it is unlikely that the illusion turns on a specific lexical ambiguity between comparative and additive *more*, as outlined in §2.3.

It is possible that the acceptability judgments we received reflect, to some degree, how likely a participant is to notice that there is no stable sense associated with an illusory comparative. One could imagine that the reduced scores in our acceptability judgment studies reflect, to some degree, participants attempting to find a sense for illusion sentences and failing, and so down-grading it as a result. For the linguist, the interest in the classic illusion is that it sounds like a well-formed sentence of the language *even while* one acknowledges that it lacks any clear sense. Something appears to be happening that is automatic and irresistible in the processing of such examples.

In our last and final experiment, we attempt to gain insight into how the illusions are interpreted.

3.4. Experiment 3: Recall

The acceptability studies reported in the previous sections provided evidence that the availability of event comparison modulates acceptability of the illusions. But how does this hypothesis relate to how they are actually perceived? To get at this question, we conducted an experiment using the sentence recall paradigm developed by Potter & Lombardi (1990) which was designed to reveal how speakers represent sentences beyond surface syntax. It is not a verbatim task, but requires selective analysis of input and planning of output which can sometimes reveal important differences in
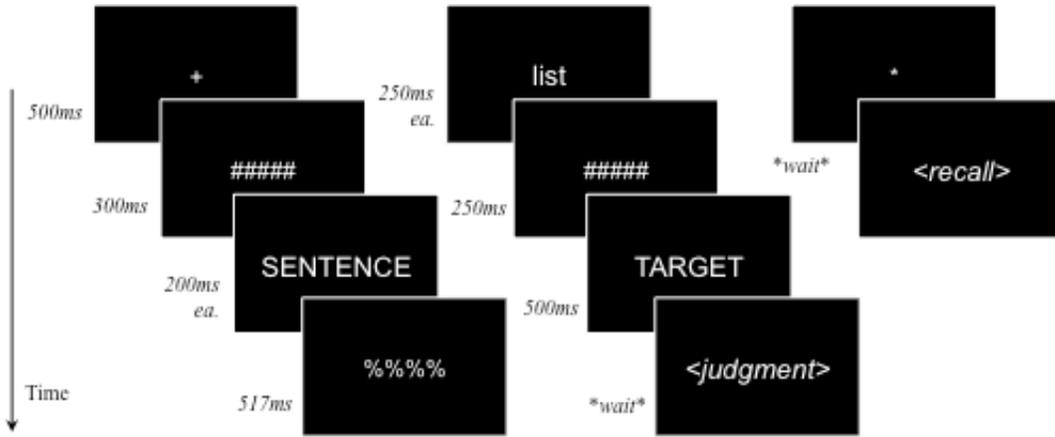
FIGURE 3. Schematic presentation of procedure from Expt. 3.

encoding. Of interest for us is the question of whether speakers are actually interpreting illusion-type sentences as stating a comparison of events. We cannot base firm conclusions about reanalysis based on judgment data alone.

The event comparison hypothesis came in two flavors, both of which would explain the acceptability judgment data. The first flavor was syntactic: perhaps speakers are syntactically reanalyzing *more* as an adverbial rather than as a determiner.[13] The second flavor was semantic: perhaps speakers semantically reanalyze the comparative quantifier as counting individuals indirectly via the events they participate in. Regardless of flavor, there are two routes available to satisfying the plurality requirement of the comparative quantifier: the first is pluralization of the VP, the second is pluralization of the subject NP in the *than*-clause.

If the syntactic version of the event comparison hypothesis is correct, then we should find evidence in this task that speakers have "moved" *more* to an adverbial position. In contrast, if the semantic version is correct, we should find evidence for pluralization in both the nominal and verbal domain; i.e., speakers will do whatever they can to reach a plurality of events. Of course, it may be that the results of the acceptability studies were something of an illusion themselves, and perhaps speakers fail to misencode such sentences entirely, and will fail to recall them in the general case.

We report results from 24 University of Maryland undergraduates, all native speakers of American English as determined in a pre-test questionnaire. An additional 10 subjects were excluded: 1 failed to respond (or responded with a non-sentence) on >75% of experimental trials; 2 failed to understand the task (e.g., recalling distractor words rather than the target sentence during the recall phase); and the data from an additional 7 were not collected due to scripting error. Participants were given course credit or $10 for participating.

---

[13]Indeed, the manipulation of *more/fewer* in Expts. 1a-c addressed this question to some degree. However, it is possible that this syntactic option doesn't rely on string identity of the comparative quantifier, in which case the lack of degradation with *fewer* would not be informative.

ASPECT

	continued		the old man was/did
Last year more young people		reading War & Peace than	
	began		old men were/did

OBJECT

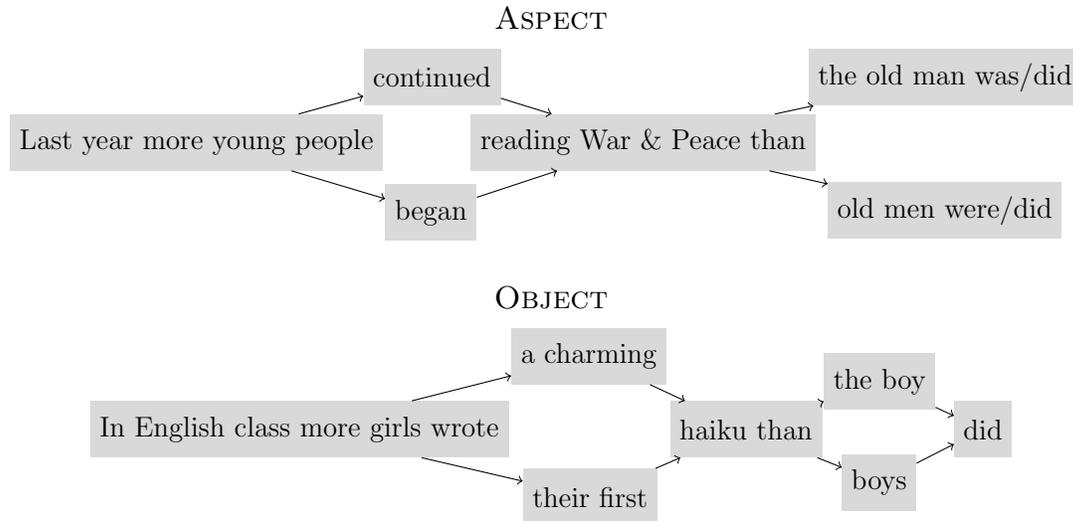|	a charming		the boy	
In English class more girls wrote		haiku than		did
	their first		boys	

FIGURE 4. Schemata for repeatable and non-repeatable items in Expt.3, representing 8 unique sentences.

3.4.1. Design

We follow the method of Potter & Lombardi (1990) quite closely. The recall task contains an initial rapid serial visual presentation (RSVP) of a sentence immediately followed by a distractor task, and finally a prompt to verbalize the initial sentence. The distractor task is designed to occupy working memory so that the sentence to be recalled has to be stored and ultimately retrieved from long term memory, decreasing the likelihood of simple verbatim recall of the sentence.

The procedure works as follows: first, a fixation cross appears for 500ms, followed by a visual mask for 300ms. In the SENTENCE phase, the words of a sentence appear RSVP with a presentation duration of 200ms each, ending with a visual mask for 517ms. In the DISTRACTOR phase, a list of five words appears RSVP for 250ms per word, ending with a visual mask for 250ms. Next, a capitalized word appears for 500ms, at which point indicates whether the capitalized word had been in the list of words, pressing 'F' for 'yes' or 'J' for 'no'. After making this judgment, the RECALL phase begins, signaled by an asterisk. Participants were given unlimited time to recall the sentence they initially read. Before the experiment proper, each participant was given 6 practice trials to familiarize themselves with the procedure. The schematic in Figure 3 summarizes these details.

Our design included two types of items that differed in their locus of (non-)repeatability. ASPECT items were non-repeatable if they had initiative or terminative aspectual verb introducing their VP (*started/finished reading*) and repeatable if they had continuative aspectual verbs or forms of *to be* (*continued/were reading*). OBJECT items were non-repeatable if they had a superlative or terminative/initiative adjective (*first/final strawberry cupcake*), and repeatable otherwise (*tasty strawberry cupcake*). Sample items are given in Figure 4. The range of words per sentence was 11-17 (mean 14.2). The *than*-clause subject was always singular, but included pronouns, descriptions, and proper names.

TABLE 5. Overview of recall for all items in Expt. 3.

	Controls			Illusions				
ALL *items*	*Obs.*	*Excluded*		*Obs.*	*Excluded*		*Binomial*	χ^2
Non-rep	138	16	11.6%	138	21	15.2%	0.0937	-	
Rep	138	9	6.5%	138	21	15.2%	0.0133	*	0.1615 -
Sum	**276**	**25**	**9.1%**	**276**	**42**	**15.2%**	**0.0113**	*****	
ALL *items*	*Obs.*	*Included*		*Obs.*	*Included*				
Non-rep	138	122	88.4%	138	117	84.8%	0.0489	*	
Rep	138	129	93.5%	138	117	84.8%	0.0380	*	0.6586 -
Sum	**276**	**251**	**90.9%**	**276**	**234**	**84.8%**	**0.0269**	*****	
ALL *items*	*Net*	*Accuracy*		*Net*	*Accuracy*				
Non-rep	122	87	71.3%	117	83	70.9%	0.0611	-	
Rep	129	96	74.4%	117	83	70.9%	0.0597	-	0.5059 -
Sum	**251**	**183**	**72.9%**	**234**	**166**	**70.9%**	**0.0427**	*****	

Lists of words for the distractor task were constructed out of sets of 5 words that were matched for character length (3-7 characters per set), and controlled for phonological and semantic similarity (to each other, and to the elements of the sentence they were paired with). The target word was present in the initial set on only half of the trials, for a target 50/50 split in 'yes' and 'no' responses. Our fillers contained no ungrammatical sentences. However, the fillers did contain a small number of other, grammatical forms of comparatives and superlatives, which we hoped would decrease the salience of the experimental trials.

3.4.2. Results

Out of a total of 276 trials each in the bare plural and illusions conditions (see Table 5), 9.1% of bare plural and 15.2% of illusion trials were excluded. Excluded data consisted of non-responses (no audible response) and failure to recall a comparative sentence. By this measure, recall of the illusions was statistically worse (bare plural 25, illusion 42, binomial $p = 0.027$), and there was no interaction with repeatability, $\chi^2(1, N = 67) = 1.96, p = 0.162$. Of the remaining trials, accuracy on the distractor task was statistically worse on illusion trials (bare plural 72.9%, illusion 70.9%, binomial $p = 0.043$), and there was no interaction with repeatability, $\chi^2(1, N = 349) = 0.44, p = 0.506$.

We coded two types of errors in recall that were not specific to either of the ASPECT or OBJECT items. The first is displacing *more* to an adverbial position. For this measure, we compared the control trials to the illusion trials, expecting, if the syntactic version of the event comparison hypothesis is correct, that *more* would be displaced more in illusions as opposed to control trials. The second error we coded was pluralizing the *than*-clause subject, only within illusion trials only.[14] Hypothesizing that sentences that are rated lower in acceptability will give rise to greater evidence

---

[14]We would not expect that participants would singularize a bare plural subject of the *than*-clause, as this would make the sentence ungrammatical. However, we did code for this: it happened a total of twice in the entire experiment.

for repair, we expected to see a greater degree of pluralization in the non-repeatable trials.

An example of the first type of error is given in (28)). Such displacement rarely occurred in either bare plural or illusion trials (bare plural 12, illusions 11; binomial $p = 0.165$) and there was no interaction with repeatability, $\chi^2(1, N = 23) = 2.15, p = 0.142$, suggesting that such displacement occurred no more in illusions than in control trials.

(28) **Moving** *more* **error:**
    Target: **More** girls ate pizza than I/boys did.
    Recalled: Girls ate pizza **more** than I/boys did.

An example of the second type of error is given in (29). We found that participants overwhelmingly pluralized the *than*-clause subject in both illusions conditions, though the difference between the two conditions was not significant (repeatable 55, non-repeatable 61; binomial $p = 0.063$).

(29) **Pluralizing** *than*-**subject error (illusions):**
    Target: More girls ate pizza than **the boy** did.
    Recalled: More girls ate pizza than **the boys** did.

We were surprised at this lack of a difference, so we examined our illusions items more closely. We found that illusions items containing singular definites as the subject of the *than*-clause drove the pluralization effect.[15] When we looked just at our 87 non-repeatable and 92 repeatable illusion trials containing singular definites, we found that participants pluralized the definite on repeatable trials at a rate of 57.6% (53/92), but at 73.5% (64/87) in non-repeatable trials! This difference was significant in the predicted direction (binomial $p = 0.031$).[16]

The next two analyses are specific to each of the item types ASPECT and OBJECT. (The summary is presented in Table 6).

With ASPECT items, we considered how often the aspectual verb was recalled with a continuative rather than initiative/terminative verb (e.g., *begin → continue*), or vice versa (e.g., *continue → begin*; see VP-number in Table 6. Examples are given in (30-31)). The first type was coded as pluralizing (as "making repeatable") and the second singularizing (as "making non-repeatable"). We found no significant difference between the bare plural and illusions with respect to how often participants made these category changes (bare plural 25, illusions 32, binomial $p = 0.051$). However, we did find a significant interaction with repeatability, $\chi^2(1, N = 57) = 19.13, p < 0.001$, driven by a boost in pluralizing in non-repeatable illusion trials.[17]

(30) **"VP-number" error (singular → plural):**
    a.   *Singular target*:   More girls **began writing** than the boy did.
    b.   *"Plural" recalled*: More girls **continued writing** than the boy did.

(31) **"VP-number" error (plural → singular):**

---

[15]Only 5 items had pronominal or proper name subjects.

[16]On a hunch, we compared these rates of pluralization with that of singular definites in filler items. We found that, of 143 filler trials containing singular definite descriptions, the rate of pluralization of the definite was 54.2% when that trial contained a clausemate bare plural (32/59), in contrast to only 3.6% in the absence of a bare plural (3/84). The rate of 54.2% is comparable with the rate of

TABLE 6. Overview of errors coded for ASPECT and OBJECT items in Expt. 3.

		Controls			Illusions					
ASPECT *items*		*Obs.*		*Errors*	*Obs.*		*Errors*	*Binomial*		χ^2
VP-number	Non-rep	60	20	33.3%	57	25	43.9%	0.078	-	
	Rep	65	5	7.7%	59	7	11.9%	0.173	-	<0.001 ***
	Sum	**125**	**25**	**20.0%**	**116**	**32**	**27.6%**	**0.051**	**-**	
OBJECT *items*		*Obs.*		*Errors*	*Obs.*		*Errors*	*Binomial*		χ^2
AP-delete	Non-rep	62	24	38.7%	60	30	50.0%	0.070	-	
	Rep	64	37	57.8%	58	36	62.1%	0.089	-	0.069 -
	Sum	**126**	**61**	**48.4%**	**118**	**66**	**55.9%**	**0.051**	**-**	

a.   *"Plural" target*:   More girls **continued writing** than the boy did.
b.   *Singular recalled*:   More girls **began writing** than the boy did.

With OBJECT items, we considered how often a critical adjective was deleted. We considered a critical adjective to be one like *first*, in contrast to one like *tasty* (AP-delete in Table 6; examples are given in (32)). If a superlative or terminative/initiative adjective like *first* or *final* is deleted, the predicate is repeatable. There is no such effect if a regular adjective like *tasty* is deleted. We found no significant difference between bare plurals and illusions in the rate of deletion of the critical adjective (bare plurals 61, illusions 66, binomial $p = 0.051$), and no significant interaction with repeatability, $\chi^2(1, N = 127) = 3.32, p = 0.069$. (But see discussion below.)

(32)   **"AP-delete" error:**
       Target: More girls ate **a tasty/their first** strawberry cupcake.
       Recalled: More girls ate **a/their** strawberry cupcake.

3.5. Discussion

The sentence recall task can provide a window into how speakers interpret sentences. Unlike verbatim recall tasks, speakers are given a working memory task after exposure to a sentence which enforces non-verbatim storage and recall. The suggestion has been that it is the meaning, and not the sentence form, that is stored. If this picture is roughly correct, then the high level of recall of the illusions itself suggests speakers have access to a non-syntactic level of representation for the illusions.

The event comparison hypothesis contends that speakers do interpret the illusions at some level, analyzing a grammatical comparison of numbers of individuals as a

---

pluralization of our repeatable illusions, and both are significantly different from our non-repeatable illusions, $\chi^2(1, N = 149) = 21.88, p < 0.001$.

[17]A potential worry is that some of our ASPECT items alternated phrases like *begin writing* and *were writing*. That is, one may worry that we contrasted non-repeatable trials with a 'more specific/contentful' verb, with repeatable trials with a 'less specific/contentful' verb (where *begin writing*/*continue writing* would be considered matched for specificity). This affected 4 out of 12 of our ASPECT items. Of the bare plural trials, 44 out of 125 trials were 'unbalanced' pairs, and 81 'balanced'. Of the illusions trials, 42 out of 116 were 'unbalanced', and 74 'balanced'. Considering 'balanced' and 'unbalanced' sets respectively, the interaction with repeatability persists in each ('unbalanced', $\chi^2(1, N = 10) = 12.40, p < 0.001$; 'balanced', $\chi^2(1, N = 22) = 8.81, p = 0.003$).

comparison of numbers of events. There are two ways this interpretation could be arrived at, one syntactic and the other semantic. We found little evidence for a purely syntactic strategy: syntactically displacing *more* to an adverbial position would make at least the repeatable illusion targets both grammatical and fully interpretable, yet speakers chose to leave *more* in a determiner position in 96.5% of illusions trials, and there was no difference between repeatable/non-repeatable trials. This suggests that participants are syntactically faithful when they perceive the illusion, and that the event comparison reading indeed comes via NP, and not VP. This result could not have been detected using acceptability data.

On the face of it, the manipulation that defined OBJECT trials failed to support the event comparison hypothesis: the difference between controls and illusions trials with respect to deletion of the adjective was marginal, as was the interaction with repeatability. In non-repeatable trials, the adjective was a semantically loaded expression like *first*, whereas in repeatable trials it was a relatively trivial adjective like *tasty*. In the control trials, we found that the trivial adjective was deleted significantly often more than the loaded adjective (repeatable/trivial 37, non-repeatable/loaded 24; binomial $p = 0.031$), suggesting to us that, in general, participants like to retain semantically loaded expressions. Interestingly, in the illusions trials, there was only a marginal difference in the deletion of the trivial and loaded adjectives (repeatable/trivial 36, non-repeatable/loaded 30; binomial $p = 0.067$). This relatively lower rate of deletion of the non-repeatable/loaded adjective compared to controls could reflect a tension between wanting to *retain* such expressions, with the fact that its retention results in a less interpretable comparative (i.e., it delivers a non-repeatable event description).

Regardless, the overall pattern of results provides clear evidence for the semantic interpretation of the event comparison hypothesis. This version says that there are two ways to get to a comparison of events: via composition of a singular *than*-clause subject with a plural event description, or a plural subject itself. Overall, we found that participants pluralized VPs significantly more often than singularized them, and this difference was magnified in the illusions: proportionally, non-repeatable predicates were pluralized more than repeatable predicates were singularized in the illusions over the controls. As there is no obvious syntactic difference between VPs like *begin reading the book* and *continue reading the book*, this suggests that the effect is semantically motivated. Participants are faithful to the syntax of nominal comparatives when processing the illusions (evidenced by the lack of displacement of *more* to an adverbial position), suggesting that the event comparison interpretation is achieved as in Krifka's *4000 ships* sentences.

One worry might be that this difference in VP repeatability doesn't yet tell us much, given how often participants pluralized the *than*-clause subject. Once the subject has been pluralized, event comparison is secured. We conducted a post-hoc analysis within the ASPECT illusions trials to see whether VP pluralization occurred more or less when the *than*-clause subject had been pluralized. The results are striking: participants in fact *increased* pluralization of the VP on non-repeatable trials when they also pluralized the *than*-clause subject (not pluralized subject 8, pluralized subject 24, binomial $p = 0.014$). This gave rise to an interaction between subject pluralizing and VP number changes, $\chi^2(1, N = 32) = 10.17, p = 0.001$.

By far the biggest effect we found was pluralization of the *than*-clause subject in the illusions trials. The extremely high rate of pluralization in illusions trials, though, could suggest that the subject plurality road to event comparison is singularly tempting. One may wonder, however, that if participants simply pluralized the NP and deleted the determiner, they have simply grammaticalized the illusions. However, this cannot be the full story: we looked at the illusions conditions to contrast retention of the determiner when the *than*-clause subject was pluralized versus not. We found that, in ∼25% of cases where the *than*-clause subject was pluralized, the determiner was retained (non-repeatable 15/61, repeatable 13/55).

4. General discussion

This paper represents the first formal attempt to understand which properties of the classic comparative illusion are essential to its phenomenology, and which inessential. We considered four hypotheses as to the source of the illusion, running the gamut of less to more commitment to grammatical processes in facilitating the illusion. The first, loosely attributed to Townsend and Bever, suggests that a shallow parser is to blame. The second, pursued by Fults and Phillips, was that a general process of repair by ellipsis could be responsible. The third held that an ambiguity between comparative and additive *more* could be at fault. None of these hypotheses were supported, as they each (as near as we can determine) rely on properties of the classic illusion that our experiments deemed inessential.

The fourth hypothesis held that the possibility of salvaging a legitimate semantics for comparison underpinned the illusion of acceptability and interpretability. The main finding from the acceptability studies we report was that plural *than*-clause subjects and "plural" verb phrases facilitated the illusion. Expts. 1a-c and Expt. 2 showed that manipulating the repeatability of the VP had a large effect on acceptability. Expt. 2 showed that the only properties of the *than*-clause subject that mattered to acceptability was number, while person and NP-type had no effect. Other linguistic factors, like whether or not the comparative quantifier was *more* versus *fewer*, whether ellipsis had been performed, or whether the *than*-clause subject was a potential member of the group denoted by the matrix clause subject, did not have an effect on the illusion.

Our Expt. 3 replicated these findings in a different experimental methodology that places very different demands on speakers. Rather than passively reading a sentence with unlimited time and offering a metalinguistic judgment, speakers were presented with a sentence presented RSVP followed by an unrelated task requiring an unrelated kind of judgment, and then asked to reproduce the sentence. Overall, recall of the illusions was high. If one hypothesized that the source of the illusion was some kind of blanket misencoding, this result is unexpected. If the source is, rather, some type of repair process, then one would expect accurate recall but evidence for repair. Our results suggest that an explanation along the latter lines is correct.

In the data from the recall task, we found evidence for an increase in repairs for sentences that were rated less acceptable in the passive judgment studies. Illusions overall were rated worse than bare plural controls, and these evidenced a significantly

greater degree of repair. Similarly, non-repeatable illusions were rated worse than repeatable illusions, and these too evidenced a significantly greater degree of repair.

The most interesting results, though, were the types of repair strategies participants utilized. Participants nearly universally respected the basic syntax of the nominal comparative, but within these confines engaged in semantic repairs of illusion-type sentences that closely track the required semantics of comparatives. Such semantic repairs came in two varieties: pluralizing the *than*-clause subject (often while retaining the determiner), and pluralizing the VP. Participants did not do things like move *more* to an adverbial position, which would, at least for repeatable illusions, solve the syntactic and semantic problems of the illusion immediately. This pattern suggests that the illusions do not stem from a process by which speakers try to get a grammatical sentence out of what they heard/read by any means possible; rather, they show a subtle pattern of syntactic faithfulness while attempting to fix up the semantics.

We take these results to suggest that speakers both parse and interpret illusion-type sentences deeply. The syntax of the illusions tells speakers that they are in nominal comparative mode, and as such they should compare cardinalities of individuals. Krifka has shown us that counting individuals sometimes amounts to counting events in which those individuals participate (counting their "participations"). Semantically, *more* requires a plural (or mass-like) domain for measurement/counting and comparison. Speakers' tendency on illusion-type trials in the recall task to pluralize the VP and/or the subject of the *than*-clause suggests that speakers make use of multiple strategies to deliver the pluralities count *more* requires.

This contribution to understanding the puzzles raised by the comparative illusions perhaps raises more questions than it answers. How do we arrive at these other interpretations, these event readings of nominals? If they are so freely available, why don't the comparative illusions *just mean* something like their corresponding adverbial comparative would? And, if we unconsciously repair strings of words to make them more meaningful, how widespread are these processes? How do they work? Methodologically speaking, how should we understand the relationship between acceptability ratings, and repair processes more generally? Is it the case that, as a general rule, lower acceptability ratings will lead to more evidence for repair, as our results suggest?

References

Barker, Chris. 1999. Individuation and quantification. *Linguistic Inquiry* 30(4). 683–691.

Bresnan, Joan. 1973. Syntax of the comparative clause construction in English. *Linguistic Inquiry* 4(3). 275–343.

Chomsky, Noam. 1977. Conditions on transformations. In *Essays on form and interpretation*, 81–162. New York, New York: Elsevier North-Holland, Inc.

Christianson, K., A. Hollingworth, J. Halliwell & F. Ferreira. 2001. Thematic roles assigned along the garden path linger. *Cognitive Psychology* 42. 368–407.

Cusic, David. 1981. *Verbal Plurality and Aspect*: Stanford University dissertation.

Ferreira, F., V. Ferraro & K. Bailey. 2002. Good enough representations in language comprehension. *Current Directions in Psychological Science* 11(11-15).

Frazier, Lyn & Jr. Clifton, Charles. 2011. Quantifiers undone: reversing predictable speech errors in comprehension. *Language* 87(1). 158–171.

Fults, Scott & Colin Phillips. 2004. The source of syntactic illusions. CUNY 2004 poster.

van Geenhoven, Veerle. 2005. Atelicity, pluractionality, and adverbial quantification. In Henk Verkuyl, H. de Swart & Angeliek van Hout (eds.), *Perspectives on aspect* vol. 1 Studies in theoretical psycholinguistics, 107–124. Springer.

Gibson, Edward & J. Thomas. 1996. Memory limitations and structural forgetting: the perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes* 14(3). 225–248.

Greenberg, Yael. 2009. Additivity in the domain of eventualities. Handout, Sinn und Bedeutung 14.

Greenberg, Yael. 2010. Additivity in the domain of eventualities (or: Oliver Twist's *more*). In Martin Prinzhorn, Viola Schmitt & Sarah Zobel (eds.), *Proceedings of Sinn und Bedeutung 14* 14, 151–167. Vienna.

Hackl, Martin. 2001. Comparative quantifiers and plural predication. In K. Megerdoomian & Leora Anne Bar-el (eds.), *Proceedings of WCCFL XX*, 234–247. Somerville, Massachusetts: Cascadilla Press.

Heim, Irene. 2000. Degree operators and scope. In Brendan Jackson & Tanya Matthews (eds.), *Proceedings of SALT X*, 40–64. Cornell University, Ithaca, NY: CLC Publications.

Kennedy, Christopher. 2003. Ellipsis and syntactic representation. In Kerstin Schwabe & Susanne Winkler (eds.), *The interfaces: Deriving and interpreting omitted structures* (Linguistics Aktuell 61), John Benjamins.

Krifka, Manfred. 1990. Four thousand ships passed through the lock: object-induced measure functions on events. *Linguistics and Philosophy* 13. 487–520.

Lasnik, Howard. 2001. When can you save a structure by destroying it? In Minjoo Kim & Uri Strauss (eds.), *Proceedings of the north east linguistic society 31*, 301–320. Georgetown University: GLSA.

Lechner, Winfried. 2004. *Ellipsis in Comparatives*. Berlin/New York: Mouton de Gruyter.

Lewis, Richard L. 1996. Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research* 25(1). 93–115.

Merchant, Jason. 2001. *The syntax of silence: sluicing, islands, and the theory of ellipsis*. Oxford: Oxford University Press.

Montalbetti, Mario. 1984. *After binding*. Cambridge: Massachusetts Institute of Technology dissertation.

Nakanishi, Kimiko. 2007. Measurement in the nominal and verbal domains. *Linguistics and Philosophy* 30. 235–276.

Pancheva, Roumyana. 2009. More students attended FASL than CONSOLE. In *Proceedings of fasl*, 1–18.

Potter, Mary C. & Linda Lombardi. 1990. Regeneration in the short-term recall of sentences. *Journal of Memory and Language* 29. 633–654.

Richards, Norvin W., III. 1997. *What moves where when in which language?* Cambridge, Massachusetts: Massachusetts Institute of Technology dissertation.

Ross, John Robert. 1969. Guess who? In Robert I. Binnick, Alice Davison, Georgia M. Green & Jerry L. Morgan (eds.), *Chicago linguistics society*, 252–286. Chicago, Illinois.

Sanford, A. & P. Sturt. 2002. Depth of processing in language comprehension: not noticing the evidence. *Trends in Cognitive Science* 6. 382–386.

Schein, Barry. to-appear. Conjunction reduction redux. Manuscript, USC.

Thomas, Guillaume. 2010. Incremental *more*. *Proceedings of SALT 20* 233–250.

Townsend, D.J. & T.G. Bever. 2001. *Sentence comprehension: the integration of habits and rules.* MIT Press.

Wason, P. & S.S. Reich. 1979. A verbal illusion. *Quarterly Journal of Experimental Psycholinguistics* 31. 591–597.

Wellwood, Alexis. to appear. Back to basics: *more* is always *much-er*. In *Proceedings of Sinn und Bedeutung 17*, .

Wellwood, Alexis, Valentine Hacquard & Roumyana Pancheva. 2012. Measuring and comparing individuals and events. *Journal of Semantics* 29(2). 207–228.