Granularity in the semantics of comparison[*]

Helena Aparicio
Cornell University

Curtis Chen
Massachusetts Institute of Technology

Roger Levy
Massachusetts Institute of Technology

Elizabeth Coppock
Boston University

Abstract This paper makes the novel observation that definite comparatives, such as *the bigger circle*, impose restrictions on the cardinality of the comparison class (CC) against which their truth conditions are evaluated. We show that the corpus frequency counts of definite comparatives sharply drop when the comparison class used for their interpretation is formed by more than two individuals. Two alternative theories of these distributional facts are considered and tested experimentally through an acceptability judgment task. According to the first theory, the **2-Individuals Theory**, definite comparatives presuppose that the CC is of cardinality 2; under the second theory, the **2-Degrees Theory**, the meaning of the comparative is evaluated against a granularity γ that maps the individuals in the CC to degrees in the relevant adjectival scale, and definite comparatives presuppose that the set of the degrees resulting from this mapping is of cardinality 2. Our results show that definite comparative descriptions are most frequent and felicitous when evaluated against comparison classes with two individuals, but also that acceptability drops off with higher cardinalities in a gradient manner that is sensitive to granularity. Taken together, these findings argue against the 2-Individuals theory of definite comparatives and lend support to the 2-Degrees theory.

1 Introduction

Definite *comparatives*—in which a comparative adjective like *bigger* attributively modifies a noun heading a definite description, as in (1a)—have similar meanings to
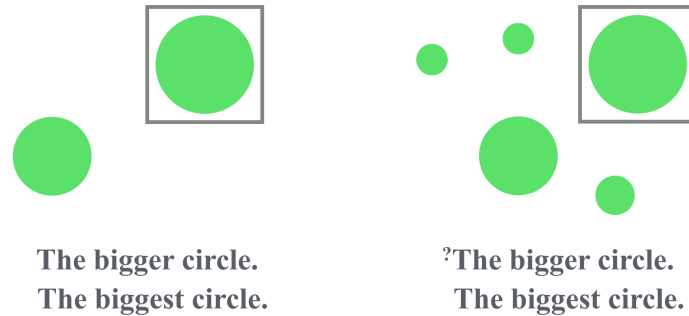

Figure 1 Left: A comparison class of two different sized circles. Right: A comparison class of five circles with different sizes. The gray box highlights the intended referent of the definite descriptions *the bigger/est circle*.

definite *superlatives* like (1b).

(1)   a. the bigger square   (definite comparative)

          b. the biggest square   (definite superlative)

Both examples in (1) pick out the largest square in some contextually given set of squares, called the 'comparison class' (henceforth CC). With a comparison class of size 2, the descriptions in (1) pick out the same referent. However, definite comparative and superlative descriptions diverge in the expectations they carry about the CC. In particular, they seem to impose different requirements on the cardinality of the CC used for their interpretation. This is exemplified in Figure 1, which represents two CCs of circles of different sizes that differ in their cardinality (2 vs. 5 circles). While definite superlatives seem to be equally acceptable in the two scenarios, the acceptability of definite comparatives decreases when judged against a CC of higher cardinality (see Figure 1, right scene). Though the literature on comparatives and superlatives is vast, to our knowledge this is the first time that this contrast between definite superlatives and comparatives is explicitly discussed in the semantics literature (with the exception of Coppock 2019).

    In Section 2, we provide corpus evidence that supports and helps us refine this novel observation. We show that while comparatives are most frequent with comparison classes of size 2, uses with CCs of higher cardinality are attested at a non-negligible rate. This observation leads us to the question: What exactly is the cardinality constraint for definite comparatives? Anecdotal evidence suggests that when the CC contains more than two individuals, definite comparatives are typically used in situations where there is one object that stands out with respect to the relevant adjectival property, while the remaining individuals are more or less on a par. For

Figure 2   Images accompanying *the larger of the three birds* (left) and *the larger of the three tombs* (right)

instance, attested uses of *the larger of (the) three* are not always paired with images, but Figure 2 shows two cases in which they are.[1] Coppock (2019) observes that in such cases, the smaller two objects of the relevant type (birds or tombs, respectively) are of similar size, and the largest one stands out. This observation raises the possibility that what matters is not the number of individuals in the comparison class *per se*, but the number of degrees along the relevant dimension (e.g., size) instantiated by members of the comparison class, at some level of granularity.

In light of this, we hypothesize that definite comparatives require that the CC instantiate two *degrees* (the **2D Theory**), rather than two *individuals* (the **2I Theory**). Section 3 introduces these two theories of definite comparatives, and discusses their predictions with respect to the role of granularity in acceptability judgments. In Section 4, we give results from an acceptability judgment task showing that acceptability is modulated by granularity in accordance with the 2D Theory.
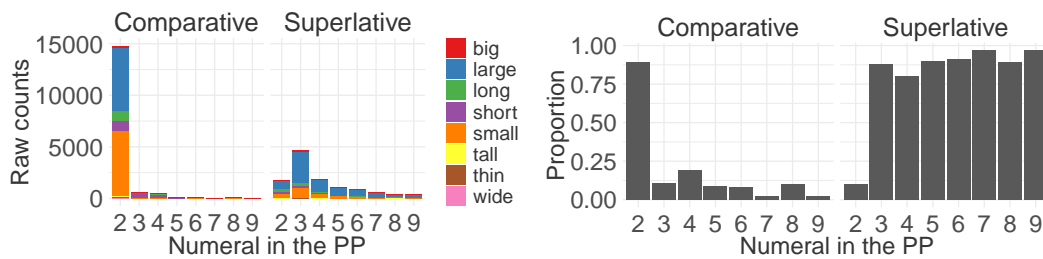
Figure 3 Internet Archive Corpus results (`the A-{er,est} of n`). Left: raw counts. Right: Proportion counts of adjective (e.g. *bigger*) to total (*bigger*+*biggest*) in a cardinality.

2 Corpus Evidence: Asymmetries in CC Cardinalities

To support the intuition that definite comparatives disprefer comparison classes with more than two individuals, we searched the Internet Archive Corpus[2] for strings of the form [*the* `Adjective` *of* `numeral`] (e.g. *the bigg{er,est} of two*), where the adjective was either in the comparative or the superlative form. Adding the prepositional phrase allowed us to target different CC cardinalities. Results are reported in Figure 3. The left panel plots the raw counts for each of the cardinalities included in our queries, i.e. 2-9, for a series of relative adjectives. The results confirm our original observation that definite comparatives have a preference for comparison classes of cardinality 2; this is shown by the stark dip in frequencies for cardinalities higher than 2. The opposite pattern is observed for superlative adjectives, with an increase from cardinality 2 to cardinality 3.

The results also show a tapering in the counts for the higher cardinalities in both the comparative and the superlative. To compare superlative vs. comparative in a way that holds constant the overall frequency associated with comparison classes of various sizes, we transformed the raw counts into proportions of a total, where the total for a given cardinality was the sum of the comparative and the superlative counts within that cardinality. These results are presented in the right panel in Figure 3. As seen in the plot, the decrease in raw counts observed for cardinalities over 2 in both the superlative and the comparative does not hold across the board once the data is visualized in the proportion space: Above 2, with the superlative, there is no decrease in proportion as the CC size increases (in fact there is a slight rise in proportion), but for comparatives, there is a decrease from 3-5 (13.58%) to 7-9 (5.15%). These

---

1 https://www.tripadvisor.com/LocationPhotoDirectLink-g297996-d3695414-i296727865-Three_Tombs-Erzurum.html, accessed August 22, 2021; https://www.gutenberg.org/files/54024/54024-h/54024-h.htm, accessed August 22, 2021.

2 https://archive.org/about/

differences, however, can only be considered trends, as Welch's t-tests did not reach significance (comp.: $t(3.99) = -2.03, p = 0.1$; sup.: $t(3.97) = 2.03, p = 0.1$).

Overall, the corpus results suggest that the meanings of the comparative and the superlative morphemes are sensitive to the cardinality of the CC used for their interpretation. In particular, definite comparatives display a preference for CCs of cardinality 2, while definite superlatives prefer cardinalities higher than 2. However, uses of the comparative at higher cardinalities are attested at a non-trivial rate in the corpus data, tapering off in a gradient fashion.

The corpus evidence we have reviewed in this section is only suggestive; given only this, it remains open whether these attested uses correspond to actual semantic and/or pragmatic constraints linked to definite comparatives, or whether we should assume them to be negligible noise in our data. The next section articulates two different answers to this question, corresponding to two possible theories of the meaning of definite comparatives. We test these theories experimentally through the study reported in §4.

3   The two theories

We now introduce two possible theories of the comparative morpheme *-er* as it occurs in definite comparatives. The first theory, formalized in (2), treats *-er* as a relation between a comparison class of individuals C, a gradable adjective A, and an individual x, the target of comparison. Given these three parameters, the meaning in (2) returns 'True' if and only if there is some individual x' in the CC whose maximal degree of A-ness is below x's degree of A-ness. Putting it differently, the target of comparison x must instantiate the adjectival property to a higher degree than the standard of comparison x', and the standard of comparison argument is existentially quantified over.

(2) $[\![\text{-}er]\!] = \lambda C_{\langle e,t \rangle} \lambda A_{\langle d,\langle e,t \rangle \rangle} \lambda x_e : x \in C \wedge |C| = 2.$ (Theory 2I)
$\exists x' \in C : max\{d|A(d)(x)\} > max\{d|A(d)(x')\}$

This meaning can be seen as a version of a 'phrasal' comparative operator à la Heim (1985) or Kennedy (1997)—taking as arguments a gradable predicate, and two individuals, the target and the standard—except that here, the standard argument is existentially quantified over, and both the target and the standard are presumed to be part of a given comparison class C. In the lexical entry above, the CC is treated as an argument of the comparative, as with superlatives. This choice is mainly for convenience of exposition, but empirical support for it comes from the fact that definite comparatives license an *of*-phrase that can instantiate this argument (as in *the bigger of the two*).

Importantly, the meaning in (2) also encodes two presuppositions: the first one states that the target of comparison x must also be a member of the comparison class C. The second presupposition, highlighted in blue, imposes the requirement that the cardinality of individuals in the CC be 2. Because this presupposition will play an important role in our interpretation of the experimental results, we refer to this theory as the **2-Individuals (2I) theory**.

The second theory we consider here diverges from the 2I theory in that the relevant cardinality presupposition is stated over degrees, not individuals. We refer to this second theory as the **2-Degrees (2D) theory**. We assume that the set of degrees instantiated by a given comparison class depends on a contextually-given granularity. We are inspired by Krifka's granularity-based theory of imprecision (Krifka 2007, 2009), and its developments and applications (Sassoon & Zevakhina 2012; Solt 2014, 2015; Solt, Cummins & Palmović 2017; Gyarmathy 2017), in which scalar meanings map individuals to scales—in this case an adjectival scale—that can be more or less coarse-grained, making more or fewer distinctions. We take γ_A to be a set of degrees of A-ness or, equivalently, a granularity of the A-scale; a subset of the A-scale. The degrees must be evenly spaced, such that for some size δ, for any two members of γ_A, the additive distance between them is some integer mutliple of δ. Two distinct granularities may have the same size; for example, if the degrees of A are integers, then $\{...,5,10,15,...\}$ and $\{...,6,11,16,...\}$ would be two distinct granularities of the A scale. A granularity is therefore characterized by both a size and and offset from the zero element of the relevant scale.[3]

Let $A_\gamma(x)$ 'the measure of x along A with granularity γ' denote the greatest element d of γ_A such that $A(d)(x)$:

$$A_\gamma(x) = \max\{d \in \gamma_A \,|\, A(d)(x)\}$$

To restrict the granularity scale to a particular comparison class under consideration, we define $\gamma_{A,C}$ as the set of values that A_γ assigns to any member of C (the image of A_γ under C).

$$\gamma_{A,C} = \{d \,|\, \exists x \in C : A_\gamma(x) = d\}$$

With this, the meaning of the comparative under the 2D Theory is given in (3).

(3) $[\![\text{-}er]\!]^\gamma = \lambda C_{\langle e,t\rangle} \lambda A_{\langle d,\langle e,t\rangle\rangle} \lambda x_e : x \in C \wedge |\gamma_{A,C}| = 2 \,.$ (**Theory 2D**)
     $\exists d' \in \gamma_{A,C} : A_\gamma(x) > d'$

As with the 2I theory, the meaning in (3) encodes two different presuppositions. The first one is shared with the 2I theory and ensures that the target of comparison x is in the comparison class C. The second presupposition, highlighted in red, imposes

---

3 For a characterization of the notion of zero element, see Elizabeth Coppock's contribution to this volume.
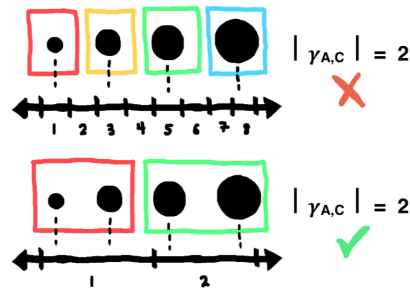
Figure 4   Two granularities representing different levels of coarseness. The upper granularity maps each of the individuals in the comparison class to different degrees, failing to satisfy the 2-degree cardinality presupposition of the 2D theory. The granularity in the bottom maps the same comparison class to only 2 degrees thus satisfying the cardinality presupposition of the 2D theory.

a cardinality requirement on the number of degrees represented in the comparison class ($\gamma_{A,C}$). The latter presupposition will be satisfied as long as the individuals in the CC, collectively, are mapped to exactly two distinct degrees.

To illustrate the role that granularities play in the presupposition satisfaction conditions of modified NPs like *bigger circle*, consider the two granularities in Figure 4, which differ in their coarseness: for the fixed range shown, the top granularity makes a 8-way distinction (i.e., it distinguishes between 8 degrees, yielding a 4-way distinction among individuals in the comparison class), while the bottom one makes a 2-way distinction (i.e., it distinguishes between 2 degrees, creating a 2-way distinction among individuals in the comparison class). Now let us consider what sets of degrees would be obtained if we were to map the comparison class represented in the figure, composed of four circles of different size arranged horizontally from smallest to biggest. Mapping the individuals in this comparison class onto the top granularity scale results in four distinct degrees, represented by the different colored boxes in the figure. Since the cardinality of the degree set is higher than two, the presupposition of the comparative fails to be satisfied, and the description is as a result undefined.

Matters are different when the bottom granularity in Figure 4 is considered. In this case, the same comparison class is now mapped to only two distinct degrees, as more than one individual is mapped to the same degree. Therefore, the 2-degree presupposition is satisfied in this context, even though the comparison class consists of more than two individuals. So far, so good, as far as the comparative is concerned. However, under this granularity the uniqueness presupposition of the definite article is violated, so *the bigger circle* as a whole is still not predicted to be felicitous under
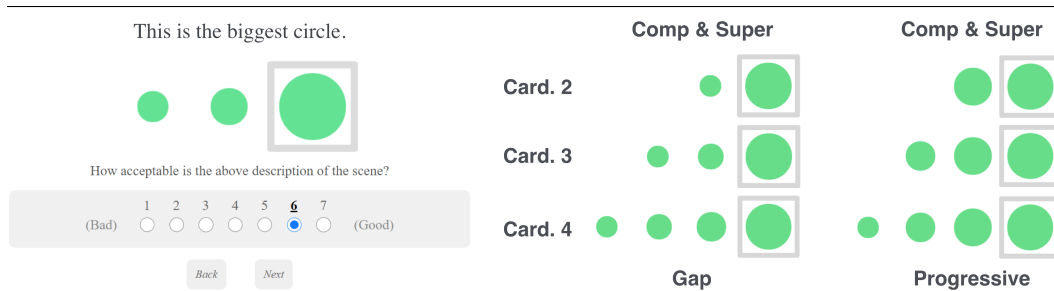
Figure 5 Left: Trial example representative of the experimental task. Right: Item example representing the 12 conditions tested in the experiment. The sizes were as follows. Gap/2: 10,30; Gap/3: 10,11,30; Gap/4: 10,11,30. Prog 2: 25,30; Prog 3: 20,25,30; Prog 4: 15,20,25,30.

the 2D theory.[4] In the experimental materials discussed in the next section, we will see examples where the comparison class consists of more than two individuals, the degree-cardinality presupposition is satisfied (due to a coarse granularity), and the uniqueness presupposition of the definite article is satisfied as well. In these contexts, the 2D Theory predicts a definite comparative to be felicitous, while the 2I Theory does not.

In some cases, whether or not the 2D Theory predicts a description to be felicitous relative to a given scene depends on the choice of granularity. It could happen that given one choice of granularity, the presupposition is met, but given another, it fails. In these cases, we predict that the degree of acceptability is related to the proportion of granularities that fail, so our predictions about acceptability will be gradient. We refer to a situation in which reference failure occurs at every granularity as *global reference failure*.

4 Experiment

We now present a study whose goal is to adjudicate between the 2I Theory and the 2D Theory. We used an acceptability judgement task where participants were asked to rate, on a 1-7 scale, statements of the form *This is the biggest circle*, against visual scenes displaying a comparison class of geometric shapes. The individuals in the comparison class differed in size but were all of the same shape and color. The intended referent was highlighted with a gray frame that disappeared if the

---

4 Note that the indefinite description *a bigger circle*, which does not presuppose uniqueness, is predicted to refer successfully under the 2D Theory when the bottom granularity is chosen, as long as the intended referent is either of the two circles in the green box.

Adjective	Noun
{big, small}	{pentagon, triangle, circle, diamond, heart}
{tall, short}	{cone, cylinder}

Table 1   Pairs of nouns and adjectives tested in the experiment.

participant hovered over it with the mouse. This was done in order to ensure that participants could easily compare the dimensions of all the shapes displayed in the scene. The size of the intended referent was kept constant across conditions (See right panel of Figure 5).

Three factors were manipulated. The first was the CARDINALITY of the set of individuals forming the comparison class. Three cardinalities (2 vs. 3 vs. 4) were tested. Second, the RELATIVE SIZE of the individuals in the scene was manipulated. This manipulation determined the degree of coarseness of the granularities licensed by the context (see §4.4). In the *progressive* condition, the size differential was kept constant across all the shapes in the comparison class, i.e., the size of the shapes (in)creased by the same factor. In the *gap* condition, there was a bigger size differential between the intended referent of the description and the second-ranked object in the display, while the remaining objects differed in size by the same factor. The six displays resulting from crossing the factors CARDINALITY and RELATIVE SIZE were paired with two different types of descriptions differing only in the ADJECTIVE TYPE, i.e., *comparative* vs. *superlative*, of the form 'this is the Adj-er/est N' (e.g. *this is the bigger/est circle*). The resulting twelve conditions are illustrated in the right panel of Figure 5.

4.1   Materials

Eighteen items were created. The adjective and noun pairs tested can be found in Table 1. All geometric shapes were automatically generated in order to ensure that the size manipulations were comparable across items. For instance, visual items consisting of squares, circles or diamonds were generated by circumscribing the shape in a circle whose size was controlled by scaling its radius. An example item can be found in Figure 5.

12 filler trials were included. Filler trials were divided into acceptable (6 trials) and unacceptable (6 trials) fillers. Most unacceptable fillers involved global reference failure. For instance, in the first class of unacceptable fillers, the shape mentioned in the description (e.g., *This is the red/bigger circle*) did not match the highlighted shape in the visual display (e.g., a square); a second type of ungrammatical fillers failed to

meet the uniqueness presupposition of the definite article in the description (e.g., the scene contained two purple circles, one of which was highlighted, and the description was *This is the purple circle*). The third class of unacceptable fillers involved cases where the adjective in the description was unnecessary to unambiguously identify the referent, thus incurring a violation of the quantity and manner Gricean maxims (e.g., the scene consisted of a blue square and a green triangle and the description was *This is the green triangle*).

For all grammatical filler trials, care was taken to construct the display in such a way that the description referred successfully. Acceptable filler trials were of three types: the first class involved descriptions that included a postnominal PP modifier, instead of a prenominal adjective (e.g., *This is the cylinder next to the cone*); a second class of grammatical fillers included negation in the description (e.g., the description read *this is not the biggest circle* and the highlighted shape was the smallest circle in the display); the last class of acceptable fillers included a color adjective in the description (e.g., the description read *This is the purple rectangle* and the scene consisted of all red rectangles with the exception of a purple rectangle, which was also the highlighted shape).

4.2   Procedure

Participants were instructed to read the description that was paired with each display, scan the visual scene and rate the acceptability of the statement as a description of the highlighted shape. Participants provided their rating by clicking on one of the seven points of the scale that appeared below the visual display. At that point, participants were free to change their rating or proceed to the next trial by clicking on the *next* button (See left panel of Figure 5). At the beginning of the experiment, participants had three practice trials to help them become familiar with the task.

4.3   Participants

Participants were 196 native speakers of English (mean age of 35 years old) recruited through the online platform Prolific. Data from 9 participants was discarded due to failure to accurately rate the filler items. A participant was discarded if they had a mean rating for the unacceptable fillers that was above 4 *or* a mean rating for the acceptable fillers that was below 4. The results reported in §4.6 pertain to the 187 participants whose responses met these independently established criteria.
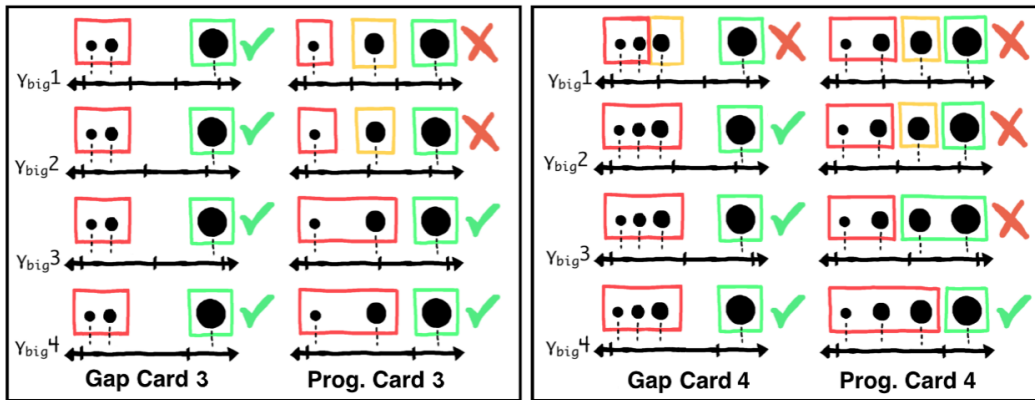
Figure 6 Reference failure patterns for the *gap* and the *progressive* conditions at cardinalities 3 (left) and 4 (right) as predicted by four representative granularities (γ1-4). Each coloured box represents a different degree.

4.4 Linking Hypothesis

We assume that the acceptability of a definite description is modulated by its **reference failure potential**, i.e., the likelihood that a given description will fail to refer in the context given the possible parametrizations of its meaning. In our experiment, there were two ways in which a description could fail to refer. The first one involves failure to satisfy the uniqueness presupposition of the definite article. The second one involves failure to satisfy the cardinality presupposition of the comparative, which can be either stated over individuals (2I theory) or over degrees (2D theory). The 2I theory predicts that the reference failure potential of a description is categorical: provided that there is no uncertainty about the comparison class, the uniqueness and the cardinality presuppositions are either violated or satisfied. On the other hand, the 2D theory predicts that the reference failure potential of a description can be gradient, since the same description can fail to refer given a scene depending on how γ is parametrized. Figure 6 illustrates how the reference failure potential of a description is modulated by the two experimental factors (i.e., RELATIVE SIZE and CARDINALITY) manipulated in the current study.

   We start with the left panel, which represents a comparison class of three different sized circles in the *gap* (left) and *progressive* conditions (right) evaluated against four different granularities (γ1-4). As seen in the figure, reference failure patterns differ across the two levels of the RELATIVE SIZE factor. Within the *gap* condition, the description *the bigger circle* is defined under the four granularities γ1-4. Despite the fact that the comparison class contains three individuals, the cardinality of the
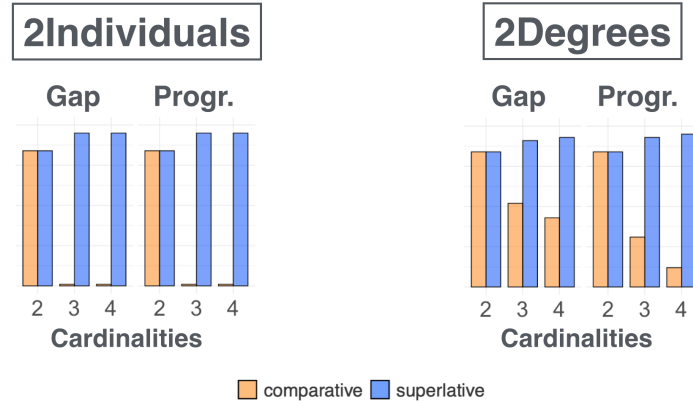
Figure 7   Qualitative predictions made by the 2I (left) and the 2D theories (right).

set of degrees represented in the comparison class is 2, since the two smaller circles are mapped to the same degree. Hence the presuppositional constraints imposed by the comparative are satisfied. The important contrast to note here is between the *gap* and the *progressive* condition. As shown in the figure, the reference failure potential given granularities $\gamma_{big}1-4$ is greater in the *progressive* condition, where two out of the four granularities under consideration result in reference failure, compared to the *gap* condition, where none of the four granularities are problematic. Therefore, we expect the acceptability of the description to be higher in the *gap* condition compared to the *progressive* condition.

Next we spell out how the cardinality of individuals in the comparison class can affect the reference failure potential of a description. The right panel of Figure 6 illustrates a comparison class of 4 individuals in the *gap* condition (left) and the *progressive* condition (right). Using the same granularities $\gamma_{big}1-4$ to evaluate the meaning of the definite description, we see that the reference failure potential is higher at cardinality 4 compared to 3 within the *gap* and the *Progressive* levels respectively. In particular, focusing on the *gap* conditions, $\gamma_{big}1$ causes the description to fail at cardinalality 4, while the description successfully refers at cardinality 3 when this granularity is adopted. Similarly, in the *progressive* conditions, the description fails in two out of the four granularities when the comparison class contains three individuals, whereas it fails in all the granularities but one when the comparison class is formed by four circles. To summarize, in our experiment there is higher reference failure potential in the *progressive* condition compared to the *gap* condition, and for higher cardinalities compared to lower ones.

	Gap 2	Gap 3	Gap 4	Prog 2	Prog 3	Prog 4
Sizes	10,30	10,11,30	10,11,12,30	25,30	20,25,30	15,20,25,30
Total γ-sizes	20	20	20	5	10	15
Failing γ-sizes	0	1	2	0	5	10
Proportion failing	0%	5%	10%	0%	50%	66.67%

Table 2   Proportions of granularity sizes that give rise to a degree set cardinality larger than two in each experimental condition.

4.5   Predictions

The qualitative predictions for the experiment are illustrated in Figure 7. We start with the predictions for the comparative conditions made by each of the two theories considered. The 2I theory claims that definite comparative descriptions should only be defined when the cardinality of the comparison class is 2, and should fail to refer otherwise. In the current experiment this should translate into high acceptability ratings in the cardinality 2 condition compared to cardinalities 3 and 4 (left panel Figure 7). The 2D theory (right panel Figure 7) predicts that the acceptability of definite comparatives should be negatively impacted the higher the reference failure potential of the description is in a given context. The reference failure potential of the description is estimated as a proportion of the granularity sizes between the smallest and largest differences between two objects in the scene that is guaranteed to result in reference failure, regardless of the offset; see Table 2.

In the *gap* condition, the reference failure potential for cardinality 3 is less severe than for cardinality 4: 5% vs. 10%. Hence a difference between cardinalities 3-4 is expected to be found within the *gap* condition, if a difference that small can be detected. The corresponding drop in the *progressive* condition is much larger: 50% vs. 66.67%. Hence we expect to observe an interaction between RELATIVE SIZE and CARDINALITY, such that the acceptability of definite comparatives should be lower at cardinality 4 compared to 3, with this drop being more accentuated in the *progressive* compared to the *gap* condition.

We now turn to the predictions pertaining to the superlative conditions. We are assuming a meaning for the superlative morpheme that does not encode any cardinality presuppositions, so the reference failure potential for superlatives is null in all the conditions tested. A consequence is that at cardinality 2, comparatives and superlatives are at a par with respect to their reference failure potential. From this perspective, we have no reason to expect a contrast between them or any penalty for either at this cardinality.
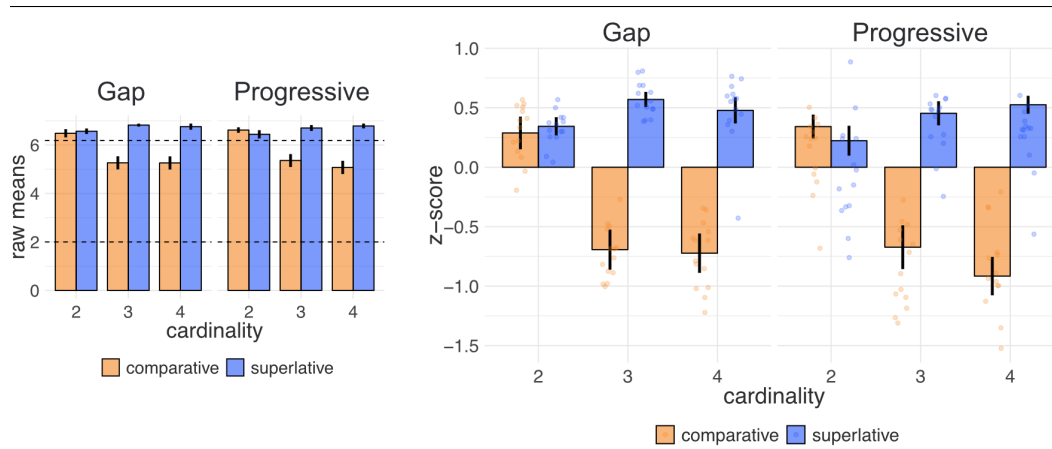
Figure 8 Experimental results. Left: mean acceptability judgements. Error bars represent 95% bootstrapped confidence intervals. Dashed lines are mean acceptability judgements corresponding to the grammatical fillers (upper dashed line) which did not involve either local or global reference failure and unacceptable fillers (lower dashed line), which always involved global reference failure. Right: acceptability judgements z-scored by subject. Error bars are 95% bootstrapped confidence intervals. Floating dots represent mean ratings across participants for each of the items tested.

Other factors might be expected to play a role in acceptability judgments above and beyond reference failure potential, however. For instance, as a result of pragmatic competition, the acceptability of the superlative might be boosted in conditions where the acceptability of its comparative counterpart decreases. If utterance competition plays a role in our results, both theories predict higher ratings for the superlative at cardinalities 3 and 4, compared to cardinality 2 in both the *gap* and the *progressive* conditions. Furthermore, if the acceptability of the superlative increases as the acceptability of the comparative decreases, then the 2I theory predicts comparable superlative ratings at cardinalties 3 and 4 (Cf. left panel Figure 7), whereas the 2D theory predicts that the superlative ratings should also increase from cardinality 3 to cardinality 4, as the comparative decreases in acceptability (Cf. right panel Figure 7).

4.6 Results

Figure 8 contains the experimental results. The raw means for each of the conditions tested were overall very high (> 5, see left panel of Figure 8), suggesting that all conditions were perceived to be acceptable by our participants. Results were *z*-scored by participant in order to normalize the ratings across subjects (right panel of Figure 8). Data pertaining to the *gap* and *progressive* conditions respectively was subsetted by cardinality (cardinalities 2-3 and cardinalities 3-4) and submitted to a series of Bayesian Mixed Effect models, where *z*-scored ratings were predicted from the factors ADJECTIVE TYPE and CARDINALITY and its interaction. All models contained the maximal random intercept and slopes justified by the experimental design. The brms R package was used to perform all the analyses. Simple effects were computed using the function hypothesis also included in the package brms.

A significant ADJECTIVE TYPE $\times$ CARDINALITY interaction for cardinalities 2-3 within the *gap* condition was detected ($\beta = 1.21$, CI $= [0.97, 1.46]$), such that there was a significant increase in the acceptability of superlative descriptions in cardinality 3 compared to cardinality 2 ($\beta = 0.25$, CI $= [0.09, 0.43]$), and a significant decrease in acceptability in the comparative conditions going from cardinality 2 to 3 ($\beta = -0.98$, CI $= [-1.19, -0.78]$). However, the equivalent 2-way interaction did not reach significance at cardinalities 3-4 ($\beta = -0.01$, CI $= [-0.28, 0.25]$).

In the *progressive* condition, there was a 2-way ADJECTIVE TYPE $\times$ CARDINALITY interaction in cardinalities 2-3 ($\beta = 1.22$, CI $= [0.96, 1.47]$). As in the *gap* condition, the interaction detected at lower cardinalities was driven by both the comparative and the superlative conditions, with superlatives increasing in acceptability in cardinality 3 compared to 2 ($\beta = 0.25$, CI $= [0.09, 0.43]$), while ratings were significantly lower for comparatives in the same comparison ($\beta = -0.97$, CI $= [-1.17, -0.74]$). However, unlike what was observed in the *gap* condition, a significant ADJECTIVE TYPE $\times$ CARDINALITY interaction was detected for the higher cardinalities ($\beta = 0.35$, CI $= [0.08, 0.64]$). This interaction was driven by significantly lower ratings for the comparative in cardinality 4 compared to cardinality 3 ($\beta = -0.29$, CI $= [-0.54, -0.06]$). No effect was detected for the superlative conditions ($\beta = 0.06$, CI $= [-0.1, 0.21]$).

Finally, while there was a clear difference between the comparative and the superlative conditions within cardinalities 3 and 4 respectively, such that the superlative conditions were overall rated higher, no main effect of RELATIVE SIZE was found for the comparative conditions at cardinalities 3-4 ($\beta = -0.03$, CI $= [-0.22, 0.15]$). Furthermore, no significant difference was detected within cardinality 2 for the comparative vs. superlative comparison in either the *gap* ($\beta = 0.04$, CI $= [-0.11, 0.20]$) or the *progressive* ($\beta = -0.13$, CI $= [-0.33, 0.05]$) conditions.
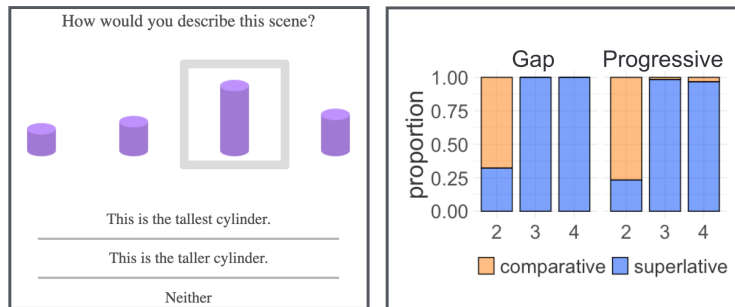
Figure 9   Left: Trial example included in the production experiment. Right: Proportions of definite comparative *vs*. definite superlative responses.

4.7   Discussion

We start with the implications of the current results for the two theories of definite comparatives under consideration. Our results are overall supportive of the 2D theory of definite comparatives. Parallel to what was observed in the corpus results presented in §2, our experimental results show that definite comparative descriptions decline in acceptability when judged against comparison classes of more than two individuals; definite comparatives with cardinalities > 2 were judged to be less acceptable than ones with cardinality 2, and less acceptable than our fully grammatical fillers (mean rating ≈ 6). However, the absolute mean ratings that comparatives received at cardinalities 3 and 4 remain significantly higher (> 5) than the mean ratings obtained for the unacceptable fillers (≈ 2), which were undefined due to global presupposition failure. These findings suggest that participants did not judge comparatives with comparison classes of higher cardinalities to involve global reference failure, contrary to what is predicted by the 2I theory. This pattern of results can be easily accommodated by the 2D theory, which predicted a non-categorical decline in acceptability ratings as the comparison class grew beyond size 2.

However, the crucial data point supporting the 2D theory is the significant decrease in acceptability ratings found for comparatives within the *progressive* condition at cardinality 3 compared to cardinality 4. This decline is to be expected under the 2D theory due to the higher reference failure potential predicted in the cardinality 4 condition, but cannot be explained by the 2I theory.

The 2D theory also predicted a 3-4 decrease in the *gap* condition, albeit a smaller one, and no such decrease was detected. A possible reason for this is that the acceptability judgment task used in the present experiment does not have enough sensitivity to detect smaller rates of reference failure potential than those involved in the progressive condition.

We now turn to the results pertaining to the superlative conditions. As predicted

under the view that comparatives and superlatives are competing expressions, our participants rated superlative descriptions higher when the cardinality of the comparison class was higher than 2, mirroring the decrease in acceptability for comparatives. No other signs of pragmatic competition were observed in the experiment, probably due to the fact that the difference in acceptability of the comparative descriptions in the *progressive* conditions at cardinality 3-4 was not big enough to be reflected in the acceptability patterns of the superlative conditions. Note that this pattern of results could not be completely explained if we assumed that the meaning of the superlative encodes a cardinality presupposition that requires the comparison class to have a cardinality of degrees or individuals greater than 2. If this was the case, the acceptability of the superlative descriptions should be on a par with ungrammatical fillers. In fact, comparative and superlative descriptions received fairly high comparable ratings at cardinality 2. In this respect, our experimental results differ from the corpus findings, which did show a strong preference for the comparative at cardinality 2 compared to the superlative (cf. Figure 3).

A strong preference for the comparative over the superlative at cardinality 2 was also found in a pilot production study ($n = 30$) consisting of a forced choice task. Participants saw the same visual displays tested in the comprehension experiment, so the data was comparable to our comprehension data. Each scene was paired with three possible descriptions (*This is the tall-er/est cylinder* or *neither*, see left panel of Figure 9). Participants were instructed to choose the utterance that best described the object highlighted by the gray box. Our findings paralleled the corpus results such that at cardinality 2 participants overwhelmingly preferred the comparative description, whereas at higher cardinalities there was a clear preference for the superlative, although the comparative was on occasion chosen as well. So there is an asymmetry between production and comprehension here: Production shows a clear preference for the comparative in cardinality 2, but superlatives and comparatives receive comparable acceptability ratings.

The production preferences indexed by the corpus and pilot results reported here receive a straightforward explanation if we assume that speakers follow the principle of MAXIMIZE PRESUPPOSITION (Heim 1991), which states that given two utterances with equivalent assertive content, speakers will prefer the utterance with the richer presuppositional contribution. From this perspective, however, it is surprising that comprehenders gave such high ratings for the superlative variant with comparison classes of size 2. After all, it is well-known that violations of Maximize Presupposition affect the acceptability of the utterance (e.g. #*a sun is shining*). One way of squaring Maximize Presupposition with the high acceptability ratings for superlatives in the cardinality-2 conditions may have to do with the strength of the comparative as a competitor. Although the ratings for comparatives with cardinality 2 are high, they are not at ceiling; they are on par with the superlatives in the

cardinality-2 condition, and significantly below superlatives at higher cardinalities, which were very near the top of the scale. It may be that the comparative is a somewhat weak competitor to the superlative, itself being slightly marked. We can only speculate about why the comparative received less than perfect ratings in the cardinality 2 conditions, but one possibility is that it competes with the positive form (e.g. *big*), which is shorter. Further research will need to determine whether Maximize Presupposition effects are modulated by the strength of competitors, and whether the viability of the positive form plays a role here.

5   Conclusion

What matters for definite comparatives in English is not the size of the comparison class, but the number of degrees it embodies. Definite comparatives require that the comparison class embodies exactly two degrees on the relevant scale, relative to a given granularity. Comprehenders flexibly search for a granularity that lets the semantic requirements of the definite comparative be met. The more likely this process is to result in failure given a scene, the less acceptable a definite comparative is in relation to that scene. This paper adds to the growing evidence that language comprehension involves reasoning under uncertainty about granularities, and contributes comparison as a new domain in which this concept plays a role.

References

Coppock, Elizabeth. 2019. Definite comparative descriptions: the more superlative-like comparative construction. Talk presented at SuSurrus, U. Mass Amherst.

Gyarmathy, Zsófia. 2017. A generalised framework for modelling granularity. *Journal of Semantics* 34(3). 483–506. doi:10.1093/jos/ffx004.

Heim, Irene. 1985. Notes on comparatives and related matters. Ms., Semantics Archive. University of Texas, Austin.

Heim, Irene. 1991. Artikel und Definitheit. In Arnim von Stechow & Dieter Wunderlich (eds.), *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung*, 487–535. Berlin: Mouton de Gruyter.

Kennedy, Christopher. 1997. *Projecting the adjective: the syntax and semantics of gradability and comparison*: UC Santa Cruz PhD dissertation.

Krifka, Manfred. 2007. Approximate interpretations of number words: A case of strategic communication. In Gerlof Bouma, Irene Krämer & Joost Zwarts (eds.), *Cognitive Foundations of Interpretation*, 111–126. Amsterdam: Humboldt-Universität zu Berlin, Philosophische Fakultät II.

Krifka, Manfred. 2009. Approximate interpretations of number words: a case of

strategic communication. In Erhard Hinrichs & John Nerbonne (eds.), *Theory and evidence in semantics*, 109–132. Stanford: CSLI Publications.

Sassoon, Galit & Natalia Zevakhina. 2012. Granularity shifting: Experimental evidence from degree modifiers. In *Semantics and Linguistic Theory (SALT)*, vol. 22, 226–246.

Solt, Stephanie. 2014. An alternative theory of imprecision. In *Semantics and Linguistic Theory (SALT)*, vol. 24, 514–533.

Solt, Stephanie. 2015. Granularity and approximating number pairs. In *Israel Association for Theoretical Linguistics (IATL)*, vol. 31, 83–100.

Solt, Stephanie, Chris Cummins & Marijan Palmović. 2017. The preference for approximation. *International Review of Pragmatics* 9. 248–268. doi:10.1163/18773109-00901010.

Helena Aparicio
Department of Linguistics
Cornell University
203 Morrill Hall/159 Central Ave.,
Ithaca, NY 14853-4701
haparicio@cornell.edu

Roger Levy
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, MA 02139
rplevy@mit.edu

Curtis Chen
Department of Brain and
Cognitive Sciences
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, MA 02139
curtisc@mit.edu

Elizabeth Coppock
Department of Linguistics
Boston University
621 Commonwealth Avenue
Boston, MA 02215
ecoppock@bu.edu