

BIG DIFFERENCES: THE STANDARD FOR ‘BIG’ AS USED BY ADULTS AND CHILDREN

ELBERT J. BOOIJ¹

GALIT WEIDMAN SASOON²

¹*ILLC Amsterdam*

²*Bar Ilan University*

Abstract

Dutch-speaking adults and children were tested to find out about their criteria to apply the relative adjective *groot* (En.: *big, large*). The hypothesis that *leaps* in the distribution trigger cutoff points between the items that will and will not be called *groot* was confirmed for the adults, but not for the children.

1 Introduction

The idea that language-users deploy a *comparison-class* when determining the extension of gradable adjectives like *tall* and *big*, in those combinations in which they occur, enjoys broad consensus (Klein, 1980; Bierwisch, 1989; Ludlow, 1989; Kennedy and McNally, 2005; Sassoon and Toledo, 2011). What this means is that the decision which entities are called ‘big’ or ‘tall’ in a context depends on the class of entities to which they belong, and how they compare to entities in this class. These so-called *relative adjectives* (Rotstein and Winter, 2004; Yoon, 1996), contrary to their absolute counterparts—like *clean* and *full*—lack any context-free standard of comparison to determine whether they should be applied to a certain item or not. The great expressive power of relative adjectives is possible exactly because of this flexibility to adjust to an enormous variety of contexts of use. The word ‘tall’ can make a sensible contribution in discourse about mountains and candles alike.

But to be functional in this way, it seems that flexibility must always be related to something that remains constant. Every time such a term is used it divides the domain of discourse—candles on one occasion, mountains on another—into those items that may be called, e.g., *tall* and the rest of them. In order to make communication possible a hearer must be able to draw conclusions. It seems that there must be some sort of *standard* that, given a comparison-class of a certain category to which the object belongs, determines what counts as *tall*. In Kennedy’s (2007) words,

the objects so called *stand out* among the other members of the contextually determined class. What this standing out in fact consists of is an issue of ongoing discussion and investigation. For a quick overview of the relatively simple proposals: the standard could be a contextually fixed *degree* of, say, height (apple-trees above 3.5m are *tall*) or a fixed *percentage* of the comparison class (the tallest 40% are *tall*) (Schmidt et al., 2009; Barner and Snedeker, 2008; Solt and Gotzner, 2012). Only in the former case it makes sense to speak of a *cutoff point*, i.e. the degree beyond which items will be called *tall*. The standard could furthermore be a function of the *distribution* (the trees beyond the mean plus 1.5 times the standard deviation are *tall*), in which case there is also a cutoff point, but this time variable (Schmidt et al., 2009; cf. also Solt, 2011, who uses the median and the median absolute deviation). Finally it has been claimed that the standard is based on similarity-relations (McNally, 2011).

In the present study we want to present and test a new hypothesis, according to which it is *leaps in the distribution* of the relevant variable that trigger the cutoff point for the use of the adjective. We shall dubb this the ‘Leap Hypothesis’. The theoretical background of the Leap Hypothesis is as follows. Although it is in the power of human language to make a very large number of distinctions, we do not nearly make all of them in practice. We make distinctions where and when we deem them *worth* making; when communicating them is in some way or other profitable. But often one distinction is more difficult to make than the other, and in such cases the benefits of being informed may be tempered by the costs of obtaining the information. There is likely to be a trade-off between *worth* and *ease* of a certain discrimination: one distinction may be in itself more valuable than the other, but so hard to make with any reliability, that going for the slightly less rewarding but safer and cheaper distinction is the thing to do (Sperber and Wilson, 1995). All else being equal, therefore, discriminations that are *easy to make* are likely to be privileged.

There are good reasons to expect that among the most economic cutoff points in this regard are those that coincide with a leap in the distribution. The suggestion is an appealing one intuitively: in a wood where the trees have a variety of heights, only leaving out the range between 6 and 10 metres, the domain, as it were, splits naturally into two categories that can be easily told apart. This intuitive picture finds its theoretical underpinning by the framework of *semi-orders* (Gaifman, 2010; van Rooij, 2010, 2011). A semi-order is an ordering relation between items that is less strong than a weak order, which is the sort of order that obtains when the whole domain is thought of as partitioned into degrees. An example of a weak order is the relation *taller than*, prevailing in a group of people based on their exact lengths. Such an ordering straightforwardly divides its domain into equivalence classes, viz. the groups of people of exactly the same (degree of) length. An example of a semi-order is *much taller than*. The pivotal difference between weak orders and semi-orders is that the \sim -relation¹ in the latter is *not* transitive. It may be that John is not much taller than Susan, who is not much taller than Mary, and yet John *is* much taller than Mary. There is an *indistinguishability threshold*, below which no ‘real’ inequality holds. Thus for a semi-order > we may have:

$$x \sim y \text{ and } y \sim z, \text{ yet } x > z$$

A piling up of such insignificant differences may add up to a difference that is distinguishable after all.

¹ $x \sim y$ iff neither $x > y$ nor $y > x$

It is this phenomenon that is responsible for the *Sorites paradox*²; in this context Gaifman and van Rooij (*ibid.*) use semi-orders. Unlike weak orders, semi-orders do not by themselves give rise to a sound categorization. For a set of items of which every one is indistinguishable from the next in size—a Sorites series—it is hard to find a division into subsets that in any way can be called natural. This is where leaps come in. Being larger than the indistinguishability-threshold of the semi-order, a leap in the distribution of sizes does provide a natural dividing line between categories.

But why, then, should semi-orders and leaps be preferable to degrees as a way of framing distinctions? Apart from the fact that every real-worldly system of perception must have an ultimate indistinguishability-threshold, the threshold that obtains at some given occasion is also a matter of *effort*. From a distance I can easily see that some house is high compared to the barn next to it, but only after looking very carefully I can see that it is also higher than the tree on the lawn in front. Notice that for none of these observations it is necessary to know *exactly* how tall any of the objects is. Nor is there a need to compute degree-based features like mean and standard-deviation, that would put heavy demands on memory and speed of calculation. Making use of semi-orders there is a simple and economic way of classifying available: choose the *coarsest* semi-order at your disposal and switch to a finer one if the former does not yield a standard.

The idea that standard selection for gradable adjectives proceeds in this way aligns very well with the fact that non-parametric statistical criteria must capture human judgements regarding the standard across a wide array of distributions (Schmidt et al., 2009), and in many cases (*sharp, happy*) averaging is not even meaningful. Semi-orders—with different thresholds—may be available if degrees are not. Thus leaps may be there even if there is no full-blown metric, and if so, they will provide cutoff points that at least have the merit of being easily determinable and reducing uncertainty, since most of the items will be conspicuously on one side or the other. The present paper is the result of an attempt to put especially this hypothesis to test.

2 Experiments

Two groups of Dutch-speaking subjects were tested: one group of adults and one of primary school children, as described in 2.1.1 and 2.2.1 respectively.

2.1 Experiment 1

2.1.1 Participants

A group of 28 adults, 14 men and 14 women, aged 50.9 ± 7.1 years, were tested. The majority of them were employees of a secondary school in Amsterdam, all but four of them native speakers, the others fluent Dutch speakers as well. They had received 14.1 ± 2.7 years of formal education on average.

2.1.2 Stimuli

The adjective to be tested was the Dutch word *groot* (En.: *big, large*), as applied to *pencils* of different size. The decision to test *groot* rather than equivalents of *tall* or *long* was based on the

²E.g., a haystack minus one straw is still a haystack. But then, repeating the argument many times, a pile of ten straws still is!

expectation that the children are perhaps more comfortable with a very general concept like *big* (cf. Barner and Snedeker’s (2008) finding that young children sometimes switch from *short* to *small*). Although in spontaneous use adults might have a preference for *lang* (En.: *long, tall*) when referring to pencils (some of the adult test subjects commented on this), *groot* is definitely not infelicitous.

In order to allow test subjects to display their criteria for *groot* in different yet comparable circumstances, sets of twelve pencils were used, each of one of the four distributions displayed in Table 1. All pencils looked exactly the same, but for their length, and nothing about them was unusual in respect of what pencils normally look like. Even the different lengths could be considered familiar given that pencils get shorter after prolonged use. The distributions were all symmetrical and can be described in the following way:

D1: all different length, linearly increasing

D2: all different length, two clusters, one gap in the middle

D3: almost like D1, but with two leaps

D4: all different length, two small clusters at the extremes, one large modal cluster

All distributions had the same mean. D2 had the biggest standard deviation, and D4, the smallest, with D1 and D3 in between and almost equal.

Table 1: Distributions of the sets of pencils

Distribution	length in cm												μ	σ
	5	6	7	8	9	10	11	12	13	14	15	16		
D1	5	6	7	8	9	10	11	12	13	14	15	16	10.5	3.61
D2	5	5.5	6	6.5	7	9.3	11.7	14	14.5	15	15.5	16	10.5	4.37
D3	5	5.5	8	8	9	10	11	12	13	13	15.5	16	10.5	3.57
D4	5	6	9.4	9.7	10	10.3	10.7	11	11.3	11.6	15	16	10.5	3.09

2.1.3 Procedure

Every test subject was presented with two of the four distributions, one after another. The assignment of distributions to subjects was done by a randomized list in such a way that the second distribution was always different from the first, but otherwise random. The pencils were given as a bunch and laid on the table before the subject in a disorderly fashion. The subjects were asked to put the *big* pencils in a basket (“*Leg de grote potloden in het mandje*”). Clarification questions (“Do you mean the *biggest*?”) were answered by: “The ones of these pencils *you* say are big”. If the task had been performed, after some pause the question “Are you ready?” was asked and the result was collected. Then the second set of pencils was given, with the same procedure.

We shall use the term *transition* for the *observed* boundary between *big* and non-*big* items. To bring all the results together, two parameters were decided to be of interest:

- the *transition rank*, i.e. the number of the first *big* pencil (counting from the smallest to the biggest).
- the *transition midpoint*, i.e. the length in cm which lies just in between the length of the first *big* pencil and that of the last non-*big* one.

The latter notion can be seen as an estimate of the putative ‘real’ cutoff point with respect to the application of *big*.

2.1.4 Predictions

Some basic hypotheses regarding the standard of use of relative adjectives, as in e.g., *the big pencils*, have been mentioned in the Introduction. To evaluate four of these hypotheses as applied to the present experiments, we can formulate the following predictions:

- H1 The criterion for *big* is to be bigger than a fixed *size*. If this is the case, then the transition midpoint should be expected to be relatively invariant across different distributions.
- H2 The criterion for *big* is to be among $x\%$ of the biggest pencils, where x is a fixed *percentage*. If this is the case, then the transition rank should be expected to be relatively invariant across different distributions.
- H3 The criterion for *big* is to be of a size beyond the mean plus a times the standard deviation ($\mu + a \cdot \sigma$). If this is so, then, for distributions with constant μ , the transition midpoint should be expected to be farther from μ when σ is bigger.
- H4 The criterion for *big* is determined by leaps in the distribution. If this is true, then such leaps should enhance the likelihood of the transition (from non-*big* to *big*) to fall in that region.

With respect to H4 the following should be mentioned. Even if H4 is *false* and there is a fixed cutoff point, leaps in the distribution will show a tendency to coincide with the transition between *big* and non-*big*, simply because of the enhanced probability of the (fixed) cutoff point to be inside the leap-region. As an illustration, suppose that, in a set of the same range of length (5-16 cm), pencil No. 3 (in order of size) is 6 cm and the next pencil is 14 cm. Then, *if* there is a cutoff point governing the application of *big*, it is extremely likely that, by mere chance, it happens to be between 6 cm and 14 cm. In the overwhelming majority of tests, therefore, the transition from non-*big* to *big* will be between pencil No. 3 and 4, making the transition rank peak at 4—but this has little to do with the leap itself. Creating a big gap is just an easy way to manipulate the transition rank. Consequently, to support H4, leaps in the distribution must be shown to affect the transition *midpoint*, rather than the transition *rank*. Only then there is reason to believe that the cutoff point gets affected by the leaps. This is why, in the graphs below, it was decided to place the length of the pencils on the horizontal axis.

2.1.5 Results

All respondents were *consistent* in their response, by which is meant that no ‘*big*’ pencil was smaller than any non-‘*big*’ pencil in the same test. The results for the transition rank and the transition midpoint are shown in Fig. 1-4, leftmost figures. The dotted line connects the rank orders of the pencils with their lengths; the areas of the bars indicate the number of respondents for whom the transition from non-*big* to *big* is in the region covered by the bar.

Mann-Whitney U tests were performed on the results for the transition rank and the transition midpoint. With respect to the transition rank, except for the combinations D1-D3 and D3-D4, the differences between the results for the four distributions were highly significant ($p < 0.01$). With respect to the transition midpoint this was only the case for the combination D2-D4; the other differences were not even weakly significant. Thus neither the hypothesis that the criterion for

being called *big* is to be above a fixed size (H1), nor the hypothesis that this criterion is to be among a certain percentage of the biggest pencils (H2) are supported by the findings, even though H1 does certainly better.

Also the idea that the use of *big* is governed by a cutoff point of degree $\mu + a \cdot \sigma$ (H3) appears to be refuted by the results for the average of the transition midpoint for the different distributions: the transition midpoint is even roughly *closer* to the mean value with bigger standard deviation (see Table 2).

The difference between D1 and D3 is of special interest because the mean and standard deviation of both are almost the same, but D3 has leaps, the rightmost of which is between 13 cm and 15.5 cm. The picture for D3 (Fig. 3, left) shows a peak of respondents whose transition to *big* was in this interval. If we take the results for D1 (where there are no leaps) to be a fair estimate of the ‘background’ likelihood of the transition to be between 13 cm and 15.5 cm, then a binomial test³ gives a (weak) significance of $p < 0.054$ for the peak associated with the leap in D3. It seems, then, that the leap indeed triggers a cutoff point where it would otherwise have been less probable (H4).

2.2 Experiment 2

2.2.1 Participants

A group of 26 primary school children, aged 8.1 ± 0.3 years, 13 boys and 13 girls, were tested in the same way. They were all pupils of a Waldorf school in Amsterdam, all classmates and all but four of them native Dutch speakers, the others fluent Dutch speakers as well.

2.2.2 Stimuli, Procedure, Predictions

The stimuli and procedure of testing were identical to those used in Experiment 1. Hence, so were the predictions.

2.2.3 Results

With the exception of two cases, in all tests the children were consistent in the sense defined above. The results are shown in Fig. 1-4, rightmost figures. There was wide variation in the results for individual children, with both the transition rank and the transition midpoint showing only a weak preference for the rightmost half of the distribution. In the graphs of D3 and D4 conspicuous peaks appear just in the middle.

The children in this experiment agreed less with each other than the adults did, and, unlike the latter, in several cases showed a transition *below* the average length of the pencils. Mann-Whitney U tests could not prove any of the differences between the distributions to be significant, either with respect to the transition midpoint, or the transition rank. If, however, the results for D1 and D4 are pictured with the dimensions switched (Fig. 5), we see that the pictures with respect to the transition rank are indeed dramatically dissimilar. If we idealize the results for D1 by hypothesizing that, in the absence of leaps, the likelihood of the transition to *big* to occur is equal anywhere between pencil No. 2 and No. 11—an assumption certainly not biased against the picture of D4—

³A binomial test is one of the simplest statistical tests. Its measure of significance are the odds against finding the observed number of ‘success’-outcomes under the hypothesis of mere chance.

then a binomial test gives a significance of $p < 0.000012$ for the two peaks associated with the leaps in D4.

The fact that leaps affect the results for the transition *rank* has no bearing on the Leap Hypothesis (H4) (see 2.1.4), but it disproves the hypothesis that the criterion for being called *big* by the children is to be among a certain percentage of the biggest pencils (H2). However, the hypothesis that pencils are called *big* if they are above some fixed size (H1) is compatible with the results. The idea that being *big* is being bigger than $\mu + a \cdot \sigma$ (H3) finds no support in the results (Table 2) although it is less strongly refuted for the children than for the adults. Finally, leaps in the distribution do not seem to have impressed the children, as can be seen when comparing D1 and D3. It is striking how the children differ from the adults in their reaction to D3 (Fig. 3): they virtually *ignore* the leap in the distribution of length. Thus, unlike what was the case with the adults, the Leap Hypothesis (H4) does not provide an explanation here.

Mann-Whitney U tests were also performed to see to what extent, for similar distributions, the results for the children differed from those for the adults. Significance could be proved in all four cases, albeit for D1 ($p < 0.02$) and D2 ($p < 0.084$) in a weaker sense than for the remaining two ($p < 0.01$).

Figure 1: Distribution D1: Results for adults (left) and children (right). The dotted line depicts the lengths of the pencils. The area of a bar indicates the number of subjects for whom the transition from non-*big* to *big* occurred in that region of length.

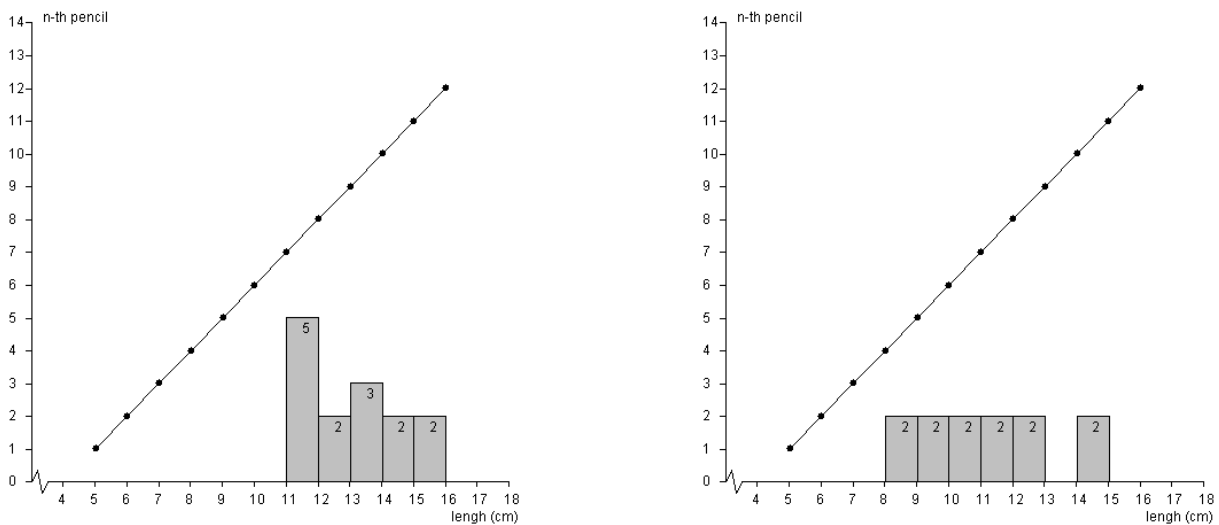


Table 2: Average value of the transition midpoint (cm)

	D1	D2	D3	D4
σ	3.61	4.37	3.57	3.09
adults (cm)	13.07	12.55	13.54	13.46
children (cm)	11.17	11.15	11.37	10.7

Figure 2: Distribution D2: Results for adults (left) and children (right). The dotted line depicts the lengths of the pencils. The area of a bar indicates the number of subjects for whom the transition from non-*big* to *big* occurred in that region of length.

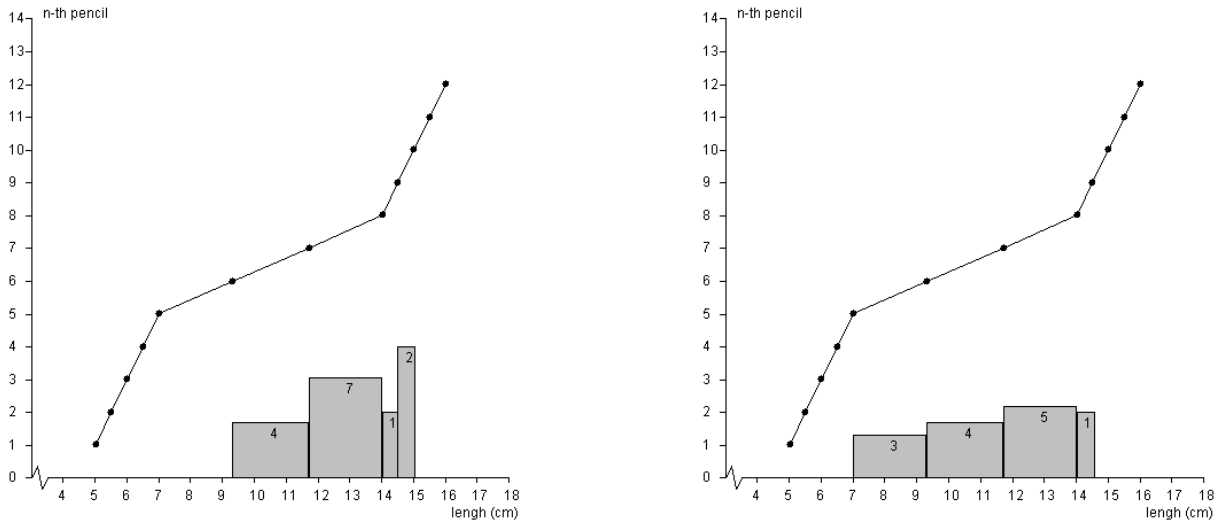


Figure 3: Distribution D3: Results for adults (left) and children (right). The dotted line depicts the lengths of the pencils. The area of a bar indicates the number of subjects for whom the transition from non-*big* to *big* occurred in that region of length.

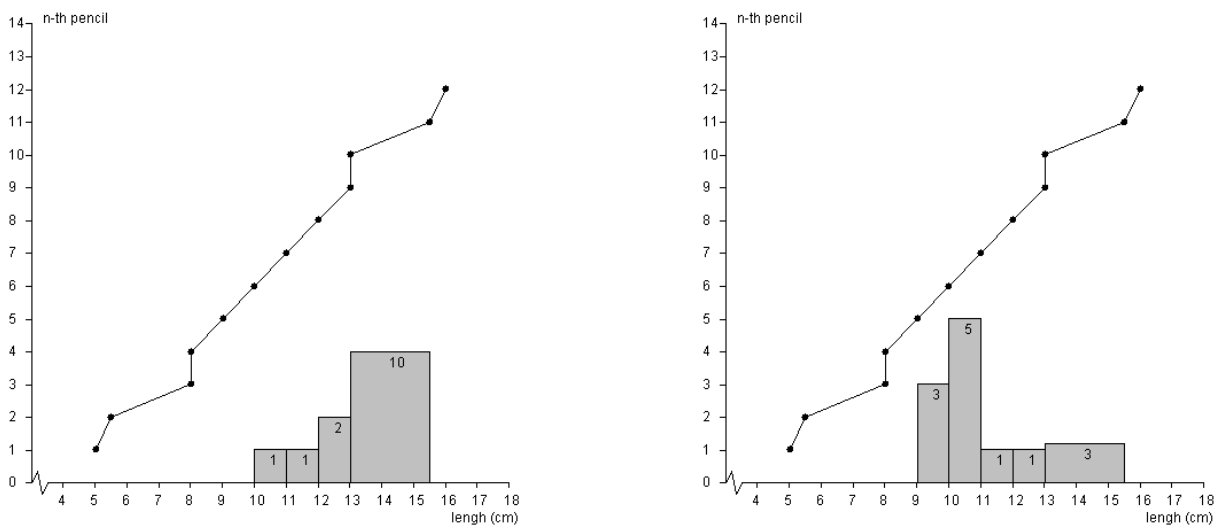


Figure 4: Distribution D4: Results for adults (left) and children (right). The dotted line depicts the lengths of the pencils. The area of a bar indicates the number of subjects for whom the transition from non-*big* to *big* occurred in that region of length.

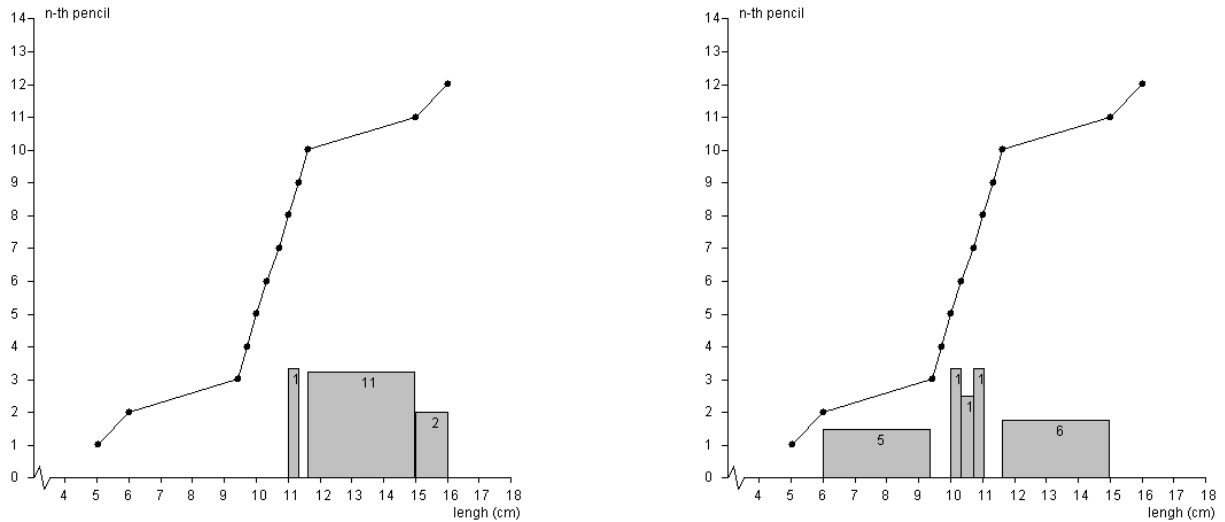
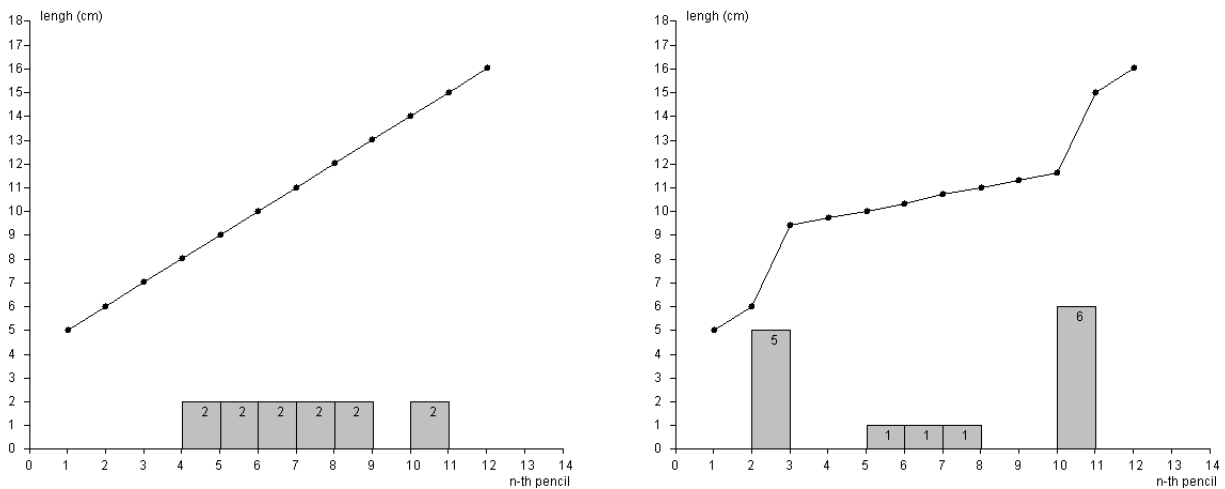


Figure 5: Children: distributions D1 (left) and D4 (right) with the number of the pencils (in order of size) on the horizontal axis. The dotted line depicts the lengths of the pencils. The area of a bar indicates the number of subjects for whom the transition from non-*big* to *big* occurred between pencils of that rank.



3 Discussion

The results obtained with the adults show that the standard for the use of the Dutch adjective *groot* (En. *big, large*), if couched in terms of a cutoff point between *big* and non-*big* of a certain length, allows of assuming this point to be relatively stable throughout different modes of stimulation. If, on the other hand, the standard is thought of as being among some fixed percentage of the biggest items, far less stability can be discerned. Therefore the assumption of a standard as a certain *degree* of length appears to be, if not entirely convincing, at least more on the right track.

Yet, if this degree of length were the whole story, this would be somewhat remarkable, since, as we are dealing with a *relative* adjective, a cutoff point that is *really* fixed—one that works throughout all contexts—is out of the question. There must be a decisive element of context-dependence involved, and there is no *a priori* reason why the range of values should be all that counts. And indeed the difference with respect to the transition midpoint between the distributions D2 and D4 demonstrates that the cutoff point in terms of length is not completely stable. One of the simpler proposals allowing the shape of the distribution to contribute to the standard for *big* is that the cutoff point should equal the mean plus *a* times the standard deviation. But such a rule finds no corroboration in the experimental results.

Leaps in the distribution, however, do seem to pull the cutoff point towards themselves, in accordance with the Leap Hypothesis. The relative stability of the cutoff point in terms of absolute length may seem at variance with this conclusion, but we suggest that both phenomena should rather be regarded as complementary. Effects of incidental leaps in a distribution are likely to occur against the background of more general tendencies to position the cutoff point, based on the same principle, but fed by experience from many preceding cases.

In the light of this finding the non-conformity of the standard for *big* to the $\mu + a \cdot \sigma$ -rule makes good sense. If a distribution has a modal cluster, comprising many items that are nearly indistinguishable, a cutoff point inside this region would cause maximal uncertainty. Consequently, where a modal cluster is present (as in D4) the use of leaps will dictate the cutoff point to be anywhere but there. But when there is a leap in the middle (as in D2), separating two clearly distinguishable clusters of items, there is every reason to place the cutoff point in the middle. Hence, in the latter case, where the standard-deviation is *bigger*, the transition midpoint will be *closer* to the mean value, as was observed with the adults. This effect is also visible in the results of Solt and Gotzner (2012:Fig. 2).

This relation between modal clusters and leaps is easily misunderstood. The Leap Hypothesis has been criticized for being unnatural in the sense that really occurring distributions are typically normal and therefore do not show leaps at all (Stephanie Solt, pers. comm.). This is a valuable point, but it must be qualified in two ways. First, there are entirely natural classes (the class of animals; the set of buildings on the skyline before us, consisting of some sky-scrapers among a large number of mansions and three-storey buildings) that are by no means normally distributed. Second, real-life classes, even if they are normally divided, are not as smooth as the textbook graph. They have many cracks and ripples, and these become larger with increasing distance from the modal cluster. Hence, for natural distributions that are approximately normal, the predictions according to the Leap Hypothesis will turn out to be in good harmony with those according to the $\mu + a \cdot \sigma$ -hypothesis.

In connection with this point the following consideration is in order. It is commonly thought (e.g., Rotstein and Winter, 2004) that the entailment-patterns of relative adjectives are such that the antonyms are not complementary:

- John is not short \Rightarrow John is tall.
- Lisa is not tall \Rightarrow Lisa is short.

An often-cited explanation for this phenomenon is that relative adjectives have *norms* for different contexts of use (Tribushinina, 2011). The positive of the adjective (or its antonym) is used for items that do not comply with the norm: they are above or below it. Now statistically a norm will typically coincide with the modal cluster of a normal distribution. Since such regions, which, as we have seen, lack leaps, are ill-fit for cutoff points, this fact provides an alternative explanation for the phenomenon of non-complementarity. Norms, after all, are an intuitive notion, but it may be questioned if they would have been that, if normal distributions had been less ubiquitous than they are.

As for the children, their standard for calling something *big* was in all likelihood not its being among a fixed percentage of biggest pencils. The option that the standard is to be beyond a fixed degree of length remains open, be it that, if so, this degree varies rather strongly among individual children. Otherwise the children behaved differently from the adults in crucial respects. Most importantly, they hardly showed any reaction to leaps in the distribution, *contra* the Leap Hypothesis. Furthermore, although they too showed an inclination to call less than half of the pencils *big*, there were, unlike with the adults, many exceptions to this trend (see e.g., Fig. 4). Table 2 shows that, consequently, the average size of the transition midpoint is far closer to the mean of the distribution (10.5 cm) with the children than with the adults. As can be seen in the results for D4 (Fig. 4), about one third of the children were willing to call everything *big* but the smallest pencils!

The peaks for the children's transition midpoint exactly in the middle of the distribution of length in D3 and D4 suggest a (minority) tendency to apply a *mid-range* standard. This phenomenon has been observed before. Tribushinina (2013) gives experimental evidence for a developmental trend whereby, by the age of about four, a mid-range standard replaces a practice where *big* and *small* only apply to the extremes of the scale. Compare also the *Innate Standard Hypothesis* as described by Barner and Snedeker (2008):

(...) upon realizing that a novel word is gradable (perhaps via syntactic cues like comparative morphology), children assume that it has an opposite and that these opposites exhaustively divide sets in two, or alternatively, result in a three-way categorization of objects. (p. 605)

The division in two is clearly the simplest assumption. Thus the mid-range standard could be a stage in children's linguistic development, following upon the stage where only the extremes are picked out, which transition is made possible by their growing ability to order items by their size (Ehri, 1976; Smith et al., 1986; Tribushinina, 2013).

We have given theoretical and empirical support for the adults's use of leaps to fix the standard for *big*. The riddle that remains to be solved, as far as we are concerned, is that the children's behaviour turns out to be so conspicuously different. It is quite conceivable that the establishment of a leap-standard happens only after the age of eight, and should be interpreted in the light of

growing world-knowledge plus the economic advantages this type of standard brings. But there are other possibilities. It could be that the antonym *small*, as an alternative, has different ways of ‘running in the background’ in children and in adults. Maybe children do not access the antonym unless it is made salient, and therefore make a two-way categorization (only *big* and non-*big*, but not *small*). The mid-range standard in this experiment would then result from failure by children to activate *small* and make room for its denotation. This also raises the option that children are in fact sensitive to leaps, but get confused by the fact that there were two of them, of exactly the same size; thus the low and high gaps were equally good candidates for the transition midpoint. By this line of thought adults automatically activate antonyms, thus using only the high gap for *big* (Singh et al., 2013; van Tiel et al., 2013).

Alternatively adults might employ more than one constraint on categorization, i.e. both a leap constraint, and an above-mid-range constraint. Children are known often to fail employing conjunctive rules because this is costly in terms of processing. So another possible explanation for our results might be that each child employed only one rule at a time: either a mid-range rule, or a leap rule. This would then explain our failing to see a preference for the high leap—or either leap—over the scale middle. Future research should therefore, apart from taking in account behaviour with respect to the antonym, investigate whether good performance in conjunctive tasks or other tasks that require executive functions matches ability to use (high) leaps (Zelazo et al., 1996, 2004; Sassoon, 2011).

References

- Barner, David, and Jesse Snedeker. 2008. Compositionality and statistics in adjective acquisition: 4-year-olds interpret tall and short based on the size distributions of novel noun referents. *Child development* 79:594–608.
- Bierwisch, Manfred. 1989. The semantics of gradation. *Dimensional Adjectives* 71:71–261.
- Ehri, L.C. 1976. Comprehension and production of adjectives and seriation. *Journal of Child Language* 3:369–384.
- Gaifman, Haim. 2010. Vagueness, tolerance and contextual logic. *Synthese* 174:5–46.
- Kennedy, Christopher. 2007. Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy* 30:1–45.
- Kennedy, Christopher, and Louise McNally. 2005. Scale structure, degree modification, and the semantics of gradable predicates. *Language* 345–381.
- Klein, E. 1980. A semantics for positive and comparative adjectives. *Linguistics and Philosophy* 4:1–45.
- Ludlow, P. 1989. Implicit comparison classes. *Linguistics and philosophy* 12:519–533.
- McNally, Louise. 2011. The relative role of property type and scale structure in explaining the behavior of gradable adjectives. In *Vagueness in communication*, ed. Rick Nouwen, Robert van Rooij, Uli Sauerland, and Hans-Christian Schmitz, Lecture Notes in Computer Science, 151–168. Berlin: Springer.
- van Rooij, R.A.M. 2010. Implicit versus explicit comparatives. In *Vagueness and language use*, ed. P. Egré and N. Klinedinst, Palgrave Studies in Pragmatics, Language and Cognition, 51–72. Palgrave Macmillan.
- van Rooij, R.A.M. 2011. Vagueness and linguistics. In *Vagueness: A guide*, ed. Giuseppina Ronzitti, Logic, Epistemology, and the Unity of Science, 123–170. Berlin: Springer.

- Rotstein, Carmen, and Yoad Winter. 2004. Total adjectives vs. partial adjectives: Scale structure and higher-order modifiers. *Natural Language Semantics* 12:259–288.
- Sassoon, Galit, and Assaf Toledo. 2011. Absolute and relative adjectives and their comparison classes. Unpublished manuscript.
- Sassoon, Galit W. 2011. Adjectival vs. nominal categorization processes: The rule vs. similarity hypothesis. *Belgian Journal of Linguistics* 25:104–147.
- Schmidt, L.A., N.D. Goodman, D. Barner, and J.B. Tenenbaum. 2009. How tall is tall? Compositionality, statistics, and gradable adjectives. In *Proceedings of the 31st annual conference of the cognitive science society*, 2759–2764.
- Singh, Raj, Ken Wexler, Andrea Astle, Deepthi Kamawar, and Danny Fox. 2013. Children interpret disjunction as conjunction: Consequences for the theory of scalar implicature? Unpublished manuscript.
- Smith, L.B., N. Cooney, and C. McCord. 1986. What is “high”? The development of reference points for “high” and “low”. *Child Development* 57:583–602.
- Solt, S. 2011. Notes on the comparison class. In *Vagueness in communication*, ed. Rick Nouwen, Robert van Rooij, Uli Sauerland, and Hans-Christian Schmitz, Lecture Notes in Computer Science, 189–206. Berlin: Springer.
- Solt, S., and N. Gotzner. 2012. Who here is tall? Comparison classes, standards and scales. In *Sfb 833, pre-proceedings of the international conference linguistic evidence 2012*, 79–83. Eberhard Karls Universität, Tübingen.
- Sperber, Dan, and Deirdre Wilson. 1995. *Relevance: Communication and cognition, second edition*. Oxford/Cambridge: Blackwell Publishers.
- van Tiel, Bob, Emiel van Miltenburg, Natalia Zevakhina, and Bart Geurts. 2013. Scalar diversity. Unpublished manuscript.
- Tribushinina, Elena. 2011. Once again on norms and comparison classes. *Linguistics* 49:525–553.
- Tribushinina, Elena. 2013. Adjective semantics, world knowledge and visual context: comprehension of size terms by 2- to 7-year-old Dutch-speaking children. *Journal of Psycholinguistic Research* 42:205–225.
- Yoon, Youngeun. 1996. Total and partial predicates and the weak and strong interpretations. *Natural Language Semantics* 4:217–236.
- Zelazo, Philip David, Fergus IM Craik, and Laura Booth. 2004. Executive function across the life span. *Acta psychologica* 115:167–183.
- Zelazo, Philip David, Douglas Frye, and Tanja Rapus. 1996. An age-related dissociation between knowing rules and using them. *Cognitive Development* 11:37–63.

