Cumulative comparison: experimental evidence for degree cumulation

Rick Nouwen and Jakub Dotlačil

In this paper we address the question whether it makes sense to assume that the domain of degrees, as used in degree semantics, consists not just of atoms, but also of *degree pluralities*. A number of recent works have adopted that assumption, most explicitly Fitzgibbons et al (2008); Beck (2014); Dotlačil and Nouwen (2016). In this paper, we provide experimental evidence for degree pluralities by showing that comparatives may express cumulative relations between degrees.

1 Adjectives and plurality

Adjectives can be distributive or collective. For instance, "tall" is related to the *height* of an individual and is, as such, intrinsically concerned with atoms only: there is no such thing as the height of John and Mary as a group. Other adjectives are different. For instance, the adjective "compatible" is inapplicable to atoms, since degrees of compatibility can only be assigned to groups of entities. This situation is familiar from the semantics of non-adjectival plurality. Predicates like "be a team" are collective in the same sense as "compatible" is, while predicates like "being wounded" are distributive in the same sense as "tall".

A dominant way of thinking about this distinction is that collective predicates have no atomic individuals in their extension, while distributive predicate have nothing but atoms in their extension. One clear advantage of this is that it enables us to account for why collective predicates are never com-

Rick Nouwen
Utrecht University, e-mail: R.W.F.Nouwen@uu.nl;
Jakub Dotlačil
Groningen University, e-mail: j.dotlacil@gmail.com

patible with singular arguments, while distributive ones *are* compatible with plural arguments, as illustrated in (1) and (2).

(1) *John is a good team.

(2) John and Mary are wounded.

To account for (2), all we need to assume is that, at least for plural cases like this, the extension of the predicate is closed under so-called *sum formation*: the operation that forms pluralities $a \sqcup b$ out of atoms a and b. The operator that enforces this kind of closure is often notated as * (after Link, 1983). What accounts for the contrast between (1) and (2) is the fact that * can create pluralities out of atoms, but it cannot create atoms out of pluralities. In other words, for a predicate with a non-empty extension, the extension of $^*P$ will always contain non-atomic entities, while it is not guaranteed to contain atomic ones.

Things are no different for adjectives. It would be natural to assume that "tall" expresses a relation between atomic entities and degrees and that "compatible" expresses a relation between non-atomic entities and degrees. Once the degree slot has been saturated, we will have a distributive predicate for the case of "tall" (as in "being tall" or "being two meters tall") and a collective one in the case of "compatible".[1]

In other words, considerations of plurality appear to have to do with how predicates map to the domain of entities and in particular to the complexity of the members of their extension. It appears then that plurality does not play a role on the side of degrees. Comparison, for instance, is a purely atomic relation, even if the adjective involved is collective. Take (3) as a case in point.

(3) John and Mary are more compatible than Peter and Sue.

What is at stake in (3) is whether the degree of compatibility of the sum of John and Mary exceeds the degree of compatibility of the sum of Peter and Sue. Clearly, even though the components of the comparative are plural, the degrees are not. The same point can be made for (4).

(4) John and Mary are taller than Peter.

[1] Whereas "tall" lacks collective readings and "compatible" lacks distributive readings, so-called additive adjectives allow both. For instance, (i) either means that the totality of boxes is heavy or that each of them individually exceeds some standard of weight.

(i) The boxes are heavy.

This observation can be straightforwardly taken into account by assuming that the underlying measure function (weight) makes sense for both atoms and plurals in the sense that $\mu(\alpha \sqcup \beta) = \mu(\alpha) + \mu(\beta)$. A sentence like (i) is now ambiguous between a distributive one in which the weight of each atomic box is at stake or a collective one in which the added up weight of all the boxes is taken into account.

Here, the predicate "to be taller than Peter" is a distributive predicate, which can be closed under sum formation using the * operator. In this way, it can apply to plural arguments even though both the relation expressed by "tall" and the comparison relation expressed by the comparative only have atoms in their extension.

So far, then, we have seen that where adjectives interact with pluralities, there is no evidence for the need for plural degrees. What is plural in all the above cases is an e-type argument of an adjectival or a comparative relation. Recently, however, there have been a number of proposals that assume the existence of non-atomic degrees. Fitzgibbons et al (2008), for instance, analyse sentences where a superlative predicate is combined with a plural subject (*John and Paul are the tallest students*) as involving a fully pluralised adjective, i.e. a relation between potentially plural entities and potentially plural degrees. Very recently, two other accounts of plural degrees were developed, both aiming for to improve on existing analyses of comparatives: Beck (2014) and Dotlačil and Nouwen (2016). In the remainder of this paper, we will present experimental evidence for such proposals. Focusing on our own account, we will start by summarising the plural degree approach to comparatives.

2 Plural degrees

At the heart of the plural degree approaches to comparatives lies a classical puzzle of the semantics of comparatives, namely how to account for the interpretation of comparatives that have quantified *than* clauses (von Stechow, 1984; Larson, 1988; Heim, 2006). The best paraphrase for a sentence like (5) is one in which the quantifier takes scope outside of the *than* clause, as in (6).[2]

(5) John is taller than every girl is.

(6) Every girl x is such that John is taller than x.

The same intuition holds for differentials: (8) is a very good analysis of (7), since it correctly predicts that (7) entails that every girl has the same height.

(7) John is exactly 2 inches taller than every girl.

(8) Every girl x is such that John is exactly 2 inches taller than x.

[2] Clausal comparatives like (5) are perhaps not entirely natural to all native speakers, presumably given the (simpler) phrasal alternative *John is taller than every girl.* However, our theory is a theory of *clausal* comparatives and so it is important that we only consider that class of comparatives. Note that in our experiment, below, we escape the potential unnaturalness of sentences like (5) by turning to subcomparatives (*the table is longer than the door is wide*). There are of course no phrasal paraphrases of subcomparatives and so these sentences are entirely natural.

The issue is that *than* clauses are islands, cf. the ungrammatical status of the following example from Larson (1988):

(9) *I wonder which door the table is longer than _ is wide.

This fact makes (6) and (8) useless as blue-prints for the semantic structure of comparatives and differentials, for they would require an island violation. Since Schwarzschild and Wilkinson (2002), semanticists have standardly observed this restriction and developed several accounts that derive the correct interpretation without the island violation.

In line with this tradition, in Dotlačil and Nouwen (2016) we also claim that the wide scope of the quantifier is an illusion. On our account, what rather happens is that the *than* clause denotes a plural degree, namely the sum of degrees containing (nothing but) the heights of every girl. The sentence is then interpreted distributively in the sense that for each atom in that sum, John's height has to exceed that atom.

Before we say a little bit more about the framework that facilitates such an analysis, we zoom out a bit. If works like this or Beck (2014) or Fitzgibbons et al (2008) are on the right track, then it suggests that the domain of degrees is no different from the domain of entities: both contain plural individuals and the relations we build on top of them are interpreted with respect to the same mechanisms, in particular, as we will see below, distributivity and cumulativity. In this paper, we follow this intuition. If degree plurality is like entity plurality, then we expect to see effects of plurality beyond the phenomena for which we designed the plural framework. That is, we should find evidence for plural interpretation beyond the simple comparatives in (5) and (7).

2.1 A framework for plural degree semantics

We take degrees to be discrete, atomic entities that are ordered by some ordering $>$. The plural degree semantics we developed in Dotlačil and Nouwen (2016) subscribes to the assumption that atomic degrees may combine to form sums. So, on top of the set of atomic degrees, there is also a set of non-atomic degrees, built from these atoms. If d and d' correspond to two different heights, then $d \sqcup d'$ is the collection that contains nothing but these degrees. Since we take degrees to be discrete, $d \sqcup d'$ equals d only if $d = d'$.

Above, we discussed how distributive $\langle e, t\rangle$-type predicates (i.e. predicates which only have atomic entities in their extension) can take plural arguments by closing the extension under sum formation using the * operator. The interpretation of * for a predicate P is as follows:

(10) a. $P \subseteq {}^*P$
b. If $\alpha \in {}^*P$ and $\beta \in {}^*P$, then also $\alpha \sqcup \beta \in {}^*P$.

c. Nothing else is in $^*P$.

To get a feel of what this definition does, let us briefly explain how, for instance, $^*\{a,b\}$ equals $\{a,b,a \sqcup b\}$. (10-a) states that $\{a,b\} \subseteq {}^*\{a,b\}$. The next condition states that $a \sqcup b \in {}^*\{a,b\}$. (10-c) adds that no other element is in $^*\{a,b\}$. In sum, $\{a,b\} \subseteq {}^*\{a,b\}$. For any atomic predicate P, the result is that $^*P$ is only true of a plurality if P is true of each of the atoms of that plurality.

In the literature, a parallel operator exists for $\langle e, \langle e, t\rangle\rangle$-type relations (Krifka, 1989; Sternefeld, 1998; Beck and Sauerland, 2000), often written as **. This is a generalisation of the * operator for sets of pairs of entities instead of just for sets of entities. For R a set of pairs:

(11) $^{**}R$ is the smallest superset of R such that if $\langle \alpha, \beta\rangle \in {}^{**}R$ and $\langle \alpha', \beta'\rangle \in {}^{**}R$ then also $\langle \alpha \sqcup \alpha', \beta \sqcup \beta'\rangle \in {}^{**}R$.

For example, $^{**}\{\langle a_1, b_1\rangle, \langle a_2, b_2\rangle\}$ equals $\{\langle a_1, b_1\rangle, \langle a_2, b_2\rangle, \langle a_1 \sqcup a_2, b_1 \sqcup b_2\rangle\}$. The effect on an originally atomic relation R is that two pluralities A and B stand in the $^{**}R$ relation if for each atom x in A there is at least one atom y in B such xRy and for each atom y in B there is at least one atom x in A such that xRy. This means that interpretative effect of ** is a cumulative reading (Scha, 1981). For instance, when (12) is interpreted as (13), it yields the truth-conditions in (14). This makes the sentence true in a situation in which one boy carried two of the boxes and the other one carried the remaining boxes. In such a situation, the distributive reading is false.

(12) The two boys carried the four boxes.

(13) [the two boys [[** carried] the four boxes]]

(14) Each of the two boys carried some of the four boxes and each of the four boxes was carried by (at least) one of the two boys.

The operations * and ** suffice to account for *distributive*, *collective* and *cumulative* readings of (12). Applying the * operator to the predicate [*carried* [*the four boxes*]] allows it to take a plural subject. The resulting reading is compatible with both a distributive and a collective understanding of the sentence (depending on whether or not the extension of the predicate already contained plurality - i.e. boys jointly carrying boxes - or not). On the collective reading, neither of the boys carried the four boxes by themselves, they only did so collectively. Note that this is different from the cumulative reading, which does not entail that any collective carrying took place.[3]

The resulting framework is a minimal plural semantics for (predicates over) the domain of entities. For the case of degrees, we now assume that: (i) the

[3] There may be reasons to think that the distributive and collective understanding are proper *readings*, in which case one needs to posit a distributivity operator that quantifies over atoms (see Lasersohn 1998 for discussion). To keep things simple, we will remain agnostic with respect to this issue, which is orthogonal to our focus below.

domain of degrees contains both atoms and sums, just like the domain of entities; (ii) predicates and relations that involve degrees can be interpreted using the * and ** operators; (iii) the degree comparison relation $>$ is a relation between atomic degrees.

2.2 Quantified than-clauses as degree pluralities

This framework can now be used to solve the puzzle of quantified *than* clauses. The idea is that *than* clauses denote potentially plural degrees, using the interpretation scheme in (15). (See Dotlačil and Nouwen (2016) for details of an underlying compositional semantics, and Beck (2014) for an alternative.)

(15) ⟦than Q/DP is tall⟧ = the smallest degree plurality that contains the height of Q/DP

For a DP like *Mary*, this scheme is going to return the smallest plurality that contains the height of Mary, which is simply the atomic degree Mary's height. For a QP like *every girl*, this scheme is going to return the smallest plurality that contains the height of every girl: girl_1's height$\sqcup \ldots \sqcup$$\text{girl}_n$'s height.[4]

If the *than* clause denotes a non-atomic degree, it is in principle incompatible with the comparative semantics, since, as we said above, only atomic degrees are ordered and, so, degree comparison is comparison of atoms only. This means that in order to interpret a comparative with a *than* clause containing a quantificational element, we need to pluralise the comparison relation. For the case of *John is taller than every girls is*, we get:

(16) John's height $^{**}\!>$ girl_1's height$\sqcup \ldots \sqcup$$\text{girl}_n$'s height

The relation $^{**}\!>$ is true of pluralities A and B if and only if each atom in A exceeds some atom in B and each atom in B is exceeded by some atom in A. If A itself is atomic, this simply boils down to this atom exceeding each atom in B, and so, for (16) to be true, John's height has to exceed all the atoms in the plurality of girl heights, which entails him being taller than the tallest girl. In other words, what in (5)-(8) seemed like a distributive quantifier taking wide scope is really the distribution over atoms in a plurality stemming from the need to pluralise an atomic relation that got given a non-atomic argument.

[4] In Beck (2014), *than* clauses with *every* denotes plural degrees by virtue of the fact that *every* DP can have group readings. This is problematic because *than* clauses with *each* DPs yield the same readings as those with *every* DPs, but it is well-known that *each* DPs do not have group readings (Beck, 2014, p. 101-102). While the paraphrase in (15) may suggest that our proposal suffers from the same problem, we should hasten to add that (15) is only a very rough paraphrase of our proposal. In Dotlačil and Nouwen (2016) we compositionally derive plural degree denoting *than* clauses with both *each* and *every* DPs, based on their *distributive* semantics.

2.3 A predicted effect: cumulative comparison

We are assuming that if degrees can be plural then all the interpretation mechanisms we observe for the domain of entities should in principle also be available for degrees. We see no reason to assume a watered-down version of plural semantics for degrees, for instance where degree pluralities exist but relation cumulativity over degree relations does not. The account we sketched for quantified *than* clauses already suggests that this assumption is on the right track. This kind of view, however, also predicts that we should be able to observe further effects of plurality. In particular, the availability of ** accounts for cumulative readings for sentences like (12) and, so, we would expect to see true cumulative readings for $^{**}>$. The interpretation (16) of *John is taller than every girl is* is not evidence for that, since that interpretation is equivalent to the distributive reading we get by pluralising a derived predicate $\lambda d.John'sheight > d$ and applying it to the plurality denoted by the *than* clause.

(17) $[^{*}(\lambda d.\text{John's height}> d)](\text{girl}_1\text{'s height}\sqcup \ldots \sqcup \text{girl}_n\text{'s})$

In order to find true cumulative readings, we need two plural arguments. The literature contains at least one influential example of where we might find such a reading.

(18) The frigates were faster than the carriers. (Scha and Stallard, 1988)

One possible interpretation of (18) is one in which there were groups of ships and in each group the frigates in that group were faster than the carriers in that group. On that reading, the subject distributivity reading is false, since there may be carriers that were faster than one or more frigates, as long as they were not in the same group.

In order to account for this reading, it is natural to resort to **. But for cases like (18), one need not assume that such an operator functions in the domain of degrees. Indeed, Scha and Stallard (1988), Schwarzschild (1996) and Matushansky and Ruys (2006) all analyse (18) as a cumulative relation between *entities*. That is, since (18) is a phrasal comparative, we can analyse it as a relation between entities (here, the frigates and the carriers) and so we can cumulate that relation using **.

This means that examples like (18) are not evidence for a cumulative interpretation of the degree relation $>$, but one could think that its clausal counterpart (19) is.

(19) The frigates were faster than the carriers were.

Clearly, (19) shares with (18) the same cumulative-like reading. However, in order to analyse (19) as a relation between entities, we would need to move out the subject of the *than* clause.[5] This is because clausal comparatives cannot be understood as relations between entities, given that one of the 'comparees' is a clause. To turn it into a relation over entities, we would somehow need to abstract over the subject in that clause, something we assume not to be a viable option, given that it would constitute an island violation. This suggests, then, that perhaps (19) does not involve a cumulative relation between entities, but one between degrees. Still, as we explain in Dotlačil and Nouwen (2016) in more detail, (19) is still not definitive proof that cumulative comparison exists. This is because we could arrive at exactly the same truth-conditions using distributivity and dependency. As Winter (2000) shows, cumulative readings are often indistinguishable from distributive readings. For (19), that reading would be along the lines of (20).

(20) The frigates EACH_i were faster than [the carriers]$_i$ were.

The idea is that the definite *the carriers* is interpreted as being dependent on the frigates. All one needs to assume is that distributivity can *bind* definites, something we need anyway to account for examples like (21), which has one reading in which each boy thinks that *he* is the tallest, instead of attributing the contradictory thought to him that all the boys are the tallest.

(21) The boys each think they are the tallest.

This means that if we want to show that cumulative comparison exists we need to use examples with two features: (i) we have to avoid phrasal comparatives, like (18), and use clausal comparatives instead, since phrasal comparatives may be understood as cumulative relations between entities, not degrees; (ii) we need to exclude the option of cumulative-like truth-conditions arising through dependent interpretation. We can accomplish the latter by resorting to distributive quantifiers. Consider for instance a minimal variation on (21): (22).

(22) The boys each think each of them is the tallest.

Whereas (21) has a reading in which *they* depends on the distributive quantification over boys, (22) lacks such a reading. The reason is that if *them* in (22) is interpreted dependently, it will refer to single boys and this renders the distributive quantification by *each* inappropriate, cf:

(23) *Each of John is sick.

Using this for our quest to find cumulative comparison, we arrive at examples like (24): This is an example of a clausal comparative, where there is no option of the subject of the comparative clause to depend on distribution over the matrix subject.

---

[5] In fact, that would not suffice to gain a relation. See Dotlačil and Nouwen (2016).

(24) The frigates were faster than each of the carriers were.

Intuitions are admittedly murky, here, and there are several complications: not least of all the fact that the distributive reading tends to be more readily available than the cumulative one, even already for the much simpler (19). For this reason, we turn to an experimental setting, in which we probe the truth-conditions participants assign to sentences of the shape in (24).

3 The experiment

We tested interpretations of comparatives in a simple verification task. The goal was to find to what extent cumulative readings of clausal comparatives are accepted and how the level of acceptance compares to other readings one might associate with clausal comparatives. The experiment was run in Dutch.

3.1 Experimental setup

In the experiment, participants were first given a cover story which told of a fictional study that compared people's ability to write (by hand) and type (on a keyboard) in a wide array of different circumstances. For each trial, this study recorded the writing and typing speeds of the participants in the cover story. Each stimulus of our experiment consisted of a fictional graph from the fictional study, depicting the typing and writing speed of three participants for a single trial. Figure 1 shows an example of such a graph. (The original stimuli were in Dutch and contained colours instead of shading.) Here, the speeds of three (fictional) participants (p.1, p.2 and p.3) are displayed. Shaded bars indicated the speed of their handwriting, non-shaded bars the speed of their typing in the trial.

Graphs like these were displayed with sentences that were supposed to provide a true statement about the trial in question. Participants in our experiment had to decide whether the statement was indeed correct.

There were two types of test items appearing with graphs. In the test items, the *than* clause included a distributive universal quantifier (glossed as dist[6]), (25-a), or a plural definite anaphor, (25-b).

[6] We gloss it as such to avoid the issue of deciding whether *elk* in Dutch corresponds more closely to the distributive quantifier *every* or to the distributive quantifier *each*. Syntactically, it behaves like *each*: it appears in partitive constructions and can function as a floating quantifier. But semantically, it express distributivity but it does not seem to force the differentiation condition associated with *each* (Tunstall, 1998; Brasoveanu and Dotlačil, 2015).

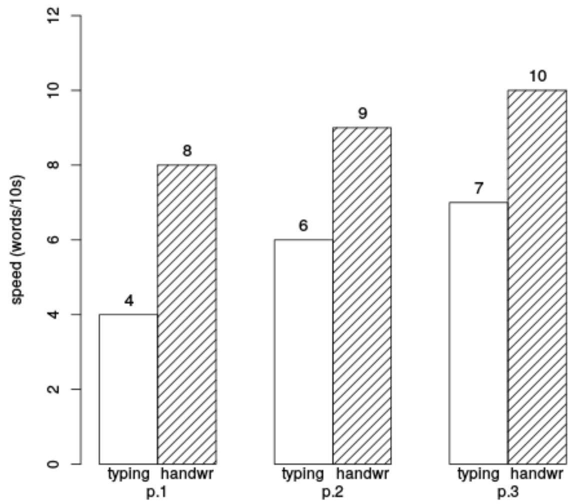

Fig. 1 An example plot used in the experimental stimuli

(25) a. De deelnemers typten sneller dan elk van hen schreef.
The participants typed faster than dist of them wrote
UNIVERSAL

b. De deelnemers typten sneller dan ze schreven.
The participants typed faster than they wrote
PLDEF

The test items had a verb in the *than* clause and consequently, they had to be treated as clausal comparatives.

Each barplot graphically summarized six data points representing the typing and writing speed of the three participants, as illustrated in Figure 1. For ease of exposition, we will represent the graphs used in stimuli by enlisting the typing speed / writing speed pairs of the three participants. For instance, the shorthand for Figure 1 is $\langle 4-8, 6-9, 7-10\rangle$.

It depended on the available readings whether a sentence was compatible with its accompanying barplot or not. We focused on two readings, the distributive and the cumulative reading. These are represented by the propositions in (26) and (27), respectively.

(26) Each of the three recorded typing speeds exceeds each of the recorded writing speeds. distributive reading

(27) For each of the three recorded typing speeds there exists a recording writing speed that is slower and for each of the three recorded writing speeds there exists a typing speed that is faster. cumulative reading

There were 5 tested scenarios for the experimental items. They are summarized in the following table:

Name	Example	Distr. reading	Cumul. reading
dist	$\langle 8-5, 10-6, 12-3 \rangle$	true	true
cumul1	$\langle 6-5, 10-7, 12-3 \rangle$	false on 1 account	true
cumul2	$\langle 8-6, 6-5, 5-4 \rangle$	false on 2 accounts	true
noreading1	$\langle 4-8, 9-5, 7-6 \rangle$	false	false on 1 account
noreading2	$\langle 7-8, 9-5, 2-3 \rangle$	false	false on 2 accounts

To illustrate the idea behind this setup let us go through the examples. First of all, the example given for *dist* clearly verifies (26), since $8 > 5$, $8 > 6$, $8 > 3$, $10 > 5$, $10 > 6$, $10 > 3$, $12 > 5$, $12 > 6$ and $12 > 3$. Since in our setup, the distributive reading entails the cumulative one, (27) is true too. In the case of *cumul1*, the distributive reading is false. This is because participant 2 wrote faster than participant 1 typed: $7 > 6$. The cumulative reading is still true though, since $6 > 5$, $10 > 7$ and $12 > 3$. That is, the fact that for each participant it was the case that the typing speed exceeded the writing speed satisfies the requirements for the cumulative reading as stated in (27). This requirement also holds in the case of *cumul2*, but here the distributive reading is false on two accounts. Firstly, the typing speed of participant 2 does not exceed the writing speed of participant 1 and the typing speed of participant 3 does not exceed the writing speed of either participant 1 or 2.

In the cases of *noreading1* and *noreading2* both the cumulative and the distributive readings are false. We distinguish two cases here. In *noreading1*, two participants satisfied the cumulative relation imposed by the comparative, while one participant violated it. In the example in the table above, the problematic participant is participant 1 since he typed slower than he wrote (4 vs. 8). In *noreading2*, two participants violated the cumulative relation imposed by the comparative (in the example above, these are participant 1[7] and participant 3). For this reason, the cumulative reading is false in *noreading1* on one account (participant 1) and false in *noreading2* on two accounts (participants 1 and 3).

[7] Why participant 1? According to the definition in (27), participant 1 should not be problematic since he types faster than some of the other participants write, and he writes slower than one of the other participants type. However, following Schwarzschild (1996), among many others, we assume that cumulative relations are also sensitive to context, which determines which typing/writing speeds are compared. In the case at hand, the context requires that each participant's typing is compared to the writing of the same participant. Since participant 1 types slower than he writes, he presents a case violating the cumulative relation.

	dist	cumul1	cumul2	noreading1	noreading2
D&N16	top	top/high	top/high	bottom	bottom
No ** over degrees	top	bottom	bottom	bottom	bottom
Leniency	top	lower	lower still	even lower	bottom

Fig. 2 Predictions in terms of proportion of responses in which participants respond that the sentence correctly describes the graph for the UNIVERSAL item

3.2 Predictions

The theory of Dotlačil and Nouwen (2016) predicts the following. For the test sentence without the distributive quantifier, i.e. the PLDEF item, both the distributive and the cumulative reading should be available. The former should be the default interpretation, derivable via pluralisation of the matrix predicate. The latter is available in two distinct ways: (i) via the cumulative operator, **; (ii) via the distributive operator in tandem with a dependent interpretation of the pronoun. For the test sentence with the distributive quantifier, the UNIVERSAL item, it should also be the case that both readings are available. However, now the option of arriving at the cumulative interpretation via a dependent analysis of the pronoun is excluded. If for some reason inserting * is preferred over inserting **, then we would furthermore predict higher rates of acceptance for *dist* than for *cumul1* and *cumul2*.

Theories that do not have the option of interpreting the comparison relation cumulatively would potentially make the same prediction as Dotlačil and Nouwen (2016) for the PLDEF item, since in the absence of cumulativity, *cumul1* and *cumul2* are still compatible with a distributivity plus dependency reading. However, for UNIVERSAL items this reading is unavailable, so here one would predict low acceptability for all conditions, except for *dist*. The only way we could imagine higher scores is if participants somehow allow exceptions on the distributive quantification. In this case, you'd expect a slippery scale from universal acceptance for the case of *dist*, lower acceptance for *cumul1* and then continuously lower acceptance for *cumul2*, *noreading1* and *noreading2*. The differing predictions are summarized in Figure 2.

3.3 Methodology

3.3.1 Participants

44 native Dutch speakers participated in the experiment. 38 of them were students from the University of Groningen who either volunteered or received a course credit for their participation. 6 participants were volunteers from Utrecht University.

3.3.2 Materials and procedure

The experiment consisted of graph-sentence pairs, as described in section 3.1. Participants had to decide whether the sentence was true or false given the situation captured in the graph. Two sentence types were tested (PlDef vs. Universal) in five scenarios (*dist*, *cumul1*, *cumul2*, *noreading1* and *noreading2*). Two items per scenario were created (10 items in total). Two lists were created out of the items, so that in each list only one sentence type was present for each item. Every participant was assigned to one of the lists.

Apart from 10 experimental items, the experiment consisted of 2 practice items and 24 fillers. The fillers were unambiguously true/false (e.g., for Figure 1 one filler might be the true sentence *Participant 3 typed slower than he wrote*). Fillers and experimental items were randomly ordered and each stimulus appeared on a separate screen (with no backtracking possible).

The whole experiment was run in Ibex and hosted on Ibex Farm (see http://spellout.net/ibexfarm/).

3.4 Results

Just one participant made more than 3 mistakes in the 24 fillers. Except for this one individual, we kept all the participants for the analysis.

Figure 3 shows the results. The percentages indicate proportionally how many participants responded that the sentence correctly describes the graph.

For the analysis, we focus on the Universal factor since this is the part at which theories make different predictions. We consider logistic regression with one factor – Reading. We consider two different models. In the first one, Reading consists of two levels, *distributive reading* (consisting only of *dist*) and *no reading* (consisting of all the other cases, i.e., *cumul1*, *cumul2*, *noreading1*, *noreading2*). This is the model that is appropriate for theories that assume no **. In the second model, *cumul1* and *cumul2* are treated as a separate factor from *noreading1/2*, that is, Reading consists of three levels. This is the model appropriate for Dotlačil and Nouwen (2016). Somewhat unsurprisingly (given the graphical summary in Figure 3), we see that using

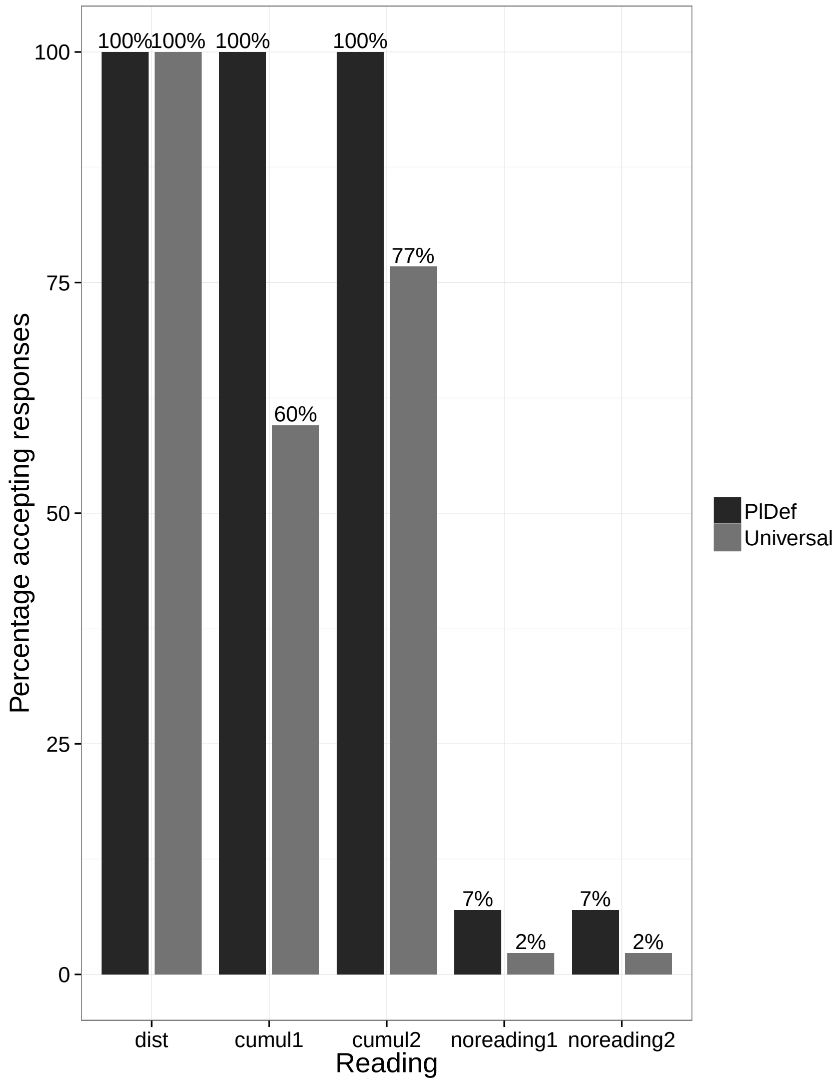

Fig. 3 Experimental results

the three-level factor of Reading improves the model fit compared to the two-level factor ($\chi^2(1) = 96, p < .001$).

As we noted in Section 3.2, the theory lacking ** for comparatives could predict higher scores in *cumul1* than, say, *noreading1* if it somehow allowed exceptions on the distributive quantification. But in that case there should be a slippery scale from universal to *noreading2*. This is not the case when we look at Figure 3. Here's one way to quantify this claim. If the acceptability decreased with the number of exceptions, we might expect that responses in readings *cumul1*, *cumul2*, *noreading1* and *noreading2* would form some form of linear function. Therefore, fitting it using logistic regression with one independent variable (the number of exceptions to make the dist reading true) would be appropriate. On the other hand, if Dotlačil and Nouwen (2016) are right, such a linear fit simplifies the picture. Instead, we can consider a general additive model, in which the model itself is left to find the best smooth function over the number of exceptions (using the mgcv package, see Wood (2006)). This logistic model has one dependent variable, the number of exceptions to make the dist reading (with 5 knots), and RESPONSE is the dependent variable. It turns out, perhaps unsurprisingly, that the logistic general regression model fits our data significantly worse than the logistic general additive model ($\chi^2(2.8) = 38, p < .001$). One potential worry is that we might be overfitting the model in the former case. Importantly, though, the logistic general additive model is not significantly better than the simple model we considered above: logistic regression with one variable, Reading, which has three levels (*distributive reading*, *cumulative reading* and *no reading*) ($\chi^2(1.8) = 2.5, p > .1$). In conclusion we can say that in our search for the right model to fit the data with the distributive quantifier, the model that assumes that there is a distributive and cumulative reading (and nothing else) is the best. This supports Dotlačil and Nouwen (2016).

Finally, we note that it is clear from Figure 3 that *cumul1/2* readings are more acceptable in PLDEF than in UNIVERSAL. This is compatible with our account under the assumption that * is preferred over ** and this preference is further corroborated by the higher acceptability of *dist* readings in UNIVERSAL.

3.5 Discussion: collective readings?

Readers familiar with Scontras et al (2012) might recall other cases in which comparatives do not seem to be interpreted distributively. For instance, one may judge (28) to be true of a depiction of blue and red dots, even if there is one red dot that is smaller than every blue dot, as long as the average size of red dots exceeds that of blue dots.

(28) The red dots are bigger than the blue dots.

The experiments of Scontras et al (2012) suggest that plural comparatives like (28) are indeed sometimes interpreted collectively. This means that subjects tend to interpret such sentences in terms of a comparison between an aggregate degree of size for the red dots and an aggregate degree of size for the blue ones.

It is not immediately clear whether the observations in Scontras et al (2012) are relevant to our present study. First of all, the sentences in their experiments we always phrasal comparatives. Our theory in Dotlačil and Nouwen (2016) is a theory of clausal comparatives and so our current experiment deliberately only contains clausal comparatives as stimuli. Second, the sentences used by Scontras et al. have definite plurals in the *than* clause. It is not clear whether the collective reading is available once this definite is replaced by a distributive quantifier, as in the crucial stimuli in our experiment.

Ignoring these questions, could our results be understood as cases of collective comparison? We do not think so. First, while dot sizes might be easily imaginable as aggregate, supporting the collective interpretation for comparatives, our setup stressed individuals' writing/typing achievements. The focus on individual performances makes it unlikely that participants would consider collective comparisons in our experiment. Second, in our items, *all* conditions, including the *noreading* ones, were created in such a way that the collective reading would be true. For a sentence like *Participants typed faster than they wrote* the total typing speed was always faster than the total writing speed, in any of the test conditions. Consequently, the average typing speed exceeded the average writing speed for such a sentence.[8] Consequently, if the average-based collective reading is an option, we would expect that no condition would be rejected. But this is not the case: both *noreading* conditions were almost universally rejected.

In some cases, however, the difference between the average speeds is hard to gauge. It could be that the collective reading only results in "correct" responses when the difference between the averages is clear enough. That is, the likely-hood of accepting the sentence increases when the difference between the average increases.

To test this, we looked at the average-differences in the test item and how these influenced responses, see Figure 4. If the average difference played a role, we would expect that the proportion of accepting responses increases with the difference. This is clearly not the case. In particular, the difference of 2 is almost fully rejected even though lower differences between averages are almost fully accepted. To test this further, we considered a logistic regression model in which the difference in averages is a linear predictor. The model is significantly worse than the model we considered above as the one supporting Dotlačil and Nouwen (2016) (i.e., the one which has a factor with three levels, *distributive reading*, *cumulative reading*, *no reading*): $\chi^2(1) = 175, p < .001$. Thus, categorizing our data into three reading types clearly has much more predictive power than considering the difference in averages.

[8] The items balanced the order of the comparison. Sometime typing was compared to writing, as in this example, but sometimes it was the other way around. In each case, however, the average speed corresponding to the verb in matrix clause exceeded the average speed corresponding to the verb in the *than* clause.

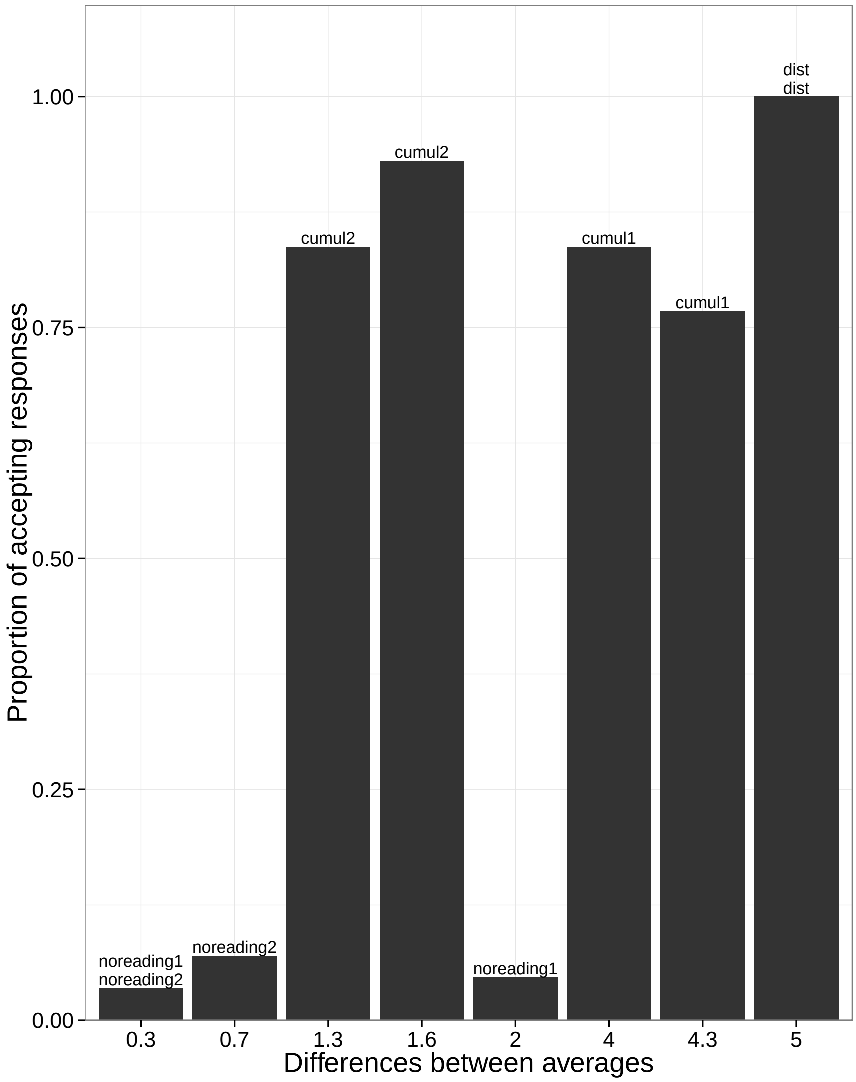

Fig. 4 Responses compared to differences between average speeds

4 Conclusion

We discussed the semantics of comparatives and a new analysis that employs plural degrees (Fitzgibbons et al, 2008; Beck, 2014; Dotlačil and Nouwen, 2016). Focusing on our own account, we argued that comparisons of plural degrees predicts a hitherto undiscussed reading, cumulative comparison. Controlling for several factors, we presented an experiment in which the relevant reading clearly surfaces. We take this as supporting evidence for the plural degree analysis of the comparative.

References

Beck S (2014) Plural predication and quantified than-clauses. The Art and Craft of Semantics: A Festschrift for Irene Heim 70:91–115

Beck S, Sauerland U (2000) Cumulation is needed: A reply to Winter (2000). Natural Language Semantics 8(4):349–371

Brasoveanu A, Dotlačil J (2015) Strategies for scope taking. Natural Language Semantics 23(1):1–19, DOI 10.1007/s11050-014-9109-1, URL http://dx.doi.org/10.1007/s11050-014-9109-1

Dotlačil J, Nouwen R (2016) The comparative and degree pluralities. Natural Language Semantics pp 1–34

Fitzgibbons N, Sharvit Y, Gajewski J (2008) Plural superlatives and distributivity, university of Connecticut / SALT proceedings

Heim I (2006) Remarks on comparative clauses as generalized quantifiers, semanticsArchive

Krifka M (1989) Nominal reference, temporal constitution, and quantification in event semantics. In: Bartsch R, van Benthem J, van Emde Boas P (eds) Semantics and Contextual Expressions, Foris, Dordrecht, pp 75–115

Larson R (1988) Scope and comparison. Linguistics and Philosophy 11:1–26

Lasersohn P (1998) Generalized distributivity operators. Linguistics and Philosophy 21(1):83–93

Link G (1983) The logical analysis of plurals and mass terms: A lattice-theoretical approach. In: Bäuerle R, Schwarze C, von Stechow A (eds) Meaning, Use and Interpretation of Language, Walter de Gruyter, Berlin, pp 302–323

Matushansky O, Ruys EG (2006) Meilleurs voeux: quelques notes sur la comparaison plurielle. In: Bonami O, Hofherr PC (eds) Empirical Issues in Formal Syntax and Semantics 6, pp 309–330

Scha R (1981) Distributive, collective, and cumulative quantification. In: Groenendijk J, Janssen T, Stokhof M (eds) Formal methods in the study of language, Mathematical Centre, pp 483–512

Scha R, Stallard D (1988) Multi-level plurals and distributivity. In: Proceedings of the 26th Annual Meeting of the ACL, Buffalo, New York

Schwarzschild R (1996) Pluralities, Studies in Linguistics and Philosophy, vol 61. Kluwer academic publishers, Dordrecht

Schwarzschild R, Wilkinson K (2002) Quantifiers in comparatives: A semantics of degree based on intervals. Natural Language Semantics 10:1–41

Scontras G, Graff P, Goodman ND (2012) Comparing pluralities. Cognition 123(1):190–197 von Stechow A (1984) Comparing semantic theories of comparison. Journal of Semantics 3:1–77

Sternefeld W (1998) Reciprocity and cumulative predication. Natural Language Semantics 6:303–337

Tunstall S (1998) The interpretation of quantifiers: Semantics and processing. PhD thesis, University of Massachusetts, Amherst

Winter Y (2000) Distributivity and Dependency. Natural Language Semantics 8:27–69

Wood S (2006) Generalized additive models: an introduction with R. CRC Press