Goldilocks and degree modification

Rick Nouwen

2nd draft, August 2020

1 Goldilocks Principles

In the story of Goldilocks and the three bears, the two parent bears and their child bear get ready to eat their breakfast, but find the porridge they prepared to be too hot. They decide to go for a stroll through the woods, leaving their breakfast bowls to cool. In their absence an impish golden-haired girl, Goldilocks, sneaks into their house. Among many other mischievous things she does, she tries out the bears' breakfast: Daddy's bowl of porridge is too hot, Mummy's bowl is too cold, but the little bear's porridge is *just right* and she proceeds to finish it off.

The Goldilocks fable does not excel in moral clarity. A common interpretation is that the story intends to show how the ill-natured selfish actions of Goldilocks affect the everyday life of the good-natured bears, where the girl's actions are particularly selfish because she takes only what is *just right*. The lasting legacy of the story, however, is simply this notion of *just right*-ness. It turns out that it is handy to have a cultural reference point for all sorts of instantiations of the ideal middle ground. The notion is similar to that of *the golden mean* from Aristotle's philosophy of virtue, but applicable much more broadly than to just the question of how to live life virtuously. In fact, the story of Goldilocks and the three bears has been a productive inspiration for quite a few scientific disciplines.

A typical *Goldilocks Principle* involves some ideal state of affairs that is positioned between states of affairs that are less ideal. Or stated differently, from the perspective of many different sciences, ideal situations tend to not be extreme situations, but rather moderate in some sense. One generally well-known example is from planetary science, where the so-called *Goldilocks zone* is the stretch of space around a star that allows liquid water to occur on a planet: it is sufficiently far from the sun to not let all water evaporate away but also not too far to only end up with ice. The details of how to establish which distance from the sun qualifies as a habital zone has subsequently become known as the *Goldilocks problem of climatology* (Kasting, Toon, and Pollack (1988)). In cognitive science, there's the *Goldilocks effect* (Kidd, 2012), the observation that the visual attention of infants tends to be highest when the stimulus is of intermediate complexity: while low complexity stimuli tend to already be familiar, too much complexity wastes valuable computational resources. In economy, there's the concept of a *Goldilocks economy*, which is an economy that has *Goldilocks growth*: growth that is prosperous but not so extreme that it causes excessive inflation.

Here, I will point out the relevance of Goldilocks to linguistic semantics. In fact, I think her to be particularly relevant, since my application will be to the semantics of *evaluation*. One way to understand Goldilocks principles is by seeing them as principles concerning

the very fabric of *evaluation*: that evaluation is sensitive to *excess*. What I mean by this will hopefully become clearer with my particular version of the Goldilocks concept. I will describe it in two separate Goldilocks principles:

The Goldilocks Principles of Evaluation

1. You can have too much of a good thing
2. You can't have so much of a bad thing that it becomes good again

These two principles together have exactly the intended effect. Being removed from the sun is *good* because it prevents the evaporation of water, but once you're too far, water will freeze and a good thing turns *bad*. And from there on, it will only get worse. Similarly, being close to the sun is *good* because it prevents the freezing of water, but, once again, being too close is *bad* because water will start to evaporate.

This connection between goodness, badness and excess can be applied much more generally, however, to the semantics of evaluative adjectives. An adjective is evaluative whenever its use conveys a positive or negative valuation by the speaker (e.g. Stojanovic (2015)). I will use the term *valence* for this polar value ascribed to something by means of an evaluative predicate.

Here's an example. Imagine some food that is really *bland* (=*negative valence*). By adding salt the food will become tasty (=*positive valence*), but if you keep on adding salt, the food will become *salty* (=*negative valence*). And no extra added amount of salt is ever going to make it taste fine again. The same reasoning works the other way around: You made a recipe using loads of salt and it was inedible (=*bad*). Next, time you reduce the salt and it is now enjoyable (=*good*). The next time you reduced the salt even further, only to discover that the food is now bland (=*bad*). Because we can go through this reasoning in either direction, we need only the two principles states here. (I could have added something like *You can have too little of a good thing*, but this is really not necessary. Not enough salt, is the same as having reduced the salt by too much.)

These principles are by no means universally valid. They tend to be true, but the world is not exclusively made up of Goldilocks-zones. For instance, sometimes more is better, ad infinitum, in breach with the first principle. A judge can't really ever be *too fair*. Ice cream can't really ever be *too tasty*. The strongest example of such a lack of excess is closeness to some ideal: if G is some ideal state, then obviously "being too close to G" is a non-sensical notion.

Crucial about Goldilocks principles is that they have to do with *excess*. The middle zone is ideal because it lacks excess: *goodness* occurs when excess doesn't. Conversely, *badness* involves any kind of excess, either of the extent to which some property holds or of the extent to which it does not hold. All this may seem pretty obvious, and I suppose it is. However, as I will show, these simple observations can explain why degree semantics is sometimes sensitive to valence. A simple illustration of this is with evaluative adjectives. Some adjectives seem to come with a fixed valence. For instance, the adjective *nice* has positive valence, and *nasty* is negative. However, for a lot of adjectives, their valence is not so fixed. Take *warm*, for instance. While warmth can naturally be seen as a source of good, Goldilocks teaches us that too much goodness can turn positive valence into a negative one. For instance, (1) is (typically) a positive evaluation, and (2) is (typically) a negative one.

(1) My coat is warm.
(2) My beer is warm.

Note now (2) is negative because the speaker is conveying an excess of warmness. If seen as a negative valence utterance, (2) expresses that the beer is *too* warm. The drink in question need not, in fact, be classifiable as *warm*. The speaker's point is that it is warmer than it is supposed to be. In line with this, (1) is positive only as long as the speaker does not intend to convey that the warmth brought by the coat exceeds what can be tolerated.

In a nutshell then, there is a correspondence between valence and degree. In this paper, I will connect that observation to adverbs of degree. I will seek to use Goldilocks to explain the degree of intensification that certain adverbs can bring about. In particular, Goldilocks explains why adverbs derived from negative valence adjectives tend to express high or extreme degree, while degree adverbs derived from more positive words express medium degree. Take for example (3) and (4):

(3) Scarlett is pretty tall.
(4) Scarlett is terribly tall.

There is a clear difference in the effect that the adverbs bring about in these two sentences. Scarlett is said to be quite a bit taller in (4) than in (3). As I will explain, this seems to be a general tendency, one that to my knowledge has received all but no attention in the literature.[1]

Given Goldilocks, the contrast between (3) and (4) makes intuitive sense. To be tall in a *terrible* way means to be excessively tall. To be tall in a positive way means that there is no excess involved. The remainder of this paper will makes this idea more precise. To start, however, I will try and establish how widespread contrasts like (3) and (4) are in English.

2 Valence and Scalar extent

When an intensifying adverb (Bolinger (1972)) modifies some gradable predicate, it expresses the extent to which that predicate holds. I will refer to this semantic effect as the *scalar extent*.[2] For instance, *very* expresses high scalar extent, while *pretty* expresses a somewhat lower extent, etc. The suggestion I made above is that scalar extent is tied to valence. The idea is that intensifiers based on positive words intensify to a lesser degree (are associated with lower scalar extent) than intensifiers based on negative words. This idea is supported in English by adverbs of what I will call *medium* degree like *fairly*, *pretty*, *reasonably* –all, arguably, of positive valence– and adverbs of *high* degree like *terribly*, *terrifically*, *awfully*, etc. –all, arguably, of negative valence. A quick look at German and Dutch yields similar intuitions. There, too, adverbs indicating medium degree typically derive from positive valence adjectives, while H-adverbs derive from negative adjectives. For instance:

[1] The only exception I know of is a short passage Rissanen (2008).

[2] This is not to say that the semantics of such adverbs necessarily *is* their scalar extent. Rather, scalar extent is a shallow abstraction of the semantic effect of applying the adverb. For instance, one may adopt the influential proposal from Wheeler (1972) on the semantics of *very*, where *being a very tall man* is given the semantics paraphrasable as *being very tall for a tall man*. The effect of this semantics will be that the scalar extent of *very* is high degree.

	Dutch	German
high degree (*negative valence*)	*verschrikkelijk* (*terrible*) *erg* (*bad*) *ontzettend* (*disrupting*)	*schrecklich* (*terrible*) *sehr* (*painful*) *fürchterlich* (*scarily*)
medium degree (*positive valence*)	*aardig* (*nice*) *best* (*best*) *tamelijk* (*fitting*)	*recht* (*right / just*) *leidlich* (*tolerable*) *ziemlich* (*fitting*)

It would be good to be able to get some sense of how general a pattern this is. To this end, I conducted a small study in which I compared measures of scalar extent to measures of valence associated to adverbs of degree. I hand-selected 24 adverbs that are derived from an evaluative adjective, avoiding adverbs with specialized functions, such as *completely* (end-of-scale) or *nearly* (approximation). I used Amazon Mechanical Turk to ask 61 participants to associate modifications of the adjective *warm* to temperatures. Particpants saw sentences of the form in (5) and were asked to indicate what temperature they thought it was, given that this sentence was produced on a spring day in New York City.

(5) It's ADVERB warm today.

Participants had to provide the temperature they associated to the stimuli in degrees Fahrenheit by moving a slider to the desired position. The slider scale ranged from -10 to +110 degrees Fahrenheit. The aforementioned 24 adverbs took the place of ADVERB in (5) and there was an additional condition in which the adjective was left unmodified. Another manipulation was that the 25 sentences that are thus constructed were also offered to the participants with a negation added into them: *It isn't ADVERB warm today*. All participants saw all 50 stimuli (corresponding to 50 conditions), which were presented in randomised order. There were no fillers.

I centered responses by converting each response of a participant to the distance to the mean of all responses by that participant. I removed all observations that were more than 2 standard deviations removed from the mean response per adverb condition. This removed exactly 100 of the 1525 observations.

The condition with negation was intended to test pragmatic strengthening effects of degree adverbs. Since no reliable differences were found for this condition, it is omitted from subsequent discussion.

The results of the positive condition in this survey give a rough indication of the scalar extent that adverbs are associated with.[3]

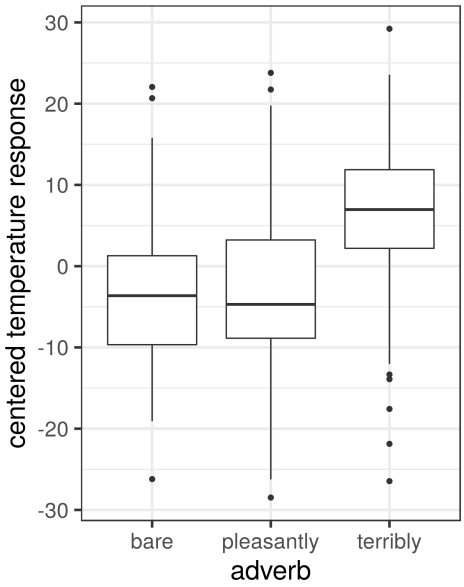

Figure 3. Example of the results for three conditions in study 2.

[3]I say "rough" because the reliability of the responses is somewhat questionable, given the artificial

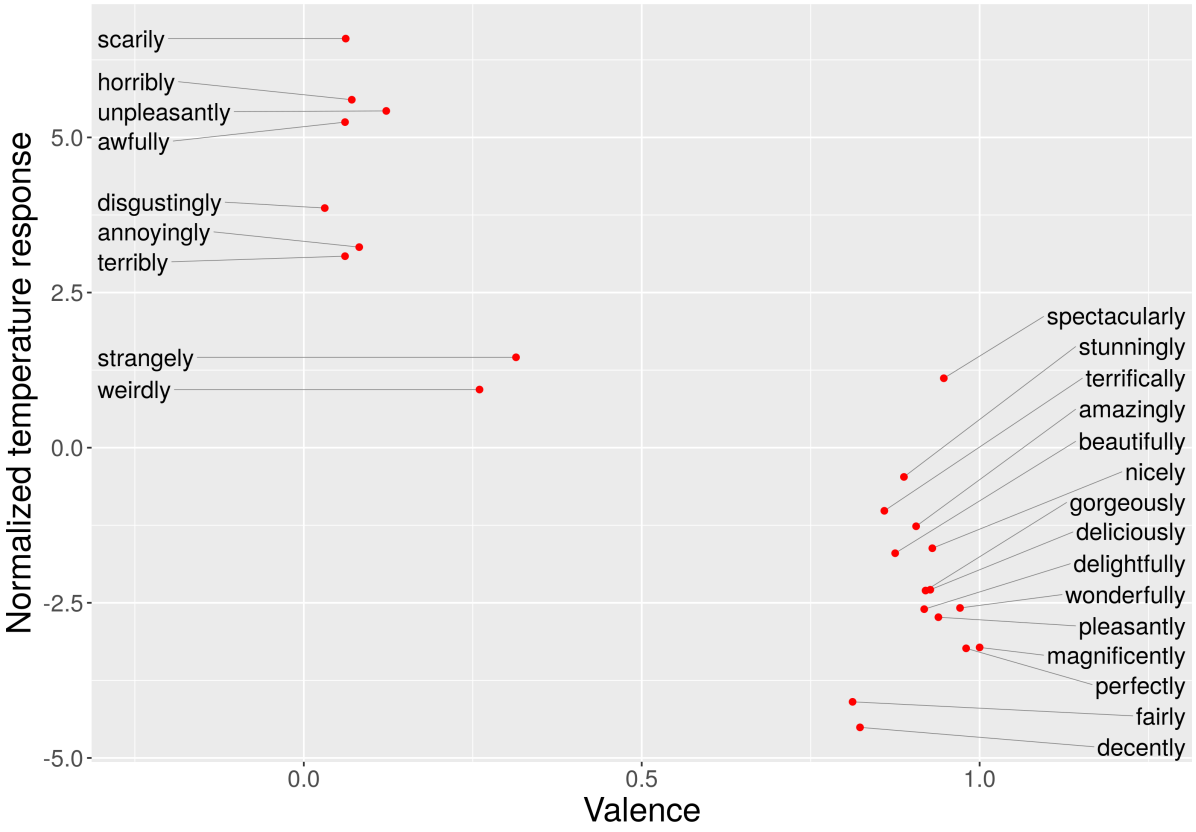

Figure 1: Mean centered temperature response versus valence

See figure 3 for an example. The next step in this study is to connect these results to valence. To this end, I used the NRC valence-arousal-dominance lexicon (Mohammad (2018)) to obtain valence scores for the adjectival stem of the adverbs used in the mechanical turk survey.[4] In the NRC lexicon, valence is quantified on a 0 to 1 scale: 0 being the extreme negative valence and 1 being the extreme positive one.

Figure 4 plots the mean centered temperature response from the mechanical turk survey against the NRC VAD valence score of the adverbs. As can be seen, there is a clear correlation between the two. The higher the valence of the underlying adjectival predicate of an adverb, the lower the scalar extent that this adverb tends to express.

What is also clear from this plot is that the adverbs form two groups. In fact, what drives the correlation seems to be the fact that most adverbs have a valence close to either 1 or 0. The former tend to express relatively low degrees, the latter relatively high degrees. And, so, we see the adverbs of high degree on the left and the adverbs of medium degree on the right. This shows that scalar extent is to a large part determined by valence, as would be expected from the Goldilocks connection between valence and excess.

nature of the task. However, despite this, this kind of setup has been proven to be useful. See, for instance, Bennett and Goodman (2018) for a study involving a similar methodology.

[4]The NRC VAD lexicon is available at: https://saifmohammad.com/WebPages/nrc-vad.html. This lexicon was built by asking annotators to rank four adjectives according to the relevant property (e.g. valence). The results is a reliably consistent annotation. See Mohammad (2018) for details on the methodology.

3 A closer look at scalar extent

Languages tend to come with a diverse array of degree-modifying adverbs. These have specialized roles like expressing approximation (*almost*, *barely*), proportion (*completely*, *totally*, *half*), or sufficiency and excess (*enough*, *so*, *too*). In what follows, however, my main concern will be an often-made distinction between three kinds of adverbs of degree, that tracks scalar extent:

- Adverbs of low scalar extent, or *L-adverbs*: e.g. *a bit*, *somewhat*, *slightly*
- Adverbs of medium scalar extent, or *M-adverbs*: e.g. *fairly*, *pretty*, *quite*
- Adverbs of high scalar extent, or *H-adverbs*: e.g. *very*, *terribly*, *extremely*

Above, I showed some empirical support for this distinction between medium and high scalar extent modifiers. Not only did the sample of adverbs I looked at split in two distinct groups in terms of scalar extent, this division coincided with a split in valence of the underlying adjective.

Adverb typologies like the one above are extremely frequent in the literature on degree. An early influential version of this way of thinking of degree adverbs originates in Stoffel (1901), who made a binary distinction between *intensives* (roughly, H-adverbs) and *downtoners* (roughly, M- and L-adverbs). A much more commonly seen way of dividing things up in the past few decades uses the three-way distinction into low, medium and high degree (Biedermann (1969), Bolinger (1972), Bäcklund (1973), Gary (1979), Van Os (1989), Klein (1998), Paradis (1997)).[5] Such classifications are standardly made on the basis of scalar extent alone. Given that, the question arises what motivates a three-partite division. Why would we divide things up in *low*, *middle* and *high*, as if we have to model our adverbs in parallel to the breakfast bowls of Goldilocks' three bears? Why not a continuum? Classification only makes sense, if the type of scalar extent of an adverb coincides with some independent property these adverbs classes have.[6]

Of course, if the distinction between medium and high scalar extent is tracked by the valence of the predicate underlying the adverb, as I suggested above, then this will give us a meaningful classification for M- and H-adverbs. Ultimately, however, I will question the value of this way of thinking about scalar extent. Before I do so, I turn to L-adverbs

[5] In fact, Klein follows van Os in assuming a separate category for degree modifiers expressing *extreme* high degrees. Note, furthermore, that the terminology used here is mine. I introduce these new terms mainly to avoid a fair amount of terminological confusion that exists in the literature. More traditional names include for instance *boosters* for H-adverbs, *moderators* or *compromisers* for M-adverbs and *diminishers* or *minimizers* for L-adverbs.

[6] Some years ago, I suggested that negation could play a role in distinguishing M-adverbs from H-adverbs. The former are awkward with negation. A sentences like (17) us only felicitous when used as a denial of an earlier claim that the towel in question was *pretty wet*. In contrast, H-adverbs are not sensitive to polarity. In fact, they tend to give rise to pragmatically strengthened readings under negation. A sentence like (18) isn't just interpreted as conveying the negation of the towel being wet to a high degree. It instead tends to be given a strong reading: the towel in question is considerably dry. There are quite a few adverbs, however, that appear to have high scalar extent, but which do not show pragmatic strengthening under negation, as for instance is the case for (19).

(17) ?This towel isn't pretty wet.
(18) This towel isn't very wet.
(19) John's essay wasn't shockingly good.

Cases like (19) make me conclude that polarity sensitivity is not neatly linked to scalar extent.

and show that it is possible to get support for the full three-partite division using the idea that negative valence is connected to excess.

3.1 L-adverbs

It is well known that L-adverbs are sensitive to scale structure (e.g., Kennedy and McNally (2005)). In particular, they only combine with adjectives associated to a lower-closed scale. For instance, *a bit* is felicitous with an adjective like *wet*, since the scale of *wetness* is lower bounded: it starts at completely dry and then involves increasing amounts of liquid. It is infelicitous with adjectives that lack a natural lower bound, such as for instance *nice*.

(6) This towel is a bit wet.
(7) *This man is a bit nice.

In contrast to L-adverbs, M-adverbs and H-adverbs are not sensitive to scale structure:

(8) This man is pretty nice.
(9) This man is very nice.

This would give us a way of predicting low scalar extent: adverbs compatible only with lower-bounded scales tend to express low scalar extent. Unfortunately, this particular way of distinguishing L-adverbs from M- and H-adverbs is not perfect. While most informants would agree there is a clear difference between (6) and (7), examples such as (10) are not so clearly infelicitous.

(10) ?The towel is a bit thick.

Examples like (10) improve when an excessive reading is plausible (Paradis (1997): 76). That is, when acceptable, (10) tends to be interpreted as the towel being (slightly) too thick. This is generally the case. While at first sight, a combination like *a bit tall* may raise eyebrows, given the right context (namely, one involving excess) it becomes felicitous. Imagine, for instance, a context where two movie casting directors are trying to find someone who could play some historical figure who happened to be quite short. One of them suggests Scarlett Johanson. The other replies:

(11) Not sure. I think Scarlett Johanson is a bit tall.

and thus communicates that they are of the opinion that she may be too tall to play a short character. To make sense of this, first notice that excess is itself a lower-bounded scale, witness the acceptability of L-adverbs with the overt excess marker *too*, as in (12).

(12) This towel is a bit too thick.
(13) *This towel is a bit thick enough.

In appendix A, I provide a formal backing for why excess creates such a lower bound. For now, what matters is simply the fact that we made parallel observations: (i) L-adverbs combine only with adjectives associated to lower-bounded scales; (ii) excess is such a lower-bounded scale; (iii) combinations of L-adverbs and adjectives that are not associated

to lower-bounded scales can be felicitous when interpreted as expressing excess, and, as such, when interpreted w.r.t. a lower-bounded scale.

There's another common observation about the distribution of L-adverbs that can be explained in a similar way. L-adverbs tend to combine more easily with negative valence adjectives than with adjectives with positive valence (e.g. Bylinina and Zadorozhny (2012), Jaspers, Craenenbroeck, and Wyngaerd (2016), Nouwen (2018)). Consider (14) - (16):

(14) That's slightly odd / ?normal.

(15) He's a bit tired / ?awake.

(16) My cat's a bit sick / ?healthy.

This observation, too, makes sense once we see negative valence adjectives as adjectives involving excess. -adverbs are felicitous with negative valence adjective and with expressions of excess. My reasoning about this observation is as follows: (i) L-adverbs have the distinguishing feature that they require bottom-closed scales; (ii) since negative valence involves excess and excess involves a scale minimum, L-adverbs are compatible with negative valence adjectives. As a result, I make the predict that L-adverbs should co-occur more with negative valence adjectives than M/H-adverbs, since the latter lack the scalar sensitivity that L-adverbs have, or the easier it is for an adjective to receive a negative, i.e. excessive, interpretation, the more often we should see that adjective pair with an L-adverb.

I conducted a corpus study to test this prediction. For this study, I used the hotel review dataset (Ganesan and Zhai (2012)), which consists of roughly 259k English language hotel reviews.[7] I extracted co-occurrences of a set of degree-related expressions (L-, M- and H-adverbs) with an adjective.[8] Subsequently, I used the NRC valence-arousal-dominance lexicon (Mohammad (2018), see above) to obtain valence scores for the adjectives in the hotel review data.

Using the corpus and the lexicon, I then compared the sets of adjectives that different degree expressions combined with. The prediction is that the mean valence of adjectives that L-adverbs combine with is lower than the mean valence of adjectives that combine with degree expressions that are not constrained to combine with lower-closed scale adjectives. For instance, combinations with degree expressions that indicate medium or high degree, which I introduced above as M-adverbs and H-adverbs, respectively, are predicted to involve higher valence adjectives.

The results are in figure 2. As can be clearly seen, L-adverbs tend to co-occur with more negative adjectives, when compared to M-adverbs and H-adverbs. These latter two classes differ significantly from L-adverbs ($t=11.79$, $p<0.001$ and $t=15.44$, $p<0.001$, respectively).[9]

[7] The dataset is available at: http://kavita-ganesan.com/entity-ranking-data/

[8] This set consisted of: *a bit, slightly, somewhat, very, so, really, too, quite, as, more, extremely, pretty, little, most, rather, absolutely, rather, incredibly, relatively, perfectly, completely*. This choice was based on the desire to have a wide range of different kinds of degree expressions and was further constrained by demanding significant frequency.

[9] Figure 2 also suggests that *rather* has a relative preference for combinations with low valence words. This is, in fact, in line with a very early intuition in Stoffel (1901): "Suppose a man under treatment at an hydropathic establishment, about to take his morning-bath, were to put his hand into the water, and say to the attendant: «The water's rather cold this morning», then he would be expressing unpleasant

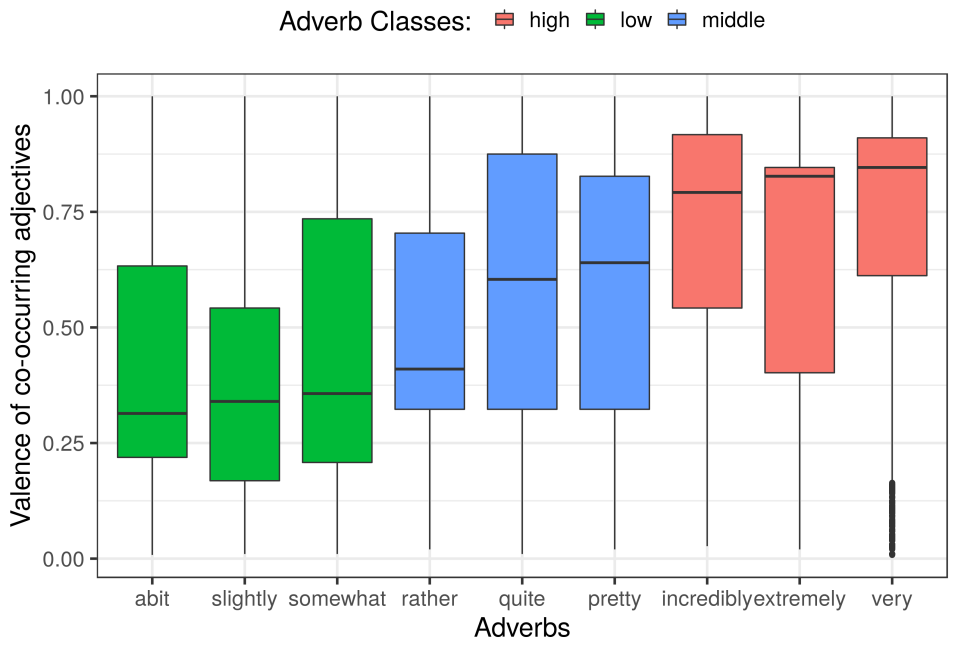

Figure 2: Mean valence of adjectives combining with degree expressions.

3.2 Bleaching

So far, I have shown that the Goldilocks way of thinking about valence and degree helps to distinguish classes of adverbs beyond just the scalar extent they are associated with. Adverbs with low scalar extent are only compatible with lower-bounded scales. Since negative valence adjectives express excess and since excess is lower-bounded, they are also sensitive to the valence of the adjective they combine with. M- and H-adverbs show no such sensitivity. Yet, they can be distinguished by the valence of their adjectival root. M-adverbs express medium degree because the underlying predicate is positive in valence. H-adverbs tend to be derived from negative predicates.

While I think that this way of thinking about adverbial degree modification is insightful, I should hasten to add that it may not be so generally applicable. In fact, ultimately, I think that scalar extent classification of adverbs is ill-conceived. This is because degree adverb typologies like the one typically adopted in the literature assume that adverbs of degree come with some (lexically) fixed expression of scalar extent. This, I will argue now, is far too simplistic. There are many intensifiers that have variable – in fact, context-dependent – scalar extent. Take (20):

(20) It is surprisingly warm today.

Without specifying a context, one may imagine this sentence being said about a day that is much warmer than usual, so much warmer in fact that the speaker is surprised. But it would be wrong to assume that this means that *surprisingly* is an adverb that indicates high degree. Imagine a different context in which the weather models predicted today to

surprise at finding the water colder than he had expected. And if the attendant were to answer: «Yes, sir, it's *pretty* cold, but then it's the doctor's oders», he would merely make an objective statement as to the temperature of the water." (page 132).

be extremely cold, much colder than normal for the time of year. It turns out that the models, which are usually very reliable, were wrong. It *is* much colder than normal for the time of year, but it is still quite a bit warmer than the models predicted. In such a context, (20) makes sense. However, the adverb clearly cannot be said to express high degree.

A similar example is (21). This expresses merely that it warmer than I hoped it would: (21) does not entail that it was very warm. Think of a context: my 7-year old son hopes that the weather will be extremely cold, for only when the temperature reaches -10C will the school be closed. It turns out that it is actually very cold, but unfortunately for my son, it is -8C. In such a context, it is felicitous for my son to assert (21).

(21) It is disappointingly warm today.

This shows that it is important to distinguish intensifiers like *very*, that specialize in expressing high degree, from intensifiers like *disappointingly* and *surprisingly* that fulfill their degree function indirectly. Another way of saying this is that *very* is *bleached*. All its meaning is directed at affecting degree. In contrast, examples like (20) and (21) do not just express how warm it is, but also that the temperature is surprising / disappointing.

A good example of a bleached intensifier in English is *terribly*. In contrast to *very*, which has lost its original meaning of *true / real* centuries ago, for *terribly* it is very clear what its non-degree meaning would be – afterall, the stem *terrible* is in use and has nothing to do with degree. But whenever we use *terribly* as an adverb of degree, the meaning of this stem is completely absent from the resulting interpretation. For instance, there is no contradiction in (22) and nothing bad ends up being said about the man in question. In clear contrast, in (23) the man is said to have two contrasting (contradictory?) properties. The example in (24), in which an adverb that has not been bleached, *disgustingly*, modifies *nice*, equally conveys a contrast or even a contradiction.

(22) He is terribly nice.

(23) That man is both nice and terrible.

(24) He is disgustingly nice.

In sum, the fact that not all adverbs of degree are bleached – adverbs like *disgustingly*, *disappointingly*, *surprisingly* and many others are what I will call *unbleached* – makes it impossible to characterize intensifiers simply in terms of their *scalar extent*. Unbleached intensifiers do not straightforwardly express such an extent. Instead, they convey the property expressed by their adjectival stem and modify degree relative to that property.

Although I won't have anything deep to say about the diachronic process of bleaching, it seems evident to me that at least some part of that process is simply that bleached adverbs end up being associated to the scalar extent their initial unbleached version was regularly associated with. If an unbleached adverb, through the lexical content of its adjectival stem, ends up typically expressing high degree, then the unmediated expression of high degree is a natural candidate for the meaning of a bleached version of this adverb. For instance, the fact that *terribly* now expresses high degree suggests that it was typically associated to high degree when it was still unbleached.

For this to make sense, however, it needs to be the case that the adjectival predicate underlying an adverb of degree has some role to play in what scalar extent is typically

expressed by the adverb. The link between Goldilocks and evaluation that I developed above predicts that whenever the adverb is associated with some valence, the scalar extent of the adverb will tend to be in line with this value: medium degrees tend to be *good*, extreme degrees (because they involve excess) tend to be *bad*.

3.3 Formal considerations

How does the unbleached content of adverbs interact with the degree semantics of adjectives? I will not be able to provide a definitive formal account of the semantics of combinations of evaluative adverbs and adjectives, but in the following I will sketch what I think a formal prosopoal would need to look like.

Wheeler (1972) proposes that unbleached degree modifiers are best analyzed as factive propositional operators. On his analysis, (25) is interpreted as *it is horrible that it is as warm as it is*. Morzycki (2008) argues against such an analysis on the basis that it would wrongly predict that the weather can be described as *horribly warm* also when it is very cold. On a freezingly cold day, most people would agree that it is indeed horrible that it is as warm as it is, but no-one would agree that it is *horribly warm*.

(25) Today is horribly warm.

Morzycki solves this by stipulating that evaluative intensifiers express extreme degrees: *horribly warm* means that it is horrible how extremely warm it is. This analysis in turn predicts that evaluative adverbs always express high degree. As I explained above, this is not the case. It can be *disappointingly warm* without being warm to a high degree. Moreover, as we have seen above, positive valence adverbs tend to express medium rather than high degree, as predicted by the Goldilocks approach. Morzycki's complaint about a Wheeler-style analysis is valid, yet the explicit encoding of a particular scalar extent in the lexical content fails to account for the way the content of the adverbs' underlying predicate determines the scalar extent. Wheeler's analysis does lay such a connection. For that reason, I will now develop an account that builds on Wheeler's idea without running into the objection raised by Morzycki. Before I can do so, let me introduce a couple of assumptions I make.

Let's assume a standard degree-based semantics of adjectives. I assume that adjectives are associated to measure functions, which I will generally write as μ. For instance, the adjective *tall* is associated to the measure function μ_{tall}, which takes an individual and returns a *degree* corresponding to that individual's height.

I follow Kennedy (2007) in assuming that adjectives in their positive form are interpreted as exceeding a threshold θ. That is, *Sue is tall* corresponds to:

$$\mu_{\text{tall}}(\text{Sue}) > \theta$$

The value of θ is such that subjects whose measurement exceeds the threshold *stand out* (Fara (2000)). The standard way of thinking about *standing out* is extensional in nature: the tallest child in her class is *tall* because she literally *stands out* when compared to her peers. Standing out can also be thought of in intensional ways, though (e.g. Fara (2000), Bylinina (2014)). When I decide that the queue in the supermarket that I am in is *long*, it doesn't necessarily mean that it stands out when compared to the length of neighbouring

queues. What I could mean is simply that the queue is significantly longer than what it normally is like, or than what I hoped it to be, etc. In general, all sorts of considerations may influence what in a certain context does and does not stand out.

Standing out is the interpretation given to an adjective's positive form. But what happens when that positive form adjective is modified? It is not uncommon to think that modifiers may influence how the positive form is interpreted, that is, how the subject is said to be standing out. For instance, *for* phrases could be seen as specifying the comparison class used to determine θ (Solt 2011). A sentence like *Sue is tall for a 4 year old* conveys that $\mu_{\text{tall}}(\text{Sue}) > \theta$ in such a way that Sue stands out when compared to 4 year olds.[10] Wheeler's (1972) analysis of *very* can be seen in the same spirit. According to that proposal, *Sue is very tall* is interpreted in just the same kind of way as *Sue is tall* is, except that *very* settles the comparison class to be made up of individuals classified as *tall*. A similar idea is implemented in a probabilistic framework in Bennett and Goodman (2018). Bennett and Goodman use the Bayesian approach to positive form interpretation of Lassiter and Goodman (2017), which could be seen as a quantitatively explicit model of how *standing out* determines a θ threshold. Bennett and Goodman take intensifying adverbs to not contribute any actual meaning. Instead, they merely add to the cost of making the utterance. The pragmatic effect of that added cost is a narrower usage, corresponding to higher degrees of the property expressed by the adjective. Once more then, the modifier is seen as not altering the *way* the adjective is interpreted, it merely fills in a parameter of that interpretation.

For the cases we have looked at above, Bennett and Goodman's claim that the lexical content of an intensifier is void does not make much sense. As I have shown above, the evaluative predicate that forms the stem of an unbleached intensifying adverb has a clear semantic effect. Moreover, this predicate (and in particular, its valence) also determines the scalar extent of the intensifier, even when the adverb is bleached. Nevertheless, I think that the idea that modifications of positive form adjectives are often modifications of the *standing out* interpretation is correct for evaluative intensifiers, too. In fact, I think that the reason why *horrifically warm* means that it is *warmer* and not *colder* than what is normally considered to be pleasant is exactly because something is only *horrifically warm* when it stands out w.r.t. how warm it is: $\mu_{\text{warm}}(x) > \theta$. Extremely cold days *are* horrible, but they do not stand out on the warm scale, since they do not exceed θ.

To implement this, I will first implement Wheeler's semantics. I take the proposition in (27) to express what Wheeler had in mind for (26).

(26) Today is horrifically / pleasantly warm.

(27) $\mu_{\text{horrifically/pleasantly}}(\lambda w.\mu_{\text{warm}}(\text{today})(w) = \mu_{\text{warm}}(\text{today})(@)) > \theta'$

I provide the measure function for *warm* here with a world argument, so that I can express the proposition *that today is as warm as it is*, coloured red in the formula. (I omit the world argument in the adverb's measure function to keep things simple.) Note that (27) is a positive form interpretation of the adverb: that today is as warm as it is stands out w.r.t. how horrific / pleasant that is.

My proposal is now that (27) is only part of the interpretation of (26) and that the adjective still receives a positive form interpretation – which it does not in (27). That

[10] This does not suffice to characterize the sentence's meaning, since the sentence also conveys that Sue *is* a 4 year old. See Solt (2011) for discussion.

is, *Today is horrifically / pleasantly warm* is simply interpreted as (28), where θ is to be determined in part on the basis of the truth of (27). In other words, an evaluative intensifier contributes a proposition like (27) and by doing so influences the way in which the subject is said to be *standing out*.

(28) $\mu_{\text{warm}}(\text{today}) > \theta$

I will leave open for now how this needs to be implemented exactly (and compositionally), but one way to think about the proposal is that the intensifier restricts the domain under consideration. Take *Today is horrifically warm* as an example. The sentence conveys both that (i) it is horrific that it is as warm as it is and (ii) that today stands out among days that satisfy (i). What kind of day does this sentence describe? According to (i), the day in question is either very cold or very warm. In other words, we are looking at days with extreme temperatures. Clearly, the extremely warms days stand out from the extremely cold ones. Since in positive form interpretation standing out is interpreted as *exceeding* a threshold, it is only the extremely warm days that can be described to be *horrifically warm*. In contrast, to be *horrifically cold* is to be extremely cold, since extremely cold days stand out because the measure of how cold they are exceeds a threshold.

An M-adverb works in the same way. *Today is pleasantly warm* conveys that (i) it is pleasant that it is as warm as it is and (ii) that today stands out among the days that satisfy (i). The result is that a pleasantly warm day is a day on the warm end of days that are pleasant because of the temperature.

What is attractive about this interpretation is that there is no direct manipulation of the degree to which it is warm. That is, the scalar extent attributed to *pleasantly* and *horribly* comes about entirely by an inference based on the fact that our horror/pleasure at the current temperature *stands out*. As such, this approach is capable of accounting for the observationis we made above. Scalar extent is tied to the content of the adverb and in particular to the valence that its underlying predicate expresses.

4 Conclusion

I have shown that valence, the polar value expressed by an evaluative act, is relevant for degree semantics. I have given two examples of this. On the one hand, the fact that negative valence involves excess makes it that negative valence predicates can turn into adverbs of high degree, while adverbs formed by positive valence predicates express medium scalar extent. On the other hand, I showed that negative valence adjectives inherit a main scalar property of excess in being lower-bounded.

Valence is not often considered in formal linguistics. This is probably because other notions of polarity seem to impact more straightforwardly on semantics, such as the presence or absence of a negation marker (*not*, *un-*, etc.) or lexical antonymity. In contrast to such polar phenomena, valence is a much more illusive notion. The valence of an adjective is not morpho-syntactically determined. I have given examples of adjectives like *warm* and *tall* that illustrate that they can receive both negative and positive readings, but even cases that appear to be more straightforward, like for instance *nice*, turn out to be quite flexible. An example like (29) will readily be interpreted as saying something bad about John.

(29) John is a bit nice.

What I hope to have shown is that behind the slippery notion of valence, there is a straightforward semantic mechanism, one that directly affects degree.

Acknowledgements

For discussion of and comments to earlier presentations of ideas in this work, I am grateful to Dominique Blok, Lisa Bylinina, Jakub Dotlacil, Nicole Gotzner, Larry Horn, Stephanie Solt, Carla Umbach, Cameron Wilson, Jan Winkowski and Guido vanden Wyngaerd.

5 References

Bäcklund, Ulf. 1973. *The Collocation of Adverbs of Degree in English*. Acta Universitatis Upsaliensis.

Bennett, Erin D, and Noah D Goodman. 2018. "Extremely Costly Intensifiers Are Stronger Than Quite Costly Ones." *Cognition* 178: 147–61.

Biedermann, Reinhard. 1969. *Die Deutschen Gradadverbien*. Universität Heidelberg.

Bolinger, Dwight. 1972. *Degree Words*. Den Haag: Mouton.

Bylinina, Lisa. 2014. *The Grammar of Standards: Judge-Dependence, Purpose-Relativity, and Comparison Classes in Degree Constructions*. Netherlands Graduate School of Linguistics.

Bylinina, Lisa, and Stas Zadorozhny. 2012. "Evaluative Adjectives, Scale Structure, and Ways of Being Polite." In *Logic, Language and Meaning*, 133–42. Springer.

Fara, D. Graff. 2000. "Shifting Sands: An Interest-Relative Theory of Vagueness." *Philosophical Topics* 20: 45–81.

Ganesan, Kavita, and ChengXiang Zhai. 2012. "Opinion-Based Entity Ranking." *Information Retrieval* 15 (2): 116–50.

Gary, Edward Norman. 1979. "Extent in English: A Unified Account of Degree and Quantity." PhD thesis, University of California, Los Angeles.

Jaspers, Dany, Jeroen van Craenenbroeck, and Guido Vanden Wyngaerd. 2016. "De Syntaxis En Semantiek van Diminishers in Het Nederlands En Het Duits." *Leuvense Bijdragen: Tijdschrift Voor Germaanse Filologie* 99-100: 87–108.

Kasting, James, Owen Toon, and James Pollack. 1988. "How Climate Evolved on the Terrestrial Planets." *Scientific American*, no. 2: 90–97.

Kennedy, Chris. 2007. "Vagueness and Grammar: The Semantics of Relative and Absolute Gradable Predicates." *Linguistics and Philosophy* 30 (1): 1–45.

Kennedy, Christopher, and Louse McNally. 2005. "Scale Structure, Degree Modification and the Semantics of Gradable Predicates." *Language* 81 (2): 345–81.

Klein, Henny. 1998. "Adverbs of Degree in Dutch and Related Languages." PhD thesis, Rijksuniversiteit Groningen; John Benjamins.

Lassiter, Daniel, and Noah Goodman. 2017. "Adjectival Vagueness in a Bayesian Model of Interpretation." *Synthese* 194 (10): 1–36.

Mohammad, Saif M. 2018. "Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words." In *Proceedings of the Annual Conference of the Association for Computational Linguistics (Acl)*. Melbourne, Australia.

Morzycki, Marcin. 2008. "Adverbial Modification of Adjectives: Evaluatives and a Little Beyond." In *Event Structures in Linguistic Form and Interpretation*, edited by Johannes Dölliing and Tatjana Heyde-Zybatow. Berlin: Mouton de Gruyter.

Nouwen, Rick. 2018. "A Note on Positive and Negative Evaluative Adjectives." In *A Coat of Many Colours*, edited by Ceupens et al. https://dj60.be/.

Paradis, Carita. 1997. *Degree Modifiers of Adjectives in Spoken British English*. Lund Studies in English 92. Lund University Press.

Rissanen, Matti. 2008. "From 'Quickly'to 'Fairly': On the History of Rather." *English Language & Linguistics* 12 (2): 345–59.

Solt, Stephanie. 2011. "Notes on the Comparison Class." In *Vagueness in Communication*, edited by Rick Nouwen, Robert van Rooij, Uli Sauerland, and Hans-Christian Schmitz. Berlin: Springer. https://doi.org/10.1007/978-3-642-18446-8_11.

Stechow, Arnim von, Sveta Krasikova, and Doris Penka. 2004. "The Meaning of German Um Zu: Necessary Condition and Enough/Too."

Stoffel, Johannes. 1901. *Intensives and down-Toners*. Heidelberg: Carl Winter.

Stojanovic, Isidora. 2015. "Evaluative Adjectives and Evaluative Uses of Ordinary Adjectives." In *LENS12*, edited by Daisuke Bekki and Eric McCready. Tokyo.

Van Os, Charles. 1989. *Aspekte Der Intensivierung Im Deutschen*. Narr Tübingen.

Wheeler, S.C. 1972. "Attributives and Their Modifiers." *Nous* 6: 310–34.

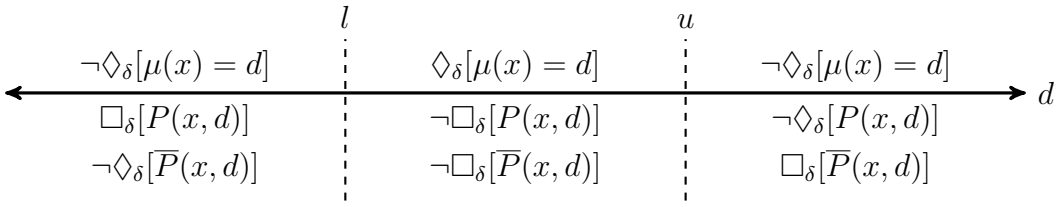

Figure 3: Sufficiency and excess in modal terms (repeated)

6 Appendix: the semantics of excess

As before, I assume a standard monotone degree-based semantics of adjectives, along the lines of (30).

(30) $P(x, d) \Leftrightarrow \mu(x) \geq d$

Let's also assume that we have an antonymic predicate $\overline{P}$, which is associated with the same measure functions, the same scale of degrees, but an inverse ordering. Correspondingly,[11]

(31) $\overline{P}(x, d) \Leftrightarrow \mu(x) \leq d$

I now formalize a notion of excess using simple modal concepts. Let $\Diamond_\delta p$ be true if p is compatible with reaching some desirable goal δ. The dual $\neg\Diamond_\delta\neg p$ is true if and only if δ is unobtainable without p being true, which we write as $\Box_\delta p$. Excess can be defined by applying the goal-modals to measure functions and gradable properties. Let μ and P be respectively the relevant measure functions and corresponding gradable property and let x be the subject we are measuring. Say now, we have a standard Goldilocks situation in which δ is obtainable only whenever $l < \mu(x) < u$. This is depicted in the top half of figure 1. The bottom half depicts the true modal statements w.r.t. the monotonic predication $P(x, d)$, i.e. $\mu(x) \geq d$, and its antonym $\overline{P}(x, d)$.

For x to be in the goldilocks zone, $\mu(x)$ has to exceed the lower bound l. This means that all worlds in which the goal is obtained are worlds in which $\mu(x) \geq d$, for all degrees $d \leq l$. So, $\Box_\delta[P(x, d)]$ is true for all these degrees. If $d > u$, then the goal is unobtainable. So $\neg\Diamond_\delta[P(x, d)]$ holds for such degrees. For the case of $\overline{P}$, things are simply flipped. We can now express excess correspondingly, using P.

$$x \textbf{ has an excess of } P \quad \Leftrightarrow \quad \exists d[\neg\Diamond_\delta[P(x, d)] \wedge P(x, d)]$$

In words, this says that x has an excess of P whenever x has P to at least some degree such that being P to that degree makes the goal unobtainable.[12] Going back to figure 1, if $\mu(x) > l$, then x has an excess of P. If $\mu \leq l$, then x has an excess of $\overline{P}$. And, so, a goldilocks zone is an interval of degrees that are neither too P nor too $\overline{P}$. The semantics above explains the second Goldilocks principle, by maintaining that excess is monotonic.

[11] It follows that $P(x, d) \wedge \overline{P}(x, d) \Leftrightarrow d = \mu(x)$ or in other words $\lambda d.\overline{P}(x, d)$ is $\overline{\lambda d.P(x, d)} \cup \{\mu(x)\}$.

[12] This way of characterizing excess is reminiscent of the proposed semantics for *too* in Stechow, Krasikova, and Penka (2004). Importantly, however, I do not intend the above as a semantics for the degree modifier *too*, but rather as the semantics of the underlying concepts of excess and sufficiency.

If we abstract away from the quantificational force in the propositional semantics of excess above, we can represent *excess* as a gradable property: degrees to which x is excessively P.

$$\lambda d.\lambda x.\neg\Diamond_\delta[P(x,d)] \qquad \text{(excessive } P\text{-ness)}$$

The see what this means, we should first try and understand what these degree predicates look like in a typical situation, like that depicted in figure 3. As can be read from that figure, for any degree $d > u$, we have that $\neg\Diamond_\delta P(x,d)$ and, so, the excessive P-ness degrees start just above u: Any degree higher than u is a degree of excess.[13]

If we look at excess from this perspective, the *scale* of excess is the interval of degrees $(u,\ldots)$. Consequently, excess becomes akin to a *minimum standard* adjective, an adjective that comes with a bottom-closed scale. This explains why excess is compatile with L-adverbs.

Let's assume that adjectives can be coerced to excessive readings, simply by combining them with the modal profile above. And let's assume that this happens through a functional head like (32) below. This is just to be able to discuss the consequences of having such readings available.

(32) $[\![\textsc{Excess}]\!] = \lambda P.\lambda d.\lambda x.P(d)(x) \wedge \neg\Diamond_\delta[P(x,d)]$

The result of all this is simply that "Pos [Excess P]" will be true of x whenever $\mu(x) > u$. The availability of Excess predicts that the distribution of adverbs compatible only with lower-closed scales is broader than exclusively combinations with adjectives that are lexically endowed with such scales, as we have indeed seen in section 3.1. In principle, any adjective could be interpreted in terms of excess, using Excess and, so, in principle, any adjective can be made compatible with adverbs requiring lower-closed scales. This distribution is wider than just adjectives like *wet*, because there is an option of interpreting adjectives in terms of excess. This explains why *Scarlett Johanson is a bit tall*, when felicitous, tends to be interpreted as the actress being too tall.

The current setup also explains why we would find a preference of L-adverbs for combining with negative rather than positive valence evaluative adverbs. If we accept the above theory of how relative adjectives are turned into minimum standard adjectives, then we should predict that this is easier for negative valence predicates than for positive valence predicates. Excess is intrinsically of negative valence and thus easier to connect to adjectives expressing properties we evaluative as being *bad*: *sick*, *nasty*, *tired*. This is not to say that we cannot form excessive readings with positive evaluatives. After all, as the first Goldilocks principle I gave says: you *can* have too much of a good thing. One just has to find a context where excessive amounts of this particular goodness becomes *too much*. Take (15), repeated here as (33) and (34)

(33) He's a bit tired.
(34) ?He's a bit awake.

[13] Note that none of the degrees in $(..l]$ are degrees of excessive P-ness. For all these degrees there exist worlds in which the goal is obtained. This is important, as it accounts for the fact that excessive P-ness is not the same as having an excess w.r.t. the antonym of P. For instance, an extremely low temperature does not count as being *excessively warm*.

It is fairly easy to find an interpretation for (33) that involves excess. In most contexts, being tired is construed as something negative, something that stops you from achieving things, something that inhibits your normal functioning. On the other hand, for (34) to make sense, we need to think of a context in which being more awake is *bad*. This is clearly harder to do. But imagine a heist context in which we trying to steal some extremely expensive diamond. We have drugged the one security guard standing between us and the precious stone. We are debating whether we can make our move yet, but we have our doubts. The drugs have made the guard sleepy, but have not been effective enough to be able to pass the guard unnoticed. In that context, I think (34) improves a great deal.