# Hyperbole is not a rational speech act

Rick Nouwen

April 2026

Hyperbole is a type of exaggeration for rhetorical effect, illustrated by for example (1).

(1) Sue is always late.

An utterance of (1) when used hyperbolically does two things. First of all, it conveys that Sue is often (but not always) late, and, second, it triggers some side effect. For instance, (1) would be a good way for the speaker to indicate that she is exasperated or frustrated with Sue, or that she believes Sue should be fired.

It is generally recognised that hyperbole is a scalar phenomenon and, so, it seems that the standard toolkit of formal, scalar pragmatics should be applicable to understand how hyperbole is interpreted. As a consequence, it is natural to think of hyperbole as a phenomenon that results from rational, cooperative behaviour. That is, just like there is a rational explanation for why in some contexts scalar terms receive a strengthened interpretation (scalar implicature), the scalar nature of hyperbole makes us suspect that a similar rational explanation should be available for the weakening effect involved in hyperbole interpretation. As I will argue, such a theory faces several problems. Ultimately, I argue, it is more promising to approach hyperbole as a phenomenon that does not involve pragmatic reasoning.

## 1 Norm-relativity

As I said, there is consensus that hyperbole is a scalar phenomenon. That is, it is generally recognised that hyperbole involves the mapping of one scalar position to another. For instance, Goodman (1968) writes: "In hyperbole [...] an ordered schema is in effect displaced downward. The large olive becomes supercollosal." In other words, in hyperbole, a term

like “supercollosal” is interpreted lower on the scale than its standard interpretation and ends up meaning (just) “large”.

Feinmann (2023) shows that hyperbole resembles scalar implicature, in that it involves the same kinds of alternatives. For example, (2) implicates that there are not (a lot) more than two hundred unread emails in my inbox, but (3) obviously does not have the same implicature, since it already entails that. Similarly, while (2) has a hyperbolic reading, (3) does not.

(2) There are two hundred unread emails in my inbox.

(3) There are exactly two hundred unread emails in my inbox.

Given the clear scalar nature of hyperbole, it may be tempting to then take the next step and declare hyperbole interpretation as the “weakening” counterpart of the strengthening that happens in scalar implicature. Where implicature involves the denial of an alternative that is stronger than the literal meaning, hyperbole involves replacing the strong literal meaning with a weaker alternative. This is, in fact, the theory that Feinmann entertains in his manuscript. It is important, however, to realise that hyperbole is more constrained than implicature: Not every scalar statement can be weakened, i.e. interpreted hyperbolically.

For instance, while (4) implicates the denial of (5), there is no hyperbolic use of (5) that gets weakened to mean (4).

(4) There are three unread emails in my inbox.

(5) There are four unread emails in my inbox.

This observation is in line with theories that maintain that hyperbole is only possible in quite specific scalar contexts. The idea is that hyperbole is always *norm-relative* (Walton, 2017; Popa-Wyatt, 2020; Nouwen, 2024a). A hyperbolic statement doesn’t just exaggerate some position on a scale in the way suggested in the quote by Goodman, above. Rather, it exaggerates how far that position deviates from the norm. So, we can state that Sue is “always late”, whenever Sue is late more often than expected or desired. By stating that Sue is “always late” we are exaggerating how large the deviation from the norm is. In contrast, since (4) does not express a deviation from the norm, we cannot express it using (5) hyperbolically.

The norm-relativity of hyperbole can be demonstrated clearly by showing how context can steer the direction of the scalar weakening involved. Consider the following two sentences (taken from Nouwen 2024a):

(6) There were a hundred people at your party.

(7) Nobody came to your party.

Consider the following context for these sentences. Scarlett has two friends, Timid and Brag. Both, independently of one another, organised a house warming party not long ago. Both invited 60 people, expecting 30 to come. While 58 people attended Timid's party, only 20 people showed up to Brag's. Timid is insecure and asks Scarlett (present at both parties) whether she thinks the party was a success, suggesting it may not have been. In response, Scarlett can utter (6) hyperbolically: "Of course it was! There were a hundred people at your party!" Brag, on the other hand, is overconfident and is boasting how good the party was. In that context, Scarlett could use (7) hyperbolically: "A success? Are you kidding me? Nobody came to your party!"

The reasoning is as follows: In both cases, there was a deviation from the norm. Timid's party had more guests than the expected 30, Brag's party had fewer. Scarlett is using (6) to exaggerate the deviation in Timid's case, and (7) to exaggerate the deviation (in the other direction) in Brag's case.[1] The upshot is this: Hyperbole interpretation involves scalar weakening, but it cannot be simply reduced to that. In particular, it involves the weakening of how a deviation from the norm is presented.

## 2 Rational repair

Norm-related theories of hyperbole like Walton (2017), Popa-Wyatt (2020) and Nouwen (2024a) tend to be *descriptive* in the sense that they are theories of what constitutes the phenomenon. They do not offer an account of the semantic and pragmatic mechanics behind hyperbolic meanings in the way that Feinmann (2023) for instance does by positing a reverse scalar implicature mechanism. As I explained above, Feinmann's account overgeneralises and, so, we are on the look out for alternative ways of deriving the weakening effect at the heart of hyperbole interpretation.

Quite generally, it is recognised that any such theory should simultaneously account for both the weakening effect and the richness of meaning that is often associated with hyperbole. An utterance of (1) does not just convey that Sue is late often. Rather, the speaker has some secondary

[1] Interestingly, Scarlett could opt to use the other sentence in each scenario as well, but only if she does so ironically. For instance, she could say to Brag: 'O yes, sure, it was a massive success. There were a hundred peple at your party. Not!' And to Timid she could say: 'Right, Timid, what a disaster your party was. Nobody came! Not!'. See Nouwen (2024a) for discussion.

agenda here. She wants to emphasise how bad it is that Sue is not very punctual. Perhaps, by uttering (1), she is highlighting how frustrated she is with Sue's attitude to showing up for work on time.

(1) Sue is always late.

In other words, hyperbolic meaning is (at least) two-dimensional. It triggers an inference about the state of the world, as well as a secondary more subjective inference. In the literature, there have been many theories that concern how these rich meanings come about. In this paper, I focus on one particular set of assumptions that is shared by a proper subset of these theories. I will call these 'theories of *rational repair*'.[2] The relevant key assumptions are summarised as follows:

**Rational repair:**

A. The literal meaning of a hyperbolic utterance is so extreme that it lacks utility and, as a result, it is in need of repair.[3]

B. To repair, the listener interprets the utterance with respect to some other dimension of meaning, where the utterance does have utility.

One example of a rational repair theory is the cognitive linguistic account of Peña-Cervel and Ruiz de Mendoza Ibáñez (2022), who claim that the "counterfactuality" that is part and parcel of a hyperbolic utterance "calls for an understanding of hyperbole in terms of a conceptual mapping involving a reasoning process of the 'treat B as if it were A' kind" (p. 17). In other words, the inherent untruthfulness of the literal meaning of hyperbole forces an interpretation in terms of a mapping from one domain, where the untruthful literal meaning resides, to another.

I won't have much to say about the theory in Peña-Cervel and Ruiz de Mendoza Ibáñez (2022), since it is hard to establish firm predictions from its proposal. Instead, my focus will be on a framework that, in contrast, makes very explicit predictions, namely the computational account presented in Kao et al. (2014), implemented in the rational speech act framework (Scontras et al., 2021). My reasons for this particular focus is that

[2]An example of a type of theory that does *not* make similar assumptions is relevance theory and, in particular, the theory of *Loose Talk* of Sperber and Wilson (1985). In that line of work, hyperbole is not seen as violating any communicative maxims. It is simply a creative way of expressing a complex thought.

[3]For instance, Fogelin (2011) holds that hyperbole involves saying 'something stronger than what I have a right to say, with the intention of having it corrected away from the extreme' (p. 13).

Kao et al. (2014) is the only framework I know that makes quantitative predictions concerning the meaning of hyperbole. All other theories explain why hyperbole exists, sketch what mechanisms are involved in its interpretation, but fail to offer precise testable predictions of the kind I will be scrutinising below.

Importantly, my arguments will not be arguments against the specifics of Kao et al. (2014). Instead, I will show that any implementation of the ideas of rational repair will fail. As such, I will show that if we want to have a theory of hyperbole that is as predictive as that of Kao et al., then we will need to construct that in an altogether different framework.

Kao et al.'s proposal is that hyperbole is a way of addressing two questions under discussion (QUDs) at once. Roughly, QUD1 asks about the state of the world, while the affective QUD2 asks how the answer to QUD1 makes the speaker feel. Since the probability of the answer given to QUD1 is very unlikely to be true, it must be that the utterance addresses QUD2. Since affect tends to be high when prior probability is low, it must be that the answer to QUD2 is that the speaker is somehow affected by whatever the real answer to QUD1 is. The subjective effect of hyperbole, in other words, comes about because the literal content is defective with respect to QUD1, since it expresses a content that is likely to be false.

As is standard in the rational speech act (RSA) framework, a question under discussion is a space of possibilities $S$. The speaker intends the listener to adopt a belief about which of these possibilities is actual. There is a prior probability distribution over this space, which expresses what is contextually expected about the QUD. RSA offers a theory of interpretation in terms of a Bayesian update of that prior, given some utterance event. For some $s \in S$:

(8) $\quad P(s|\mathit{utterance} = U) \propto P(s)P(\mathit{utterance} = U|s)$

This says that the interpretation of $U$ is given by the Bayesian update $U$ can bring about, relative to the prior expectations concerning the question under discussion. A crucial idea in RSA is that the likelihood term in this Bayesian update is a model of the speaker: the likelihood of the speaker uttering $U$, if her intention was to refer to possibility $s$.

The innovation in Kao et al. (2014) is that there is a second space of possibilities that plays a role in considerations of production and interpretation. This is the space of affective possibilities: whether or not the speaker is 'affected'. Call this space $A$ = {affected, not affected}. This space also comes with a prior, which is linked to the other space of possibilities. For each $s \in S$, this prior specifies $P(A = \text{not affected} \mid s)$ and $P(A = \text{affected}$

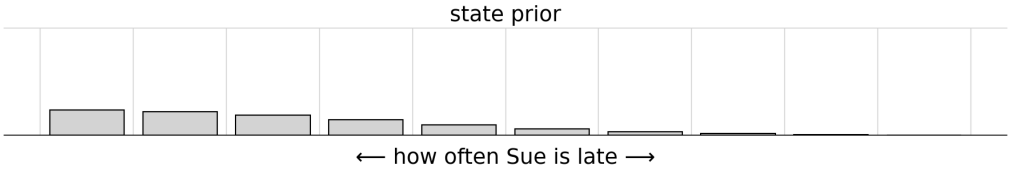
$| s)$.

So, for $s \in S$ and $a \in A$ we get:

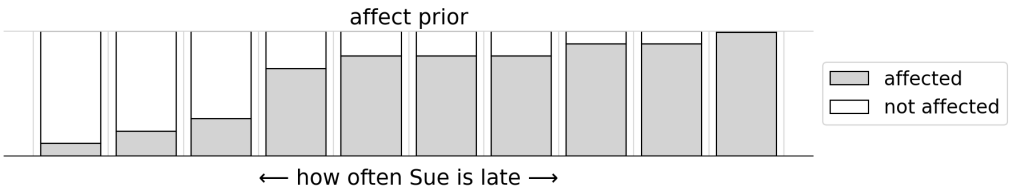
(9) $P(s, a|\mathit{utterance} = U) \propto P(s)P(a|s)P(\mathit{utterance} = U|s, a)$

To understand what this does, we should look closer at the likelihood term. This is the likelihood that a speaker utters $U$, if she wants to refer to possibility $s$ and/or indicate that her affective state is $a$. Such likelihood is calculated on the basis of the speaker's utility for choosing a certain utterance. For example, in the original RSA setup in (8), the utility of choosing an utterance $U$ that is false in $s$ to refer to $s$ is extremely low, since it is unlikely that a hearer will recognise the speaker's intention in that case. Things are different for (9), however. Now, saying something that is false may have higher utility, if the speaker's intention is to point at her affective state.

Let us now take example (1) and assume that we have a QUD 'How often is Sue late?' with prior expectations for that prior as follows:



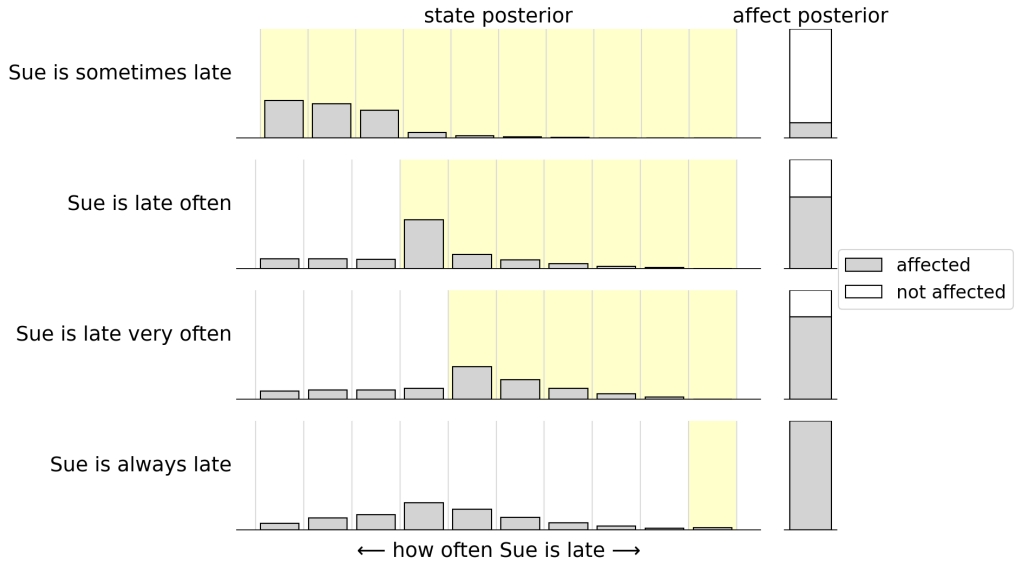

This shows a scale that I've divided into 10 discrete possibilities. On the extreme right is the possibility where Sue is always late. The probability that this possibility is actual is expected to be virtually 0. Towards the left, Sue is more and more often on time. So, according to this prior, the context is such that being late every now and then is expected, while being late (literally) all the time is something that would be extremely unexpected. Connected to this, let us say that we have the following affect priors $P(A = \text{affected}|S = \ldots)$, $P(A = \text{not affected}|S = \ldots)$:




This says that the probability of the speaker being affected by the state of affairs is high when Sue is late very often (or always) and low otherwise.

Note that this is just something I made up, but that is probably a feasible way of characterising certain situations: It is annoying when people are late more often than what is normally expected from people. Also, this way of setting up the priors actually follows examples reported on in Kao et al. (2014) quite closely.

If we now apply the model to this example to calculate posterior distributions, we get the following results.[4]




Here is how to read this: The rows enumerate the utterances available to the speaker in the model. For each utterance, the predicted posterior distribution with respect to the 10 possibilities for how often Sue is late are displayed, as well as the posterior predictions for the probability that the speaker is affected. The yellow shading indicates the possibilities that are compatible with the literal meaning of an utterance.[5] Let's first look at a neutral utterance: 'Sue is sometimes late'. What the posterior shows is that it is interpreted literally: most of the probability mass is in its literal interpretation. This is to be expected, since most of the probability mass of the literal interpretation corresponds to the mean of the prior. There is no reason, therefore, to suspect that the speaker is not being truthful, so it is likely that she is using the semantics of her utterance to refer to what

[4]The details of how these and other model predictions presented in this work came about can be found on: `https://github.com/rnouwen/hyperbole_rational`.

[5]As can be seen, to keep things simple I hand-coded a discrete meaning for '*often*' and '*very often*'. This is obviously not very realistic. It shouldn't affect the points I am making here and later, though.

she believes to be the actual state. As the affect posterior shows, once this utterance is done, there is an overwhelming probability that the speaker is not affected.

Now look at an extreme utterance: 'Sue is always late'. Here, we see that most of the probability mass is not in its literal interpretation, but rather further away from the scalar extreme. In other words, the model predicts that this sentence is likely to be interpreted non-literally. For now, don't pay too much attention to the details of the posterior distribution. What is important (for now at least) is the trend that is shown here: extreme utterances are interpreted non-literally, neutral utterances are interpreted literally. Note that the reason why an extreme utterance like (1) is interpreted non-literally is clear from the affect posterior: after witnessing 'Sue is always late', the listener is sure that the speaker is affected by this content.

This shows the potential of approaching hyperbole interpretation multidimensionally via a strategy of *rational repair*. Extreme utterances have utility beyond their literal meaning by virtue of being very suitable for conveying information w.r.t. some other question under discussion. In the example above, this results in weakening: If you look at the state posterior, you can see that 'Sue is always late' is interpreted as conveying something weaker than its literal meaning. In fact, if we see the prior as the norm, then the model shows that utterances expressing an extreme deviation from the norm, are interpreted as indicating a less sizeable deviation from the norm. This is in line with the norm-relatedness as described by Walton (2017), Popa-Wyatt (2020) and Nouwen (2024a).

The idea of rational repair as sketched here can only work, however, if there is some connection between the dimensions that make up the multidimensional meaning of the hyperbolic utterance. Kao et al. (2014) is not explicit at all about how this works. So in order to scrutinise rational repair more, I will first work out that part of the proposal.

# 3 Multidimensionality and relevance

That hyperbole triggers meanings beyond simple weakening is an old observation, but one that has been characterised in a multitude of different shapes and forms. For instance, de Mendoza Ibáñez (2014) remarks that speakers use hyperbole to 'draw attention to their own *attitude* about a situation' (p. 189, my emphasis). Peña-Cervel and Ruiz de Mendoza Ibáñez (2022) highlight the *emotive* character of the attitude expressed: 'the special meaning impact of hyperbole [. . . ] generally [indicates] a strong emo-

tional reaction on the part of the speaker' (p.4). Other authors use the somewhat generic term *affect* (for instance, Kao et al. 2014, Popa-Wyatt 2020, Feinmann 2023). Carston and Wearing (2015) describe the effect more specifically as *evaluation* when they claim that hyperbole is 'characterised by the blatant exaggeration of a relevant scalar property for the purpose of expressing an evaluation of a state of affairs' (p.79). Yet other parts of the literature highlight the argumentative and persuasive role that hyperbole can play (e.g. Henkemans 2014, Henkemans 2017), while at the same time acknowledging its affective or evaluative nature.

For an account like Kao et al.'s to work, all possible roles played by hyperbole in discourse should be captured. So, rather than having the model directly target affective, evaluative or persuasive effects of hyperbole, there should be a way of doing all that. I believe that we can, by taking the discourse role of questions under discussion seriously. I will assume three things, none of which I think are controversial: (i) every utterance comes with its own local question under discussion (which can be gleaned from the utterance information structure, Roberts 2012); (ii) any utterance may address additional questions under discussion; (iii) the potential of an utterance to answer questions under discussion other than its own is what makes an utterance relevant. Take the following example as an illustration:

(10) We should fire Sue. She is always late.

The example in (10) is a typical case of a claim-argument discourse relation. The first sentence provides a claim, which is supported by the second sentence. In terms of QUDs, this works as follows. Each of these sentences comes with its own local QUD. Respectively, the question whether or not we should fire Sue and the question how punctual Sue is. The first of these local QUDs, however, is also addressed by the second sentence. Presumably, typical listeners do not just adopt beliefs willy-nilly. So, after processing the first sentence, its QUD is still very much unresolved. As a result, when processing the second sentence, the listener will attempt to interpret it with respect to both available QUDs.

In a nutshell, then, individual utterances provide answers to their own local QUDs and use these answers to potentially contribute to other QUDs. Crucially, this is only possible if the QUDs are somehow linked to one another. A sentence can only contribute to two QUDs at once if an answer to one QUD is relevant to answering the other QUD. For this to be the case in the Bayesian setup of RSA, the context must provide a prior that connects the two QUDs in a non-trivial way. Let $\alpha$ be a space of possibilities and $\beta$

be another space of possibilities. We can now define expectations about the effect of $\alpha$ on $\beta$ as follows:

$$P(\beta = \ldots \mid \alpha = \ldots)$$

As long as there are answers $a$ to the QUD constituted by $\alpha$, such that $P(\beta = \ldots \mid \alpha = a)$ has low entropy, then $\alpha$ is relevant to $\beta$. In the example modelled in the previous section, we had a case where $\beta$ = {the speaker is affected, the speaker is not affected} and $\alpha$ is the space given by the question how often Sue is late. $P(\beta = \ldots \mid \alpha = \ldots)$ had low entropy for any value for $\alpha$. In case Sue was always or at least very often late, there was near certainty the speaker was affected, while in all other cases there was near certainty that the speaker was not affected.

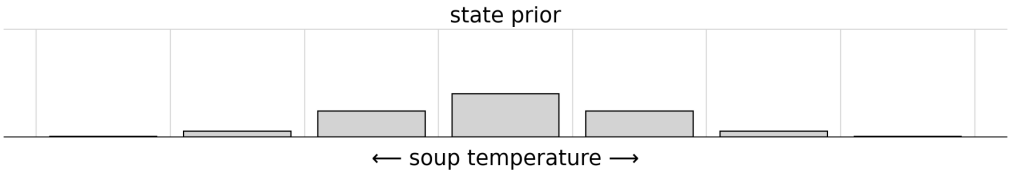
If we go back to the argumentation example in (10), we could model things in an entirely parallel way. Now, we have a prior where $\beta$ = {We should fire Sue, We should not fire Sue} and $\alpha$ is as before. In fact, we could imagine that the prior is exactly as before as well: If Sue is always or at least very often late, then this creates confidence that she should be fired. In other cases, the probability mass shifts to her not needing to be fired. As such, Kao et al.'s model would be able to deal just as well with a case like (10) as with the more affective case before.

Examples like these have in common that the correspondence between two questions under discussion is not erratic, but rather systematic. The more often Sue is late, the stronger the conviction will be that Sue needs to fired. And the more often she is late, the more annoyed the speaker will be. Such connections are crucial if the idea of *rational repair* is to work, since according to that account, hyperbole exists because the lack of utility in one domain translates to an increased utility in another domain. What I am highlighting here is something that Kao et al. do not highlight - the prior conditional distribution connecting the two questions under discussion does a lot of heavy lifting. Without a prior that guarantees relevance, the model will not work. With this in place, we can zoom in on this role played by relevance (i.e. the link between two questions under discussion) in the interpretation of hyperbole. The example discussed above displayed a very specific pattern of relevance. What happens when relevance looks rather different?

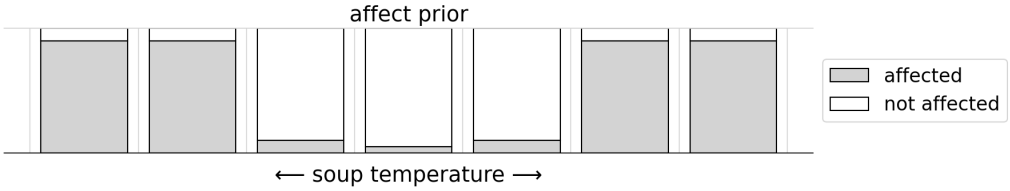
# 4 The problem of symmetrical relevance

Consider a new example, with a different kind of conditional prior. Say, we are eating soup and are discussing the temperature of this soup. Let's

say we have the following space of possibilities regarding the temperature, with prior expectations as indicated.



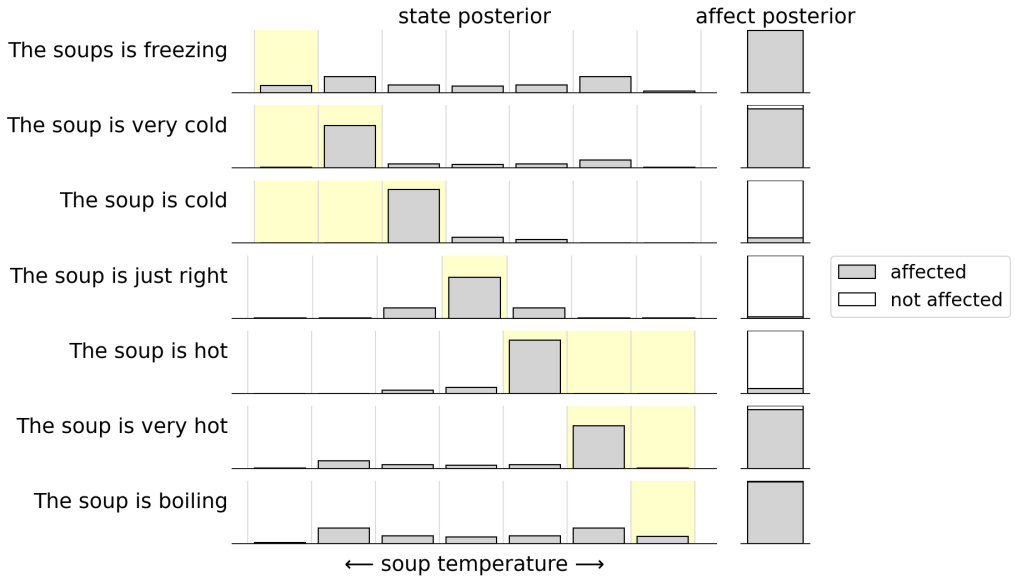

So, the middle possibility in this scale is the typical temperature of soup. To the left, are colder soups and to the right hotter ones. Let's say that we are in a context where we are discussing the temperature of the soup with the aim of establishing how enjoyable the soup is. We have the following conditional prior P( affect = ... | ... ):




This says that we are fine as long as the soup has a relatively normal temperature. But as soon as the soup is too cold or too warm, we will end up affected by it (meaning: the soup is not going to be to our satisfaction). The expectations about affect here are symmetrical. Both extremely cold soup and extremely hot soup will affect the soup-drinking speaker.[6] This is very different from the model that we looked at before, where the affect was associated to one side of the scale of Sue's punctuality and non-affect was associated with the rest of the scale.

The predictions of the model are now as follows.

[6] I think this is conceivable in this situation, but if you don't think it is, just think of some other parallel example where relevance is symmetrical in this particular way.




What the model gets right is that the extreme sentences *'The soup is freezing'* and *'The soup is boiling'* are unlikely to be interpreted literally. Also, the model predicts that given such an utterance, there is a high (posterior) probability that the speaker is affected. What goes wrong is the details of the state of affairs posterior. The appropriate interpretation would be a weakening of the presented deviation from the norm. For instance, 'the soup is freezing' should end up meaning that the soup is colder than the ideal temperature, but not so cold as to be literally freezing. Instead, it is interpreted as equally likely referring to a state of affairs that is hotter than desired than one that is colder than is desired. This is because either end of the scale is associated with the same probabilities for affecting the speaker. To the model, that is all that matters: once a literal interpretation is unlikely, the probability mass will be spread among options that are associated with affect, no matter where on the scale they are. As a result, the model wrongly predicts that 'the soup is freezing' ends up meaning almost the same as 'the soup is boiling'. (The only difference is that there is a small residual probability that these utterances are interpreted literally and refer to their end of the scale.)

One detail that might help understand the problem is that the model has no information about the ordering of the scale. To the model, the scale is just a set of possibilities. The only thing that vaguely creates a little bit of structure is the prior. If the prior is normally distributed, then it is easy to distinguish extreme scalar points from points closer to the middle. However, as became clear above, the prior has no information about direction.

Ultimately then, the rational repair account of Kao et al. (2014) manages to account for two important aspects of hyperbole interpretation: (i) that extreme utterances (utterance expressing extreme deviations from the norm) are interpreted non-literally, and (ii) that such utterances trigger affective inferences. What Kao et al.'s proposal does not account for is the nature of the non-literal, non-affective interpretation. This is a general property of the rational repair strategy. If the literal meaning of an utterance U lacks utility, because it is improbable then the repair involves the use of relevance to connect to a secondary QUD. The answer found to that QUD will be used to inform the listener about the answer to utterance QUD. If relevance is symmetrical, i.e. opposing scalar ends in the utterance QUD point to the same answer to the secondary QUD, then the result is a disjunctive reading, instead of simply weakening. The inference is that there is a deviation from the norm, without specifying the direction this deviation is in.

One could avoid this problem by banning symmetrical relevance. It is hard to see, however, what a non-symmetric link would be for the example above. For example, a question under discussion like *Are you enjoying the soup?* is answered in the same way by both (11-a) and (11-b).

(11) a. No, the soup is freezing!
b. No, the soup is boiling!

This suggests that both scalar extremes point to the same answer, while the scalar middle does not point to that answer. Hence, it suggests that there is symmetry.

# 5 The problem of multi-dimensional hyperbole

Kao et al.'s implementation of multi-dimensionality is rather coarse. The secondary QUD is simply whether or not the speaker is affected. There is no inference about the degree to which she is affected. A richer notion of the multi-dimensional meaning for hyperbole can be found in the cognitive semantic literature. de Mendoza Ibáñez (2014) and Peña-Cervel and Ruiz de Mendoza Ibáñez (2022) propose that hyperbole involves conceptual mappings in a similar way as has been proposed for metaphor. In particular, the idea is that the improbable state of affairs that is asserted to be the case is mapped onto some other conceptual domain and the hyperbolic statement ends up conveying something about that domain instead. So, stating that Sue is always late conveys that the speaker's affective state is the way it

would be if Sue's lack of punctuality were as extreme as it is asserted to be. This is obviously another variant on the rational repair account I sketched above. However, Peña-Cervel and Ruiz de Mendoza Ibáñez (2022) make clear that the target domain in their account is scalar, just as the source domain is. They discuss the case of *'This suitcase weighs a ton'* in the following way:

> "In the case of *This suitcase weighs a ton*, the mapping constrains the reasoning process in such a way that we think of the speaker's emotional reaction when trying to lift an excessively heavy suitcase as if it were the hypothetical emotional reaction that he would get when trying to lift and carry an imaginary one-ton suitcase." (Peña-Cervel and Ruiz de Mendoza Ibáñez 2022, section 4.1)

I believe there is problem with such a way of addressing the two-dimensionality of hyperbolic meaning, one that is reminiscent of discussions about metaphor (e.g. Fogelin, 2011). Metaphor is often thought of as a form of comparison in which the properties in one domain are mapped onto another. So, to say that "Sue is a rock" is to compare Sue's properties to that of a rock. In itself, however, this does not resolve the puzzle of metaphor. Metaphor is not just comparison, it is *figurative* comparison: Yes, Sue has properties that rocks have, but she has these properties *in a figurative way*. For instance, the fact that you can't move a rock, means that Sue is steadfast and reliable. So, two steps are involved: first the metaphor points to properties in the other domain, then those properties are interpreted figuratively, yes, metaphorically!

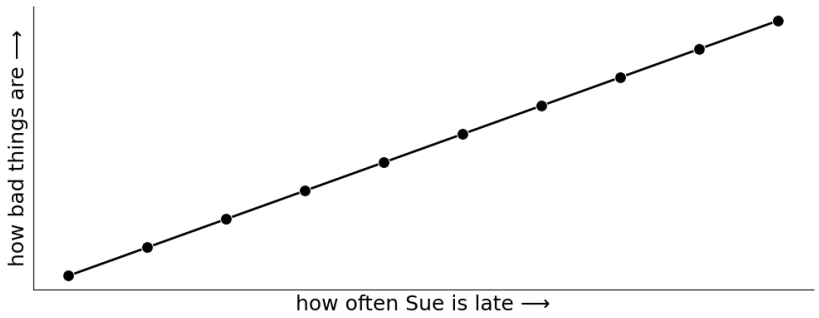
Similarly, in hyperbole, it does not suffice to say that the exaggeration creates a mapping to some subjective domain. According to the theory above, to say that I have a million unread emails in my inbox is to say that I feel the way I would feel if I had a million unread emails in my inbox (de Mendoza Ibáñez, 2014). But this cannot be the whole story, because this comparison in itself is hyperbolic. The speaker doesn't literally feel like they have one million unread emails. Presumably, that situation would be quite different from (and more extreme than) whatever the hyperbole is expressing. Similarly, when I exclaim that the suitcase I am carrying 'weighs a ton', it can't be that I am conveying that carrying this suitcase is literally like carrying an actual one ton suitcase. What I am conveying instead is that my situation is in a hyperbolic way like carrying such a fictitious suitcase. In other words, the domain mapping does not resolve hyperbole - it simply adds another layer of exaggeration.

This is connected to an issue discussed by Feinmann (2023). Some cases of hyperbole are literally impossible. Take a hyperbolic sentence like (12):

(12) I've been waiting forever for this day to arrive.

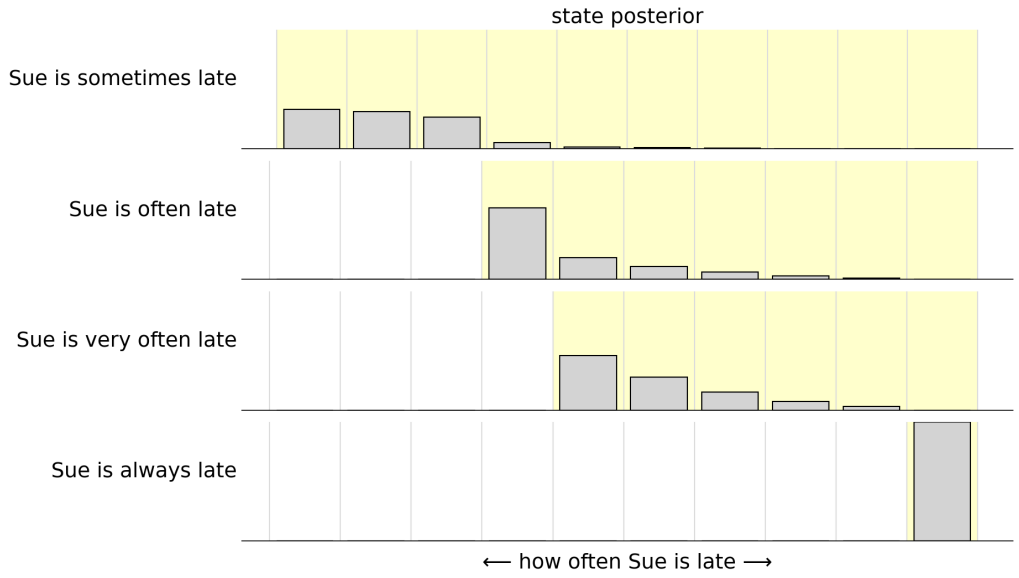
It can't be that the affective meaning of (12) comes about via some mechanism that can be paraphrased as: *the speaker feels as if she's been waiting forever for this day to arrive*. That paraphrase only makes figurative sense. We have no realistic understanding of what it feels to wait forever, so the paraphrase does not help. It is itself hyperbolic, so an account of two-dimensionality needs more than just a mapping between domains. The key feature of hyperbolic exaggeration is still weakening, so we still need an account of that.

We can make these issues more apparent, by implementing the conceptual mapping idea in the rational repair model of Kao et al. (2014). To do this, let's go back to the 'Sue is always late' example and assume there is a one-to-one connection between a scale of evaluation and the scale for how often Sue is late. So, for a given state of affairs in which Sue is d-often late, we have a corresponding idea of how good/bad that is.[7] The following graph illustrates this idea:




Given such a correspondence, where there is a simple conceptual mapping between two domains, the model's predictions become entirely trivial. This can be seen in the posterior probabilities predicted by the model:

[7] See Nouwen (2024b) for general considerations of phenomena in which objective scales and evaluation interact. See Geurts (2013) for discussion of the role correspondences like the one below could play a role in framing and decision making.




As becomes clear from this, the model is predicting literal interpretation now. This is because there is now utility in conveying the affective meaning that is associated with the extreme scalar position that *'Sue is always late'* is associated with. Because other scalar positions are associated with lower degree of affect, we now have a situation in which an answer to one question under discussion directly maps to an answer to another question under discussion. As a consequence, each answer now has utility, even if it is entirely improbable. Let's paraphrase the most extreme position on the evaluation scale 'things are improbably bad'. Since the prior probability of things being improbably bad is overwhelmingly high given the (highly unlikely) case of Sue being always late, effectively, the message 'Sue is always late' becomes synonymous with 'this is improbably bad'.

Note that the combination of meanings (Sue is always late, this is improbably bad) is extremely unlikely to be true. If we inspect the joint prior probability for the whole meaning space, the joint probability of that two-dimensional meaning is 0.0026. However, because apart from the possibility in which Sue is (literally) always late, there are no other possibilities in which things are improbably bad, this combination of meaning is the only option for the model as an interpretation of 'Sue is always late'. As a consequence, the model fails to predict hyperbolic interpretation.

# 6 Repair and composition

So far, we have seen that there are specific cases where the repair theory of Kao et al. (2014) fails to account for the multidimensional meaning of hyperbole, and that it lacks a way to account for the weakening effect, the main interpretive effect of exaggerating deviation from the norm. Putting these issues aside, I now turn to the question of what it is that a repair theory repairs. A repair theory most naturally takes a Gricean view on what needs repairing: It is the utterance that is defective and, so, the hyperbole meaning is a result of adjusting *utterance meaning*. In that sense, hyperbole is thought to be entirely pragmatic. It is a post-semantic, utterance-level mechanism. In this section, I discuss cases where hyperbole appears not to be utterance-level, but appears to be embedded.[8] Consider the following example:

(13) I've got a million things to do.

An utterance of (13) is interpreted as meaning that the speaker has a lot (but not a million) things to do and is somehow affected by that (e.g. she is overwhelmed.) Now consider (14).

(14) Sue thinks that I've got a million things to do.

This means that Sue believes the speaker to be so busy to be overwhelmed. In other words, it attributes to Sue a belief that corresponds to the hyperbolic interpretation of (13). Crucially, (14) does not entail that the speaker is busy, nor that she is overwhelmed. This is a problem for the repair theory, since on that theory the utility of expressing something improbable is because it allows us to express an affective meaning. The effect of (14), however, is not that this affective meaning is expressed. At least, it is not expressed at utterance level. So, the utterance meaning picture of repair cannot hold.

There are further reasons to question repair theories from the perspective of composition. Consider (15).

(15) Not everyone is always late.

I think there is a hyperbolic meaning to this sentence, in the sense that *'always late'* can be understood as 'late very often', so that the whole utter-

[8] I am not the first person to consider cases of embedded hypebrole. My thoughts in this section benefitted from discussion with Benjamin Spector. Some earlier work on embedded hyperbole can be found in work by his student Sossidi (2023).

ance ends up meaning that not everyone is late very often. How could this meaning come about?

For a repair theory to work for (15), there will have to be an incentive to repair. So far, such an incentive has always been the sheer improbable nature of the utterance. That cannot work in this case, since in contrast to earlier examples, this sentence is extremely unlikely to be *false*. In other words, for a repair theory to work for (15), we could need to assume that it is the fact that this sentence is almost certainly true that would trigger the repair. So far, so good.

A problem appears, however, when we then look at how the resulting meaning could be accounted for. To 'repair' the trivial literal meaning, we need the utterance to trigger an inference in an affective domain. Before, this was easy. Since the speaker is more affected, the more often Sue is late, we can easily infer affect from the exaggerated 'Sue is always late'. But with (15) the QUD is different now. For instance, it could be 'how punctual is your staff?' or 'Is everyone as late as Sue?' Crucially, the literal meaning of the utterance does not come with affect. In fact, its trivial nature makes it that it is compatible with a whole range of emotional states of the speaker. The sentence doesn't tell us anything in particular about the work force - it could be both good and bad. Even if we have a simplistic view on the QUDs involved as in (16). It still doesn't help us deriving the correct meaning.

(16) $\alpha$ { staff = good, staff = bad }
$\beta$ {Everyone is always late, Not everyone is always late}

To see this, look at the conditional prior probabilities. We are likely to associate the possibility in which everyone is always late with the evaluation 'staff = bad'. So, P('staff = bad' | everyone is always late) = 1. But what about P( $\cdot$ | not everyone is always late)? This would need to be a flat distribution. Knowing that not everyone is always late, is not knowing very much. As such, the consequences for evaluating the staff are non-existent. It is just as probable that the staff is good as that the staff is bad. But if that is the case, then there is no utility in using the hyperbole 'always late'. As a consequence, the repair theory will fail to predict that hyperbolic weakening should take place.[9]

You might think that this is a good prediction if you can convince yourself that (15) is not a case of hyperbole. Perhaps (15) is some kind of negative strengthening (Horn, 1989), akin to (17), which is interpreted as expressing something stronger than just the denial of Sue's performance

[9]In this case, weakening the hyperbole's literal meaning provides a stronger meaning at matrix level.

being very good, namely that is was rather bad:

(17) Sue's performance wasn't very good.

Unfortunately, I know of no theory that can account for negative strengthening, so this does not help much. Furthermore, there are cases of embedded hyperbole where negative strengthening seems unlikely:

(18) If someone tells me that one of my employees is always late, I ring HR.

Let us assume that the likelihood of someone telling the speaker that one of her employees is always late is very small. (This is a prerequisite for the repair theory to work). In that case, (18) is practically a tautology. Consequently, an account of the meaning of an utterance of (18) (that the speaker rings HR whenever she hears about one of her employees being late very frequently) runs into the same problems as we discussed for (15).

An even greater problem occurs with examples like (19):

(19) Employees who aren't always late may have their contract renewed.

In its literal meaning this sentence is *not* a implausibly true or false. Since it will be extremely rare for an employee to (literally) never be on time, the literal meaning of (19) is close to saying that all employees may have their contract renewed. That is a perfectly contingent proposition. As such, apart from the problem that embedding brings with deriving the intended hyperbolic meaning, in this case we have the additional issue that there is no incentive to repair to begin with.

# 7 Towards a non-rational theory of hyperbole

What could a non-rational theory of hyperbole look like? I will sketch one in three stages. First, I will propose a simple interpretation mechanism that weakens extreme deviations from the norm to meanings closer to the norm. Second, I will show that with this mechanism in place, we can resolve the issues that the rational repair approach ran into. The downside of this approach is that the interpretation mechanism is hand-coded to mimic hyperbole meaning. It is therefore entirely descriptive, not explanatory. Finally, I will speculate on how such an interpretation mechanism could be derived from a more general cognitive effect.

Let us start by thinking what a simple way of reducing a(n extreme) deviation of the norm would look like. One way to quantify deviation from the norm is to take the surprisal of a meaning $m$, as given by a prior probability $P$:

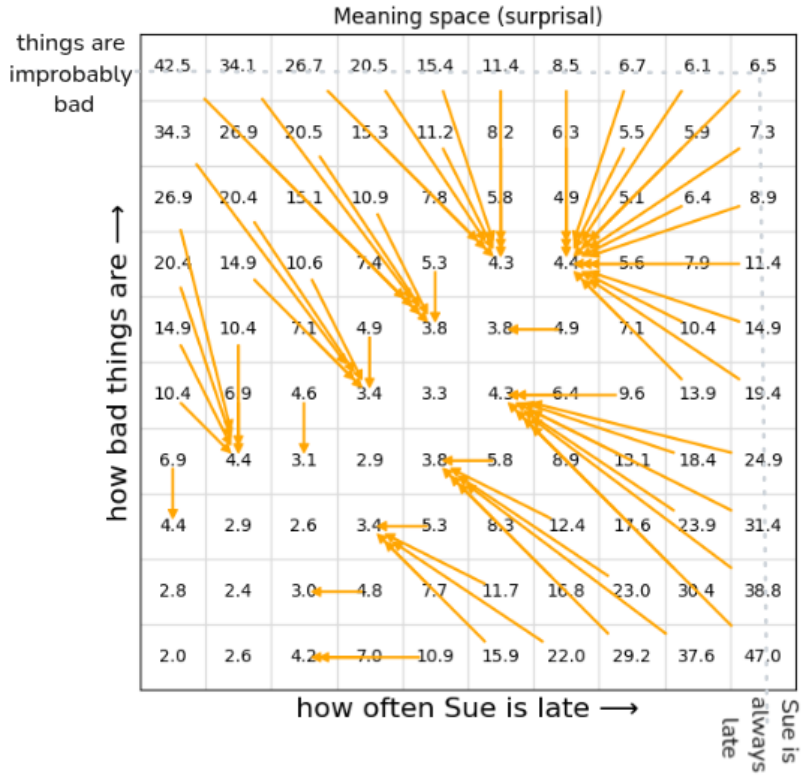
$$Dev(m) = -\log(P(m))$$

The idea now is to use this measure of norm-deviation to inform us how to view the meaning space. Meanings whose norm-deviation exceeds some threshold are mapped onto the closest meaning in the space that does not exceed the threshold. Let $d$ be the distance between two meanings in a meaning space, then:

$$\Omega(m) = \text{argmin}_{m':Dev(m')<\theta}\ d(m, m')$$

$\Omega(m)$ returns $m$ whenever its deviation from the norm does not exceed $\theta$ and it returns the closest meaning to $m$ that does, otherwise. The target use of $\Omega$ will be that it maps the literal meaning of a hyperbolic utterance to the nearest meaning that is norm-deviating within the bounds of what is conceivable. The table in (20) illustrates how $\Omega$ works. The table shows surprisal values for meanings in a two-dimensional meaning space. These are based on probabilities of these meanings, given some expectations similar to the priors in the rational repair theories. Given some threshold $\theta$ it maps outlandish meanings to closely related meanings that are sufficiently conceivable.

(20)




The mappings have the required weakening effect in the sense that it creates the norm-relative signature of hyperbole interpretation: hyperbole expresses a deviation from the norm, by exaggerating that deviation (Walton, 2017; Popa-Wyatt, 2020; Nouwen, 2024a). For example, 'Sue is always late' receives the following literal interpretation, where I represent multi-dimensional meaning by combining propositions with the operator ∘:

$$\mathrm{I}(\text{`}\textit{Sue is always late}\text{'})$$
$$=$$
$$\text{Sue is always late} \circ \text{Things are improbably bad}$$

This meaning is shifted to a more reasonable deviation of the norm by applying Ω:

$$\Omega(\mathrm{I}(\text{`}\textit{Sue is always late}\text{'}))$$
$$=$$
$$\text{Sue is late very often} \circ \text{Things are very bad}$$

Obviously, this is a cheat. It is way too specific and arbitrary to be a viable alternative account with explanatory value. But let us first of all just assume that there is such a weakening mechanism and see how it solves

the problems. Subsequently, I will speculate about the source of such a mechanism.

Since $\Omega$ operates on the multi-dimensional meaning space, this mechanism avoids the problem that was inherent in theories that try to explain hyperbole as a result of the conceptual mapping between domains. The weakening that $\Omega$ brings about is not just in the domain of the literal meaning, but also in the connected affective (or other secondary) domain. So, quite straightforwardly, $\Omega$ avoids this issue.

Can the weakening function also help with cases where hyperbole is embedded? Let us first consider (21) and (22).

(21) I've got a million things to do.

(22) Sue thinks I've got a million things to do.

Let's say that the interpretation of (21) is the two-dimensional meaning that combines the proposition that the speaker has 1,000,000 things to do and the proposition that the speaker is improbably overwhelmed. We can now quite straightforwardly take this complex meaning and use it to compositionally derive the meaning in (22). We then get that Sue believes that the speaker is very busy in a way that makes her very overwhelmed.

In the case of (22), composition involved embedding a full clause. What happens in cases where hyperbole involves sub-clausal structures? The first step is to recognise that $\Omega$ can apply to compositional meanings just as well as it can apply to possibilities. For instance, we could have a space defined as follows:

$$S \;=\; \{x \text{ is late } d\text{-often} \mid d\} \times \{\text{things are } e\text{-bad} \mid e\}$$

Here, one of the dimension contains a free variable, one that can be bound at a later stage in the composition. For this to work, we need to figure out what is the probability of $P(x \text{ is late } d\text{-often})$ for some $d$? I am assuming there is such a concept or that it is derivable.[10] In particular, I assume that the meaning like '$x$ is always late' (so where $d$ has the most extreme value) has an extremely low probability, while the probability of evaluation conditioned on that meaning is highest for an extremely bad evaluation (which I again paraphrase as 'things are improbably bad'). From there, we can apply $\Omega$ as follows:

$$\Omega(x \text{ is always late} \;\circ\; \text{things are improbably bad}))$$
$$=$$
$$(x \text{ is very often late} \;\circ\; \text{things are very bad})$$

[10] We could for instance just take some average $\frac{1}{|X|}\sum_{x\in X} P(x \text{ is late } d\text{-often})$.

We can now use this meaning to compose in the following way:

(23) Not everyone is always late.

(24) [Not everyone] $\lambda x$ [ ($x$ is very often late $\circ$ things are very bad) ]

The resulting meaning can perhaps best be paraphrased as follows: not everyone is such that they have the property that they are so very often late that thing are very bad. The account correctly predicts that the affect does not project, but it also predicts that at some embedded level the frequency of being late is connected to some kind of affect.

This proposal is admittedly rough and it is not my intention to submit a full fledged theory in this paper. For one, I have no worked out account of what the $\circ$ operator does. In how I discuss thing above, it is basically a simple conjunction operator, but that would seem suspiciously simplistic. Also, by paraphrasing a lot of the evaluative meanings, I am probably missing a lot of the nuance in the multi-dimensional meaning. The connection between the dimensions seems tighter. For instance, it is not just that 'Sue is late very often' pairs with 'things are very bad', but that at the same time it seems salient that there is a causal connection between the two. So perhaps a better way to approach things is to have meanings like '(Sue is late very often $\circ$ it is very bad that Sue is late very often)'. Moreover, it seems entirely probable that the meanings we are dealing with here are many-dimensional. They are rich concepts and the 2D illustrations I provide here do not do justice to that. (The same can be said for most other approaches to hyperbole in the literature, including Kao et al. 2014).

Despite the many questions I am leaving unanswered about this account, I do intend to show with the $\Omega$ mechanism that there is potential in a non-rational approach to hyperbole interpretation and that such an approach, once it is fully worked out, will avoid the issues that I raised above. However, this still leaves open the question where this mechanism comes from. Without a reasonable understanding of why $\Omega$ would exist, this proposal is nothing but a hack.

I think the answer lies in the observation that exaggeration, and in particular exaggeration of deviation from the norm, is a phenomenon that outside of the realm of language occurs throughout the natural world. The clearest connected notion is that of *supernormal stimuli*. Super-normality (Tinbergen, 1953; Staddon, 1975) refers to the brain's tendency to respond more strongly to a stimulus that is more extremely removed from a norm. The classical example stems from Niko Tinbergen's experiments with herring gull chicks. Such chicks peck a red dot on the parent's beak to make the parent regurgitate food. Tinbergen found that a very abstract model of

a beak that exaggerates key features of a real gull's beak - basically a thin rod with two contrasting colours - triggers more pecking behaviour than a naturalistic stimulus. An example of a super-normal stimulus in human everyday life is the wearing of make-up such as red lipstick. Female faces tend to have a greater contrast in luminance. Lipstick is a super-normal stimulus that exaggerates the facial contrast and thereby, despite its artificial nature, highlights femininity (Russell, 2009). Another example is caricature. Despite their inherent physical inaccuracy, caricatures are often recognised faster and more reliably than photographs (e.g., Ryan and Schwartz 1956; Rhodes 1996). Importantly, this phenomenon depends on norms: a caricature identifies those features in a face that deviate most from the norm and then amplifies them (Perkins, 1975).

It would seem then that from a cognitive perspective, the function of exaggeration is to highlight norm-deviation. There are some clues that this may be due to how information is represented in the brain. There is evidence that after learning, stimuli are not encoded as absolute values in the stimulus space, but rather as vectors encoding the difference of the stimulus relative to the average stimulus (e.g. Tanaka and Simon 1996; Giese and Leopold 2005), or as containing both specific and categorical information such as identity (Hill et al., 2019). This kind of encoding facilitates efficient recognition. At the same time, this explains for example why caricature recognition is so easy, because caricature exaggeration involves exactly the norm-deviating features. Supernormality, in other words, maximises discrimininability in a neural architecture.

Whatever accounts for our effortless interpretation of unnaturalistic stimuli like caricatures, this could be what instantiates a mechanism like $\Omega$. I am speculating then that hyperbole interpretation is not just non-rational, but that it is not linguistic. Hyperbole interpretation is nothing more or less than how our mind processes rich multidimensional concepts.

# 8 Conclusion

This paper has argued that hyperbole is not a rational speech act. While the scalar nature of hyperbole suggests that it should be amenable to explanation through rational pragmatic reasoning - much in the same way as scalar implicature has traditionally been approached - I have demonstrated that a theory that relies on a rational repair mechanism fails in systematic ways. These issues suggest that rather than seeking a rational explanation for why speakers exaggerate deviations from the norm, hyperbole interpretation is neither semantic nor pragmatic. Hyperbole, in this view, is not a

uniquely linguistic phenomenon but a manifestation of how exaggeration can help identify deviations from a norm.

# References


Carston, R. and Wearing, C. (2015). Hyperbolic language and its relation to metaphor and irony. *Journal of Pragmatics*, 79:79–92.

de Mendoza Ibáñez, F. J. R. (2014). Mapping concepts: Understanding figurative thought from a cognitive-linguistic perspective. *Revista Española de Lingüística Aplicada/Spanish Journal of Applied Linguistics*, 27(1):187–207.

Feinmann, D. (2023). Is hyperbole a scalar inference? *lingbuzz/007398*.

Fogelin, R. (2011). *Figuratively Speaking*. Yale University Press, New Haven and London.

Geurts, B. (2013). Alternatives in framing and decision making. *Mind & Language*, 28(1):1–19.

Giese, M. A. and Leopold, D. A. (2005). Physiologically inspired neural model for the encoding of face spaces. *Neurocomputing*, 65:93–101.

Goodman, N. (1968). *Languages of Art. An approach to a theory of symbols*. Indianapolis.

Henkemans, A. (2014). The use of hyperbole in the argumentation stage.

Henkemans, A. F. S. (2017). Strategic manoeuvring with hyperbole in political debate. *Contextualizing pragma-dialectics*, 12:269.

Hill, M. Q., Parde, C. J., Castillo, C. D., Colon, Y. I., Ranjan, R., Chen, J.-C., Blanz, V., and O'Toole, A. J. (2019). Deep Convolutional Neural Networks in the Face of Caricature: Identity and Image Revealed. *Nature Machine Intelligence*, 1(11):522–529.

Horn, L. (1989). *A natural history of negation*. University of Chicago Press, Chicago.

Kao, J. T., Wu, J. Y., Bergen, L., and Goodman, N. D. (2014). Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences*, 111(33):12002–12007.

Nouwen, R. (2024a). Meiosis and hyperbole as scalar phenomena. In Liefke, K., Anans, R., Gutzmann, D., and Scheffler, T., editors, *Sinn und Bedeutung 28*. Universität Konstanz.

Nouwen, R. (2024b). The semantics and probabilistic pragmatics of deadjectival intensifiers. *Semantics & Pragmatics*, 17:1–39. Accepted for publication.

Peña-Cervel, M. S. and Ruiz de Mendoza Ibáñez, F. J. (2022). *Figuring out figuration*. John Benjamins Publishing Company.

Perkins, D. (1975). A Definition of Caricature and Caricature and Recognition. *Studies in the Anthropology of Visual Communications*, 2(1):1–24.

Popa-Wyatt, M. (2020). Mind the gap: Expressing affect with hyperbole and hyperbolic figures. In *Producing Figurative Expression*, pages 449–468. John Benjamins Publishing Company.

Rhodes, G. (1996). *Superportraits: Caricatures and Recognition*. Essays in Cognitive Psychology. Psychology Press, Hove, East Sussex, UK.

Roberts, C. (2012). Information structure: Towards an integrated formal theory of pragmatics. *Semantics and pragmatics*, 5:6–1.

Russell, R. (2009). A sex difference in facial contrast and its exaggeration by cosmetics. *Perception*, 38(8):1211–1219.

Ryan, T. A. and Schwartz, C. B. (1956). Speed of Perception as a Function of Mode of Representation. *The American Journal of Psychology*, 69(1):60.

Scontras, G., Tessler, M. H., and Franke, M. (2021). A practical introduction to the rational speech act modeling framework. *arXiv preprint arXiv:2105.09867*.

Sossidi, E. (2023). Hyperbole in embedded environments: Investigating interpretations and theoretical implications. ENS Msc thesis.

Sperber, D. and Wilson, D. (1985). Loose talk. In *Proceedings of the Aristotelian society*, volume 86, pages 153–171. JSTOR.

Staddon, J. (1975). A note on the evolutionary significance of“ supernormal” stimuli. *The American Naturalist*, 109(969):541–545.

Tanaka, J. W. and Simon, V. (1996). Caricature Recognition in a Neural Network. *Visual Cognition*, 3(4):305–324.

Tinbergen, N. (1953). The herring gull's world: a study of the social behaviour of birds.

Walton, K. L. (2017). Meiosis, hyperbole, irony. *Philosophical Studies*, 174(1):105–120.