Hyperbole is not a rational speech act

Rick Nouwen

November 2026

1 Introduction

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.

One way to descriptively characterise the first of these effects is as *weakening*. The literal meaning of the hyperbolic utterance expresses an extreme position on a scale and its figurative meaning points to a less extreme position. The question is how this *weakening* effect comes about and how it is connected to other aspects of hyperbolic meaning. In this article I will look at what I believe to be the most well worked out theory in the literature that addresses this issue, Kao et al. (2014), couched in the rational speech act framework (Goodman and Frank, 2016). This account treats hyperbole as a phenomenon where figurative meaning emerges as a result of the literal meaning being defective. As such, the *weakening* effect is a rational side effect of standard considerations of meaning and use, albeit with the added complexity that hyperbole addresses multiple conversational goals at once.

In this paper, I aim to do two things. First of all, I will show that accounts of the kind presented in Kao et al. (2014) cannot make correct general predictions about the interpretation of hyperbole. Some of the arguments I put forward overlap with arguments that were independently presented earlier in Nouwen (2024a) and Feinmann (2026). I extend these arguments and, in contrast to these works, I will pinpoint the details where the rational speech act framework account fails, in part by using novel simulations. Additionally, I will make a general argument that the type of account provided in Kao et al. (2014), an account I will classify as a *repair theory*, cannot hold for hyperbole. Instead, I

will argue, we should consider a non-rational theory of hyperbolic utterances. The goal is not to provide such a theory in detail, but to show that hyperbole is a phenomenon that is unlikely to be captured by rational pragmatic frameworks.

2 Scalarity and norm-relativity

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". In this way, it is natural to think of hyperbole as a *scalar* phenomenon, in the sense that the literal and (part of) the figurative interpretation of the exaggerated utterance are associated to the same scalar dimension: the size of the olive for Goodman's example, the frequency of Sue being late for (1).

Feinmann (2026) argues that hyperbole is scalar in nature in a stronger sense than that: hyperbole is based on the kinds of scalar alternatives that are familiar from scalar implicature.[1] He does so by discussing contrasts between examples like the following. While (2) implicates that there are not (a lot) more than two hundred unread emails in my inbox, (3) obviously does not have the same implicature, since it already entails that. Similarly, (2) has a hyperbolic reading, while for (3) it seems substantially harder to interpret it non-literally.

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

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

From examples like these, Feinmann (2026) concludes that hyperbole is a phenomenon that involves scalar alternatives. He conjectures that only when S is a weaker scalar alternative to S' can S' express S hyperbolically. I think there are cases of hyperbole that do not involve scalar alternatives in this sense. Here is a first example.

(4) Stop nagging! That's the 10th time you asked me this!

The example in (4) involves the ordinal *'10th'*. Unlike bare numerals, ordinals do not support lower-bounded readings. *'This is the 4th time you asked me this'* does not entail *'This is the 3rd time you asked me this'*. Correspondingly, the latter does not implicate the denial of the former. Rather, it entails it. In this respect, ordinals behave just like the *'exactly'* numerals in (3). Like *'exactly two hundred'*, an ordinal statement has no weaker scalar alternatives. This

[1] In an earlier draft of this paper, my discussion of Feinmann's paper was based on a 2023 draft. I was unaware of the published version (Feinmann, 2026). The current version of my paper has an updated and more accurate discussion of Feinmann's points.

is Feinmann's explanation of why (3) resists a hyperbolic reading. By that same reasoning, (4) should resist such a reading too. Yet it clearly does not: intuitively, (4) can be used to convey that the addressee has asked something a good few times before, rather than that this is literally the tenth time.

This suggests that the reason why (3) is not easily read hyperbolically has nothing to do with the absence of a scalar alternative-inducing term, but rather with a manner-style inference triggered by the presence of *'exactly'*. In fact, adding *'exactly'* to (4) similarly suppresses the hyperbolic reading.[2]

(5) Stop nagging! That's exactly the 10th time you asked me this!

While I will have no definitive account of why examples like (5) and (3) lack hyperbolic readings, I think (4) clearly demonstrates this has nothing to do with the absence of a term triggering scalar alternatives. A slightly different example where hyperbole does not seem to involve scalar alternatives is the following:

(6) (How's your new ultra-thin laptop? Do you carry it around with you all day?)

Yes, that's no problem at all. It weighs 3 grams!

The example in (6) is clearly not meant literally. It conveys that the laptop is remarkably light, say, a few hundred grams. Here, the hyperbole goes in the opposite direction to the cases Feinmann discusses. How does this fit in with his conjecture? In his analysis of (2) and (3), Feinmann (2026) assumes that bare numerals have lower-bounded truth conditions, with the upper bound derived as a scalar implicature. On this view, (6) literally says that the laptop weighs at least 3 grams, which is simply true of a laptop of a few hundred grams. The meaning that is conveyed, that the laptop weighs a few hundred grams, entails this literal meaning. In other words, it is not a weaker but a *stronger* meaning than the literal one. Feinmann's conjecture is formulated for cases in which the hyperbolic meaning is weaker than the literal meaning, and so it says nothing about (6). Feinmann approaches hyperbole as a weakening problem, but (6) is a genuine case of hyperbole that does not involve weakening in his sense. At best, then, an account of hyperbole in terms of scalar alternatives is incomplete. Bringing (6) under the conjecture would require assuming that, in contexts like this one, numerals are upper-bounded, so that the scale is reversed, an assumption that is not without complications (Geurts, 2006).

There is, however, a sense in which (6) does involve weakening, but it is not weakening in terms of entailment. The literal meaning of (6) describes

[2]One reviewer indicates they think it is possible for examples like (3) to have hyperbolic readings. This would support the idea that *'exactly'* invites an optional manner implicature that removes the hyperbole. Further research should clarify the precise status of these examples.

an extreme deviation from what we expect a laptop to weigh. Its figurative meaning conveys a less extreme, more conceivable deviation: the laptop is unusually light, but not absurdly so. What is weakened, then, is not the literal meaning, but the deviation from expectation that it expresses. As I will argue now, this is how hyperbole works in general.

Here I believe it is important to highlight the role norms have been argued to play in hyperbole (Walton, 2017; Popa-Wyatt, 2020; Nouwen, 2024a). A hyperbolic statement doesn't just exaggerate some position on a scale in the way suggested for instance 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. Similarly, the two hundred unread emails said to be in the speaker's inbox in (2) clearly deviate excessively from what is normal and desired. Hence, it can be read hyperbolically as expressing that the actual number of unread emails equally deviates from the norm, though it is not two hundred. In contrast, since (7) does not express a deviation from the norm, we cannot express it using (8) hyperbolically.

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

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

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):

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

(10) 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 (9) 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 (10) 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 (9) to exaggerate the deviation in Timid's case,

and (10) to exaggerate the deviation (in the other direction) in Brag's case.[3] 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.

3 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 so, we are on the look out for 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 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*'.[4] The relevant key assumptions are summarised as follows:

Rational repair:

- The literal meaning of a hyperbolic utterance is so improbable that it lacks utility and, as a result, it is in need of repair.[5]

[3] 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 people 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.

[4] 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.

[5] For instance, Fogelin (2011) holds that hyperbole involves saying 'something stronger than

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

A rational repair theory is rational in a narrow sense, namely that meaning comes about through a rational mechanism that reconstructs the speaker's intentions by reasoning about the utility of utterances. The incentive for this reasoning is the fact that the literal meaning has no utility in the context of use.

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 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 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.

what I have a right to say, with the intention of having it corrected away from the extreme' (p. 13).

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 partitioned 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 $s \in S$:

(11) $P(s|utterance = U) \propto P(s)P(utterance = U|s)$

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 $A = \{\text{affected, not affected}\}$. 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 $P(A = \text{affected} \mid s)$.

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

(12) $$P(s, a|utterance = U) \propto$$
$$P(s)P(a|s) \sum_{g \in G} P(g)\, P(utterance = U|s, a, g)$$

Here G is the set of conversational goals the speaker may have: she may want to convey the state of the world ($g =$ *state*), her affective state ($g =$ *affect*), or both ($g =$ *both*).[6] The listener does not know which goal the speaker has, and so the likelihood term is averaged over the possible goals, weighted by the *goal prior* $P(g)$.

To understand what (12) 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, depending on her goal g. Such likelihood is calculated on the basis of the speaker's utility for choosing a certain utterance, together with a rationality parameter α that determines how strongly the speaker favours the utterance with the highest utility. For example, in the original RSA setup in (11), the utility of choosing an utterance U

[6] Since utterances are only about the state, and affect depends only on the state, the goal of conveying both state and affect yields exactly the same speaker behaviour as the goal of conveying the state alone: the factor $P(a \mid s)$ is the same for every utterance and cancels out in the speaker's choice. Effectively, then, the only parameter of the goal prior that matters is the probability of the affect goal. I decided to keep the 'both' option in the model to stay as close as possible to Kao et al.'s proposal.

state s	1	2	3	4	5	6	7	8	9	10
$P(s)$	.235	.222	.188	.143	.097	.059	.032	.015	.007	.003
$P(\text{affected} \mid s)$	.10	.20	.30	.70	.80	.80	.80	.90	.90	.99

goal prior: $P(\textit{state}) = P(\textit{affect}) = P(\textit{both}) = \frac{1}{3}$ rationality: $\alpha = 2$

Table 1: Model settings for the QUD 'How often is Sue late?'. States run from rarely late (1) to always late (10). The state prior is a discretised Gaussian ($\mu = 0$, $\sigma = 3$) over the ten states.

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 (12), 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?'. For this example we will simulate the interpretation of utterances according to the model, using the values given in table 1.

I am assume there is a scale of 10 discrete possibilities as to how often Sue was late. On one extreme (state 10) is the possibility where Sue is always late. The probability that this possibility is actual is expected to be virtually 0. For states < 10, 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.

The affect prior 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 calculate posterior distributions according to the model, we get the results in table 2.[7]

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 given, as well as the posterior predictions for the probability that the speaker is affected. Bold values indicate the possibilities that are compatible with the literal meaning of an utterance.[8] Let's first look at a neutral utterance: 'Sue is sometimes late'. What the posterior shows is that it is interpreted literally. This is to

[7] The details of how these and other model predictions presented in this work came about can be found in the accompanying python notebook.

[8] 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.

	state posterior $P(s \mid u)$										affect
'Sue is …'	1	2	3	4	5	6	7	8	9	10	P(affected)
… sometimes late'	**.35**	**.32**	**.26**	**.04**	**.01**	**.01**	**0**	**0**	**0**	**0**	.17
… late often'	.06	.06	.06	**.53**	**.13**	**.08**	**.04**	**.02**	**.01**	**0**	.68
… late very often'	.05	.06	.06	.07	**.35**	**.21**	**.12**	**.06**	**.02**	**0**	.77
… always late'	.06	.11	.14	.24	.19	.11	.06	.03	.01	**.04**	1.00

Table 2: Posterior distributions for the QUD 'How often is Sue late?'. Bold: states of which the utterance is literally true.

be expected, since the literal interpretation of this sentence is predicted to be true according to 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 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.[9] 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 multi-dimensionally 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).

It seems then that this kind of repair theory appears to capture the basic pattern of the pragmatics of hyperbole. Not only does it account for weakening, it also has a build-in norm-relativity. The example *'Sue is always late'*,

[9] In fact, the weakening is rather too strong: on average, 'Sue is always late' is interpreted as describing less frequent lateness than 'Sue is very often late'. We will see in section 5 that this is symptomatic of a more general problem.

however, shares features with the kinds of hyperbolic utterances studied in Kao et al. (2014). As will become clear below, however, these features are rather specific. For that reason, it would be good to test these models beyond the narrow domain that they were tested on so far.

The normal way to approach a computational model like this is by testing it on experimental data. However, the data required for the kinds of sentences I will be considering here are quite complex. Not only do we need information about what inferences participants draw about some objective scale, we also need information about the affective interpretation. This is not easy, and for that reason experiments that support models like this often opt to simplify the operationalisation for measures of meaning. For instance, in Kao et al. affect is binary - it is either there or it is not. As I will argue below, this is too coarse: it leads to wrong predictions and it cannot capture meanings that address the degree to which a speaker is affected. The model still makes predictions for more finegrained notions of affect, though. However, in that case ensuring proper construct validity in a human experiment is increasingly difficult. How do you measure affect magnitudes? How do you make sure everyone in the experiment associates the same kind of affect (how bad it is? how disappointing? how surprising? etc.) with a stimulus?

Another example of this issue is the goal prior. I question whether measuring these priors empirically is ever realistic. Language users do not normally have access to theoretical notions like conversational goals. Some proxies may be tried, but it will be difficult to know the reliability of these.

If we want RSA models, like those proposed in Kao et al. (2014), to play a role not just as models, but also as entities with theoretical relevance, then there should be ways to scrutinise the predictions they make, even for cases that we only have armchair data on. There is a rich history of developing ideas about pragmatics non-experimentally. While the past 25 years or so have shown that there is enormous value in experimental pragmatics, there is equally value in non-experimental pragmatic theorizing. RSA can, and should, play a role in both these sides of pragmatics.

In the remainder of this paper, I will test the RSA account of hyperbole without bringing in novel or existing experimental data. Instead, I will show that there are observations about hyperbole that can be assessed introspectively which are not and cannot be predicted by an account like that of Kao et al. (2014). I will do this by simulating predictions, argue that these predictions are wrong, and that this negative outcome is independent from the particular way the simulation was set up. Thus, I aim to make claims that hold generally for the kind of theory Kao et al.'s model implements, namely the family of theories I dubbed rational repair.

4 Multidimensionality and relevance

The idea of rational repair as sketched above is dependent on there being some connection between the dimensions that make up the multi-dimensional meaning of the hyperbolic utterance. Hyperbole is interpreted non-literally, since the literal meaning is associated with an affective meaning that has utility. Kao et al. (2014) is not explicit at all about how such connections work. So in order to scrutinise rational repair more, I will first work out that part of the proposal.

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 emotional 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 2026). 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 dimensions of meaning that hyperbole may express 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 the rational speech act framework can easily do this 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:

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

The example in (13) 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 α be a space of possibilities and β be another space of possibilities. We can now define expectations about the effect of α on β as follows:

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

We can now say that α is relevant to β if learning the answer to α changes our expectations about β, that is, if $P(\beta = \ldots \mid \alpha = a)$ differs from $P(\beta = \ldots)$ for at least some answers a. (Formally, α is relevant to β to the extent that their mutual information is greater than 0.) In the example modelled in the previous section, β = \the speaker is affected, the speaker is not affected\ and α is the space given by the question how often Sue is late. There, the probability that the speaker is affected increases steadily with how often Sue is late: it is high when Sue is late very often or always, and low when she is only sometimes late.

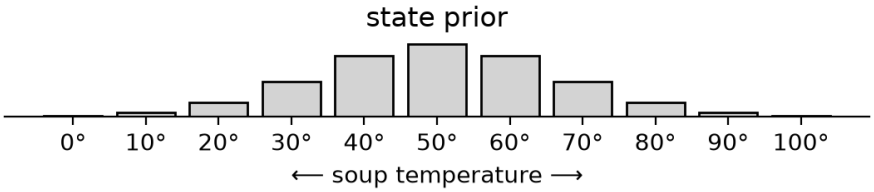
If we go back to the argumentation example in (13), we could model things in an entirely parallel way. Now, we have a prior where β = {We should fire Sue, We should not fire Sue} and α is as before. In fact, we could imagine that the prior is exactly as before as well: If Sue is always or at least very 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 (13) 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?

5 The problem of symmetrical relevance

Consider a new example, with a different kind of conditional prior, one of the kind also discussed in Nouwen (2024a) and Feinmann (2026) (section 2.1.1).[10] 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 and that we hope the soup's temperature is in some Goldilocks zone, neither too hot, nor too cold. So, 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 negatively (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.[11] 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 in table 4, with table 3 providing the details of the model settings. What the model gets right is that the extreme

[10] There is clear overlap between the general point made in this section and Feinmann (2026). There seems to have been a simultaneous convergence to the same idea. That is, Feinmann (2026) argues in much the same way as I do in this section that symmetrical relevance is problematic to Kao et al. (2014), though he does not develop the notion of symmetrical relevance in the same way. Below, however, I improve the force of the argument beyond what is in Feinmann (2026) by engaging with what may seem to be ways of addressing the issue within the RSA framework.

[11] 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.

temperature (°C)	0	10	20	30	40	50	60	70	80	90	100
$P(s)$	.003	.013	.047	.117	.200	.240	.200	.117	.047	.013	.003
$P(\text{affected} \mid s)$	.90	.90	.74	.26	.08	.05	.08	.26	.74	.90	.90

goal prior: $P(\textit{state}) = P(\textit{affect}) = P(\textit{both}) = \frac{1}{3}$ rationality: $\alpha = 2$

Table 3: Model settings for the soup example. The state prior is a discretised Gaussian centred on 50°C ($\sigma \approx 16.7$°C).

	state posterior $P(s \mid u)$, in °C											affect
'The soup is …'	0	10	20	30	40	50	60	70	80	90	100	$P(\text{affected})$
… freezing'	**.09**	.05	.15	.13	.08	.07	.08	.13	.15	.05	.01	.93
… very cold'	**0**	**.16**	**.56**	.05	.04	.04	.04	.05	.04	.01	0	.75
… cold'	**0**	**0**	**.01**	**.30**	**.53**	.07	.06	.03	0	0	0	.12
… just right'	0	0	.01	.04	.10	**.71**	.10	.04	.01	0	0	.03
… hot'	0	0	0	.03	.06	.07	**.53**	**.30**	**.01**	**0**	**0**	.12
… very hot'	0	.01	.04	.05	.04	.04	.04	.05	**.56**	**.16**	**0**	.75
… boiling'	.01	.05	.15	.13	.08	.07	.08	.13	.15	.05	**.09**	.93

Table 4: Posterior distributions for the soup example. Bold: states of which the utterance is literally true.

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'*.

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.

If you look closely at the table, you will notice that the posterior distributions for *'The soup is freezing'* and *'The soup is boiling'* aren't completely

goal prior			expected temperature (°C)	
$P(\textit{state})$	$P(\textit{affect})$	$P(\textit{both})$	'freezing'	'boiling'
.45	.10	.45	35.6	64.4
.40	.20	.40	42.4	57.6
$\mathbf{\frac{1}{3}}$	$\mathbf{\frac{1}{3}}$	$\mathbf{\frac{1}{3}}$	**45.9**	**54.1**
$\frac{2}{3}$	$\frac{1}{3}$	0	45.9	54.1
0	$\frac{1}{3}$	$\frac{2}{3}$	45.9	54.1
.50	.50	0	47.9	52.1
.25	.50	.25	47.9	52.1
.125	.75	.125	49.3	50.7
.05	.90	.05	49.8	50.2

Table 5: Expected temperature $E[T \mid u]$ for 'The soup is freezing' and 'The soup is boiling' under different goal priors ($\alpha = 2$). Bold: the goal prior used in the simulation.

the same. The difference is that there is a small residual probability that these utterances are interpreted literally and refer to their end of the scale. So, the *'freezing'* utterance has a slightly higher probability for the 0 °C, while the *'boiling'* utterance has a slightly higher probability for the 100 °C. This is because, if at least part of the conversational goal is to convey the temperature of the soup, then some impact of the literal interpretation is to be expected.

An anonymous reviewer suggests that this is perhaps the solution to the symmetry problem I present. Perhaps we should evaluate the model on *the expected state*, given the posterior distribution over states. Then, if there is more mass at the top of the scale in the *'boiling'* utterance and more mass at the bottom of the scale in the *'freezing'* utterance, we have broken symmetry.

This much is true. If we calculate expected states based on the above simulation, then we get different values for different utterances: $E[s|$ *'the soup is freezing'* $] = 45.9$ °C and $E[s|$ *'the soup is boiling'* $] = 54.1$ °C. While this is a break with symmetry, the resulting values are not really extreme enough to do justice to the intuitive meaning of the hyperbolic utterances. These expected values are closer to the mean than the expected states based on the posteriors for the other utterances.

Of course, this could just be because I made up the goal prior. To investigate this, I ran more simulations, where I systematically changed the goal priors. Table 5 summarises the results of these simulations in terms of expected temperature according to the posterior.[12] The results indicate that the goal priors do indeed have an impact on the expected temperature given pos-

[12] See foornote 6 for an explanation why $P(both)$ is immaterial to the posterior predictions.

teriors for the hyperbolic utterances. Expected temperatures only move noticeably towards the extremes for a low setting of $P(\textit{affect})$, and even then not very far. However, here we get a result that moves in the right direction for the wrong reason. The meaning is dragged towards the literal meaning here because the non-literal meaning is less likely to surface because of the low probability $P(\textit{affect})$. So the effect is based on literal interpretation, not non-literal. To see this, consider what the actual posterior distribution in this case looks like in table 6. As we can see here the interpretation is one that points to uncertainty about on the one hand either a literal interpretation and on the other hand an interpretation that is compatible with most of the rest of the scale. This is not what we are after. A true hyperbolic interpretation is one where it places most of the posterior mass in the area of the scale between the norm and the literal interpretation. That is, an interpretation that rules out the option that the true temperature is what is literally expressed. That is quite different from what is predicted here, despite the expected degree pointing to a temperature in that range.

In general, I do not think that $E[s|u]$ is a good value to represent the model's predicted interpretation. Consider the semantics of just any lower-bounded scalar quantifier. For instance, we are interpreting the simple sentence in (14):

(14) Some of the 10 marbles are red.

Say, then that we have experiment data on how participants interpret (14). This could be best guesses as to how many marbles are red, bet distributions along the scale [0,10], or any other way of looking at this question experimentally (see e.g. Franke et al. 2016). We summarise the experimental results in terms of the expected value (for instance by taking the mean of the best guesses, by weighing the bets distributed over the scale, etc.). We then build an RSA model and compare the expected values according to the posterior of the model to the expected value given the experimental data. Can we now say that the model captures the interpretation of the sentence when the expected values match? Obviously not. If the participants come to an expected value of 5. I could have a very simple model that puts half the probability mass on none of the 10 marbles being red and the other half on all of them being red. This would provide the correct expected value, but it would be hopelessly inappropriate as a model, since it suggests that some participants entertain the option that *'some'* is compatible with *'none'*. In summary, expected values mask details of scalar interpretation that matter and this is why I avoid using them in the evaluation of model predictions.[13]

[13] Another example of why expected outcome is an unsuitable measure of meaning could be disjunction. If we provide the model with a disjunction, we want the model to capture the uncertainty expressed by that sentence. So, (†) should not mean that the microwave heats everything to 50°C, but rather that it either makes things too warm, or leaves things too cold.

	state posterior $P(s \mid u)$, in °C											affect
'The soup is …'	0	10	20	30	40	50	60	70	80	90	100	P(affected)
… freezing'	**.30**	.04	.12	.10	.06	.05	.06	.10	.12	.04	.01	.93

Table 6: Posterior distribution for 'The soup is freezing' under the goal prior that best separates 'freezing' from 'boiling': $P(\textit{state}) = .45$, $P(\textit{affect}) = .10$, $P(\textit{both}) = .45$, $\alpha = 2$. Bold: states of which the utterance is literally true.

Going back to the simulations in table 4, my conclusion is that the small amount of probability mass that the model assigns to the literal meaning, does not help to obtain a hyperbolic interpretation. The only way this could work is if this mass tilts the model towards a suitable interpretation whenever there is more inclination for a non-literal meaning. This is not what we observe. Something approaching the hyperbolic interpretation can only be obtained when the utterance is mostly read literally. This is a wholly unintuitive result. Moreover, figurative meaning is only obtained under the problematic assumption that expected values are suitable indicators of interpretation.[14]

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 inter-

In this case, however, one could argue that this is because of a naive treatment of disjunction, where the proper treatment involves epistemic states, rather than states of the world.

(†) This microwave is rubbish. The soup is either freezing or boiling.

[14] A natural response to the symmetry problem, and to the arguments of this section more generally, may be to turn to a different kind of rational model, one with *lexical uncertainty* (Bergen et al., 2016; Potts et al., 2015). In such models, the listener is not sure which lexicon the speaker is using and reasons about this alongside reasoning about the state of the world. Hyperbole could then be treated as a kind of ambiguity: *'the soup is freezing'* is ambiguous between its literal meaning and a range of weaker meanings, and the listener infers which of these the speaker intended. Note, however, that this would require a substantial departure from existing lexical uncertainty models. In those models, the lexica a listener considers are refinements of the literal meanings: they can strengthen what a word means, but never weaken it. For hyperbole, we would instead need uncertainty about weakenings, and since nothing in the literal meaning of *freezing* determines how far it may be weakened, or in which direction, the space of candidate meanings would have to be given by the prior over lexica. It is this prior that would then do the explanatory work. If it only contains meanings on which *freezing* covers soup that is colder than normal, then the symmetry problem disappears, but only because the ordering of the scale and the direction of weakening have been written into the lexicon. What remains rational in such an account is the resolution of an ambiguity, just like the resolution of any other lexical ambiguity. The hyperbolic meaning itself is not derived by pragmatic reasoning. It is presupposed by it. Moreover, on a lexicon on which *freezing* means something like 'rather cold', it says much the same as *cold*, and the speaker would have no reason to prefer the hyperbolic expression over its literal counterpart. The rhetorical effect of hyperbole would have to come from elsewhere, for instance from an affective QUD, which would essentially bring in the mechanisms proposed by Kao et al. (2014).

preted non-literally, and (ii) that such utterances trigger affective inferences.[15] 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. For instance, Nouwen (2024a) proposes that the correspondence between some scale and the evaluation of positions on that scale should always be monotonic. In other words, the idea would be that affect priors (for whatever reason) are always of the kind in table 1. To me this seems too strong an assumption. It is hard to see, for instance, what a monotonic 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 (15-a) and (15-b).

(15) 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.

Another possible solution for the symmetry problem could be to move from binary affect ($\$the speaker is affected, the speaker is not affected$\}$) to an affect space that is more informative: $\$the soup is bad because too cold, the soup is fine, the soup is bad because too hot$\}$. It seems to me unlikely, however, that a discourse can set up such a specific partition, one that mixes evaluation with the cause of that evaluation. Evaluative QUDs are rather open-ended: a question can ask for an evaluation without anticipating which aspect of the thing evaluated the answer will address. When out of the blue I ask someone *'How is the soup?'*, there are many, many ways to evaluate the soup negatively: too hot, too cold, too bland, too salty, too sour, too sweet, too spicy, too watery, too thick. All these ways of evaluating the soup (and countless others) would have to jointly make up the QUD: $\$too hot and too watery but not too salty ..., too cold and too salty but not too watery ..., ...$\}$. And whether such a space is fixed by the question or inferred from the answer, its cells are no longer purely affective: it is information about the state of the

[15] Because affect is tied to the notion of a QUD, Kao et al. could furthermore in principle account for other kinds of secondary inferences apart from 'affect'.

soup, not affect, that breaks the symmetry.

6 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.[16] Yet, degree is part of what hyperbole conveys: *'This suitcase is heavy'* and *'This suitcase weighs a ton'* both express that the speaker is bothered, but not to the same degree, and a binary affect variable cannot represent that difference. 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; Wilson and Carston, 2007).[17] 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

[16] There *is* a prediction about the probability that the speaker is affected. But there being a .5 probability that she is affected is not the same as her being affected to some medium degree.

[17] This is sometimes referred to as the 'emergent property problem'.

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. First of all, neither I nor my addressees will have much experience of what it is like to carry a one ton suitcase. So, it will be quite unclear what I am expressing. Second, even if I did have that experience, it will surely be different from carrying a suitcase that is simply 'extremely heavy'. What I am conveying instead is that my situation is in a hyperbolic way like carrying a fictitious one ton suitcase. In other words, the domain mapping does not resolve hyperbole - it simply adds another layer of exaggeration.

This issue becomes even more clear if we look at cases of hyperbole where the literal meaning is literally impossible.[18] Take a hyperbolic sentence like (16):

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

It can't be that the affective meaning of (16) 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 us consider the example from Peña-Cervel and Ruiz de Mendoza Ibáñez (2022), *'This suitcase weighs a ton'*, and assume there is a one-to-one connection between a scale of evaluation and the scale for how heavy the suitcase is. So, for a given state of affairs in which the suitcase's weight falls within some range, we have a corresponding idea of how good/bad that is. For instance, let us partition the scale of suitcase weights into four cells: 0-19kg, 20-49kg,

[18] Feinmann (2026) notes an independent problem for Kao et al. (2014) with examples like these, namely that Bayes law is undefined for events with probability 0.

kg	0–19	20–49	50–99	100–
$P(s)$	.742	.258	0	0
$P(\text{A1} \mid s)$	1.00	0	0	0
$P(\text{A2} \mid s)$	0	1.00	0	0
$P(\text{A3} \mid s)$	0	0	1.00	0
$P(\text{A4} \mid s)$	0	0	0	1.00

goal prior: $P(\textit{state}) = P(\textit{affect}) = P(\textit{both}) = \frac{1}{3}$ rationality: $\alpha = 2$

Table 7: Model settings for the suitcase with a one-to-one evaluation. The state prior is a normal distribution ($\mu = 15$ kg, $\sigma = 7.5$ kg), truncated at 0 kg; $P(100 \text{ kg or more}) \approx 5{\cdot}10^{-30}$.

	state posterior $P(s \mid u)$, in kg				evaluation posterior			
'My suitcase …'	0–19	20–49	50–99	100–	A1	A2	A3	A4
… is light'	**1.00**	0	0	0	1.00	0	0	0
… is heavy'	0	**1.00**	**0**	**0**	0	1.00	0	0
… is very heavy'	0	0	**1.00**	**0**	0	0	1.00	0
… weighs a ton'	0	0	0	**1.00**	0	0	0	1.00

Table 8: Posterior distributions for the suitcase with a one-to-one evaluation. Bold: states of which the utterance is literally true.

50-99kg and at least 100kg. Additionally, let us assume that any suitcase in the first cell will be evaluated as A1, suitcases in the second cell are evaluated as A2, in the third cell as A3 and in the fourth A4. I will keep these evaluations abstract. A1 just refers to whatever evaluation is fitting for having to carry a suitcase under 20kg. So, it's something like 'this is okay', whereas A2, A3 and A4 point to increasingly more negative evaluations. In the model there are four possible utterances: *'This suitcase is light'* (only true for the 0-19kg state), *'This suitcase is heavy'* (true for 20-49kg, and upwards), *'This suitcase is very heavy'* (true for 50-99kg, and upwards) and *'This suitcase weighs a ton'* (only compatible in its literal meaning with the at least 100kg state). The model details are in table 7 and the posterior predictions are in table 8.

Because there is a one-to-one correspondence between non-affective meaning and affective meanings, there is no longer an incentive in the model to interpret non-literally. In this setup, each literal interpretation is associated with its own unique affect and, so, affective and non-affective discourse goals are settled in the same way. For that reason, the model predicts no weakening effect for *'This suitcase weighs a ton'*, despite the fact that the prior probability of this sentence being true on its literal interpretation is minute.

The setup in table 7 is of course rather unrealistic. It is likely that the affect we associate with suitcase weights is (i) more finegrained and (ii) involves a

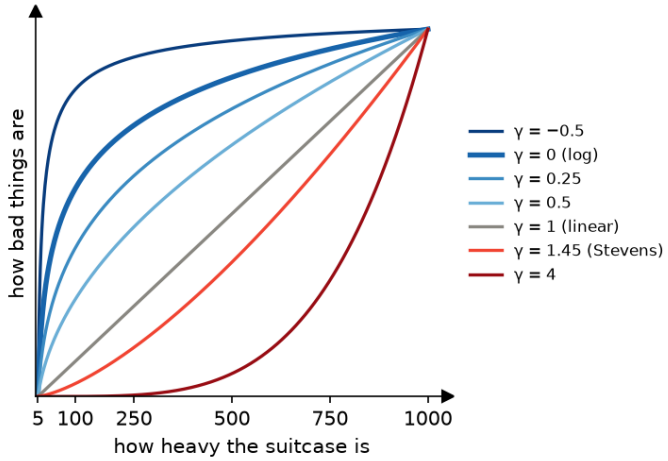


Figure 1: A family of power functions linking weight and affect.

degree of uncertainty. So let us now investigate what the model predicts for QUDs that define a finegrained space of possibilities and a correspondence between non-affective and affective spaces that is noisy.[19]

It is not immediately obvious what the mapping from weights to affect states should look like. To cover a multitude of options, I define a family of power laws:

$$f_\gamma(\mathit{weight}) = k \cdot \mathit{weight}^\gamma + c$$

The idea is that f maps a weight to a measure of how bad it feels to carry that weight. I choose k and c so that $f(5\ kg) = 0$ and $f(1000\ kg) = 1$, so as to keep the evaluation scale constant.[20] (See figure 1 for an illustration of this family of functions.) Since there is likely to be some uncertainty in any kind of evaluation, I assume that actual evaluations (which I will express by e) are normally distributed around the value assigned by f:

$$e(\mathit{weight}) \sim \mathcal{N}(f_\gamma(\mathit{weight}), \sigma)$$

Using e, I now conduct a range of simulations using different weight-affect mappings (γ) and different levels of noise in those mappings (σ).

Figure 2 shows the probability that a hyperbolic interpretation arises for *'This suitcase weighs a ton'*, where hyperbolic meaning is operationalised as

[19] 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.

[20] $k = \frac{1}{(1000^\gamma - 5^\gamma)}$; $c = -5^\gamma \cdot k$

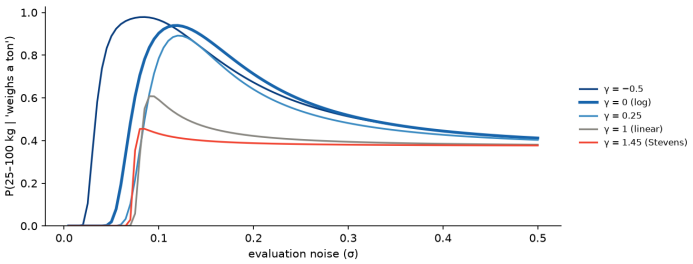


Figure 2: How posterior predictions depend on the state-affect mapping and on the noisiness of that mapping.

the total posterior probability that the weight is between 25kg and 100kg for different values of γ and σ.[21] Note, first of all, that for very low standard deviations the probability for a hyperbolic reading is 0. This is exactly what we noticed in the simple simulation in table 8. If there is a(n almost) one-to-one correspondence between evaluation states and states of affairs, literal meanings become very good at conveying highly specific affective meanings, even when they are extremely improbable. Figure 2 shows that once the standard deviation exceeds some value (thereby creating overlap in which evaluation is associated to which state of the world) the model starts interpreting the sentence hyperbolically. The probability of a hyperbolic reading drops once the noise becomes too much, because in that case the model spreads the probability mass also to even lower weights.

We see that in order for the model to predict hyperbolic interpretation, very specific values of γ and σ are required. First of all, only for values of $\gamma \leq 0.25$ does the model produce sensible predictions. It is an empirical question whether power functions of this kind are appropriate to describe the connection between weight and what weight feels like. There are some hints that this may not be so. In his seminal magnitude estimation studies, Stevens (1957) found $\gamma = 1.45$ for weight. (See the red line in figure 2.) Second of all, the band of possible standard deviations that triggers appropriate interpretation is very narrow. So, experiments would somehow need to establish that the human evaluation of states of affairs comes with this specific kind of noise.

For now, however, let us be charitable and assume that we can empirically

[21] Model details were as follows. In all simulations $\alpha = 2$, and the goal prior is uniform over the state, affect and combined goals. The non-affective space consists of 200 states, from 5kg to 1000kg in steps of 5kg, with a log-normal prior (median 15kg, log-SD 0.5). The affect space consists of 201 evenly spaced values between 0 and 1. For each state s, $P(e \mid s)$ is a discretised normal distribution over these values with mean $f_\gamma(s)$ and standard deviation σ. There are four utterances: 'This suitcase is light' (true of weights up to 10kg), 'This suitcase is heavy' (20kg or more), 'This suitcase is very heavy' (30kg or more) and 'This suitcase weighs a ton' (true only of the 1000kg state).

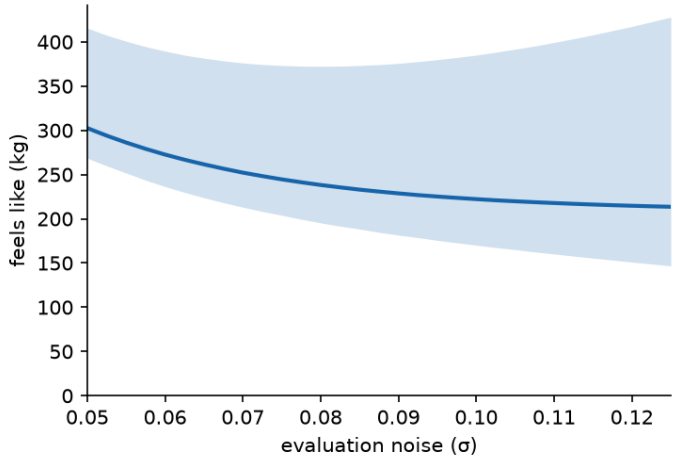


Figure 3: Predictions for affective interpretation. Evaluations are expressed as the weight of a suitcase that typically feels that bad, $f^{-1}(e)$. Line: median; band: central 95% of the posterior.

establish a $\gamma \leq 0.25$ with a $0.075 \leq \sigma \leq 0.15$ in an experiment setup suitable for the task of interpreting sentences like *'This suitcase weighs ...'*. Then, let us zoom in on the affective interpretations these kinds of settings trigger. Figure 3 shows the predicted affect for an utterance *'This suitcase weighs a ton'* for the specific case of $\gamma = 0$. What we can observe in this plot is exactly the argument I made against the domain mapping approach of Peña-Cervel and Ruiz de Mendoza Ibáñez (2022). When the model predicts a weakening of the literal meaning (*'This suitcase weighs a ton'* is understood as the suitcase being very very heavy), the affective meaning that accompanies that weakening is not itself sufficiently weakened: the evaluation of the suitcase is the evaluation that goes with a (still improbable) suitcase weighing around 220kg. In other words, the evaluative dimension of meaning is still hyperbole.

7 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. Consider the following example:

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

An utterance of (17) 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 (18).

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

This can mean 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 (17). Crucially, (18) 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 (18), 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.[22]

There are further reasons to question repair theories from the perspective of composition. Consider (19).

(19) 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 utterance ends up meaning that not everyone is late very often. How could this meaning come about?

For a repair theory to work for (19), 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 (19), we could 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

[22]Note that there are variations within the rational speech act framework that can account for implicature in embedded positions by implementing uncertainty in lexical semantics (Potts et al., 2015). To apply this to hyperbole, however, the lexica under consideration would have to include weakened meanings of the hyperbolic expression, which, as I argued in footnote 14, means that the hyperbolic meaning is stipulated in the prior over lexica rather than derived.

easily infer affect from the exaggerated 'Sue is always late'. But with (19) 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 (20). It still doesn't help us deriving the correct meaning.

(20) QUD α: { staff = good, staff = bad }
QUD β: {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, $P(\text{'staff = bad'} \mid \text{everyone is always late}) \approx 1$. But what about $P(\cdot \mid$ not everyone is always late)? This would need to be an almost flat distribution. Knowing that not everyone is always late, is not knowing very much. As such, the consequences of this claim for evaluating the staff are non-existent. The truth of this claim doesn't particularly swing the evaluation to the staff = good or the staff = bad side of things. 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.[23]

You might think that this is a good prediction if you can convince yourself that (19) is not a case of hyperbole. Perhaps (19) is some kind of negative strengthening (Horn, 1989), akin to (21), which is interpreted as expressing something stronger than just the denial of Sue's performance being very good, namely that is was rather bad:

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

Unfortunately, I know of no theory of negative strengthening that can be extended to capture the hyperbolic interpretation of (19), so this does not help much. Furthermore, there are cases of embedded hyperbole where negative strengthening seems unlikely:

(22) 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, (22) has a prior probability ≈ 1. Consequently, an account of the meaning of an utterance of (22) (that the speaker rings HR

[23] Note that, in this case, weakening the hyperbole's literal meaning provides a stronger meaning at matrix level.

whenever she hears about one of her employees being late very frequently) runs into the same problems as we discussed for (19).

An even greater problem occurs with examples like (23):

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

On its literal meaning this sentence does *not* have a prior probability of either ≈ 0 or ≈ 1. Since it will be extremely rare for an employee to (literally) never be on time, the literal meaning of (23) 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.

8 Towards a non-rational theory of hyperbole

As I explained at the start of this paper, rational repair theories are rational in a narrow sense. They involve reasoning about utterance utility, triggered by a lack of utility of the literal meaning. There are plenty of mechanisms in the literature that account for pragmatic phenomena without such a (narrowly) rational mechanism. One example is the account of scalar implicature as triggered by covert syntactic operators (Chierchia et al., 2012). This approach to scalar inference holds that alternatives-based implicature is integral to grammar and void of utility-based reasoning.[24] Other non-rational approaches think of the mechanism behind some pragmatic phenomenon as being further removed from language-specific modules. For example, Sperber and Wilson (2002) argue that it is unlikely that a lot of pragmatics involves conscious, reflective reasoning about what would have been rational for a speaker to do, given the sheer amount of reasoning that that would require for even a single utterance. Instead, they argue for the existence of intuitive heuristics that provide fast inferential procedures in a certain domain. In this section, I would like to invite the reader to consider such an automatic procedure for hyperbole, in direct comparison with the rational approach discussed above.

There are reasons to assume the existence of an intuitive procedure for dealing with exaggeration. In the animal world, stimuli that deviate from a norm have consistently been found to trigger automatic responses that at face value would not be expected. Such stimuli are known as *supernormal stimuli* (Tinbergen, 1953; Staddon, 1975) and the finding is that the brain tends to respond more strongly to a stimulus that is more extremely removed from a norm, despite the fact that its extremeness makes the stimulus unnatural. The

[24] That said, versions of this approach require disambiguation of alternative derivations and that disambiguation could in turn be a process that requires a rational mechanism (Franke and Bergen, 2020).

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 supernormal 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 supernormal 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 that deserves (extra) attention, and there is an intuitive procedure in place that readily infers the norm-deviation from the supernormal input. 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.

Whatever accounts for our effortless interpretation of unnaturalistic stimuli like caricatures, this could be the basis for an alternative, non-rational account of hyperbole. The idea would be as follows. Any hyperbolic utterance refers toi a multi-dimensional meaning that is norm-deviating in an extreme way (e.g. the speaker is carrying a one ton suitcase and is affected accordingly). Such a meaning counts as a supernormal stimulus. This triggers an automatic inferential response that treats the stimulus as if it were a stimulus that likewise was norm-deviating, but that was also naturalistic. One way to think about this is that any meaning that is compositionally derived passes through some module Ω. If the meaning is an extreme deviation of the norm, then Ω treats it as a supernormality. One way to make this more explicit is to describe the computational mechanism in Ω. For instance, norm-deviation could be thought of in terms of surprisal. For example, take a function Dev that assigns a norm-deviation value to a meaning as follows:

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

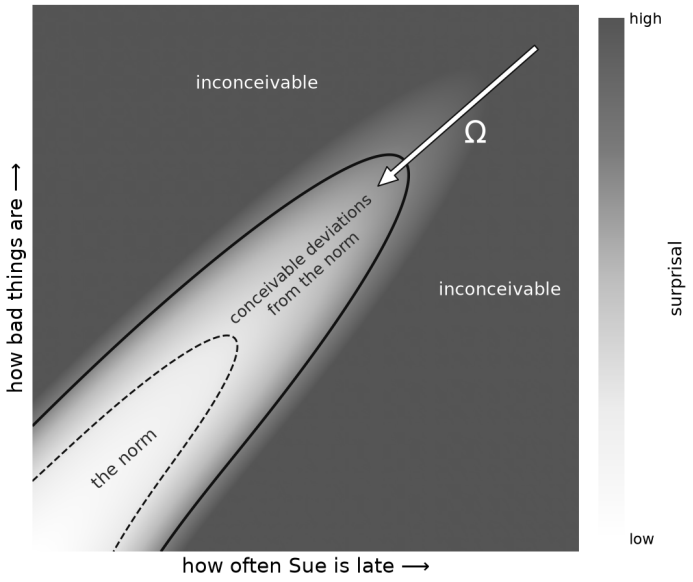


Figure 4: Illustration of how Ω maps extreme meanings to less extreme ones.

The supernomality module then simply adjusts meanings on the basis of extreme deviations. Let $C(m) = \{m' : Dev(m') < \theta \text{ and } d(m, m') \text{ is defined}\}$ be the set of conceivable meanings at some distance from m. Then:

$$\Omega(m) = \begin{cases} \text{argmin}_{m' \in C(m)}\, d(m, m') & \text{if } C(m) \neq \emptyset \\ m & \text{otherwise} \end{cases}$$

$\Omega(m)$ returns m whenever its deviation from the norm does not exceed θ, so, in other words, this is a procedure that only carries an effect with extreme stimuli. When the Dev of a meaning does exceed the threshold, then Ω returns the closest norm-deviating meaning to m that is not extreme. So, Ω will map the meaning of a hyperbolic utterance to the nearest meaning that is norm-deviating within the bounds of what is conceivable. Figure 4 illustrates how Ω works. The figure shows a two-dimensional meaning space combining frequency of Sue being late and an evaluation in terms of how bad things are. The shades indicate the surprisal associated to these meanings, based on a context where it is normal for people to be sometimes late, but not extremely often. (This is very similar to the priors we have seen for the RSA models.) Given some fixed threshold θ, Ω now maps outlandish meanings to closely related meanings that are sufficiently conceivable. Note that the idea is that meanings pass through Ω across the board. This is not an operator that may or may not apply, triggering reasoning about whether or not it does. This is simply a fast mechanism that automatically applies.

The definition of Ω presupposes the presence of a scale, by imposing there to be a usable notion of scalar distance d: $d(m, m')$ is defined only if m and m' differ in degree along the same scales. Meanings that differ categorically (e.g. p versus $\neg p$) are not at any distance from each other. If there is no suitable notion of scalar distance, then Ω just returns its input. This reflects that the supernormal procedure does not apply to just any outlandish situation. Take for example an utterance of (24) out of the blue:

(24) I won the lottery!

According to any reasonable prior, this statement is unlikely to be true. As such, one could think that this is reason to interpret an utterance of (24) hyperbolically. However, someone who actually wins the lottery can perfectly well express this, using the literal meaning of (24). The reason for this is that there is no scale that divides the meaning space up in gradations of norm-deviation. Basically, if you win the lottery you deviate from the norm and if you do not, then you do not. Supernormality requires the norm-deviation to be exaggerated, but this is simply not possible in this situation, since there is no scale. The notion of distance does not make sense in this case. Note that as soon as there *is* a more fine-grained, scalar norm, hyperbole resurfaces. For instance, if I ask *'How lucky in love have you been?'* and someone answers (24), then this answer (metaphorically) points to an extreme deviation from the luck norm and we infer that the speaker's luck in love is considerable (though not as outlandish as the luck needed to win the lottery).

The mapping brought about by Ω has 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). I am assuming here that the literal interpretation of an utterance points to a multi-dimensional meaning, namely the proposition that is expressed combined with the evaluative / affective inferences that are associated to that proposition. (This is likely to be a many-dimensional meaning, but for simplicity, I'll keep things at two dimensions.) So, the literal intepretation of 'Sue is always late' points to a multi-dimensional reading like the following (where I represent multi-dimensional meaning by combining propositions with the operator $\circ$):

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

This meaning is the top-right corner of figure 4 and it is shifted to a more conceivable deviation of the norm by applying Ω:

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

Since Ω 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 Ω brings is simultaneously applied to several dimensions, not just in the domain of the literal meaning, but also in the connected affective (or other secondary) domain. So, quite straightforwardly, Ω avoids the issues discussed in section 4.

Can the weakening function also help with cases where hyperbole is embedded? Let us first consider (25) and (26).

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

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

Let's say that the interpretation of (25) is the two-dimensional meaning that combines the proposition that the number of things the speaker has to do is 1,000,000, 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 interpretation of (26). However, given that the embedded meaning is a supernormal meaning, Ω will weaken it. This is possible, because Ω is not an utterance-level operator. It is not even a linguistic operator. It is a meaning adjustment mechanism. The belief ascribed to Sue is a supernormal belief and, so, the end result of interpreting (26) will be that Sue will be said to believe the Ω of that supernormality.

In the case of (26), composition involved embedding a full clause. What happens in cases where hyperbole involves sub-clausal structures? The first step is to recognise that Ω can apply to compositional meanings. 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.[25] 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 Ω as follows:

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

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

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

(27) Not everyone is always late.

(28) [Not everyone] λx [(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 things 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, as I said before, 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).

Note that a non-rational account does not help us understand examples of considerable norm-deviation where no hyperbolic interpretation surfaces. All else being equal, if hyperbole is a general automatic inference procedure, then we would expect hyperbolic interpretation for both Feinmann's (2) and (3).

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

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

As discussed above, it seems intuitively much harder to get a hyperbolic reading of (3) compared to (2). This is unexpected on the non-rational account just outlined.[26] If having two hundred unread emails is an inconceivable deviation from the norm, then so is having exactly two hundred unread emails. As I suggested above, perhaps this utterance is subject to some kind of manner implicature: the presence of *exactly* signals that the meaning one would associate with the simpler alternative in (2) is not the intended one. Somewhat ironically, then, the reason why hyperbole is harder for (3) is that it invites the application of a rational mechanism that I have argued to be absent for utterances that *are* interpreted hyperbolically. Note, however, that such an inference

[26]Note that the rational repair theories discussed above equally lack an explanation of this contrast.

cannot block Ω, which applies across the board. What it can do is change the input to Ω. A speaker who says *exactly two hundred* presents herself as having counted. This is evidence that the literal meaning is not outlandish after all: it raises $P(m)$, so that $Dev(m) < \theta$ and Ω returns m. In other words, the manner inference is an inference about the norm, not about the output of Ω. I leave it to a future occasion to investigate more precisely what kind of interpretations examples like these receive and how these can be reconciled with the approach outlined here.

Despite the many questions I am leaving unanswered about this account, the potential of an account along these lines should be clear. If hyperbole interpretation is not the effect of a general utility-based reasoning, but rather the automatic response to a supernormal meaning, then we readily resolve the issues the rational account ran into.

9 Conclusion

This paper has argued that hyperbole is not a rational speech act. I have demonstrated that a theory that relies on a rational repair mechanism fails to account for how exaggeration is interpreted in systematic ways. These issues suggest that hyperbole interpretation is not the result of a rational utility based reasoning. A promising alternative is to see the particularities of hyperbole interpretation as the result of a general, non-linguistic inferential procedure for how to deal with extreme deviations from a norm.

References

Bergen, L., Levy, R., and Goodman, N. (2016). Pragmatic reasoning through semantic inference. *Semantics and Pragmatics*, 9:84pp.

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

Chierchia, G., Fox, D., and Spector, B. (2012). Scalar implicature as a grammatical phenomenon. In *Handbücher zur Sprach-und Kommunikationswissenschaft/Handbooks of Linguistics and Communication Science Semantics Volume 3*. de Gruyter.

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. (2026). Is hyperbole a scalar inference? *Journal of Logic, Language & Information*, pages 1–31.

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

Franke, M. and Bergen, L. (2020). Theory-driven statistical modeling for semantics and pragmatics: A case study on grammatically generated implicature readings. *Language*, 96(2):77–96.

Franke, M., Dablander, F., Bennett, E., Degen, J., Tessler, M. H., Kao, J., Goodman, N. D., et al. (2016). What does the crowd believe? a hierarchical approach to estimating subjectivebeliefs from empirical data. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 38.

Geurts, B. (2006). Take five: the meaning and use of a number word. In Vogeleer, S. and Tasmowski, L., editors, *Non-definiteness and plurality*, pages 311–329. Benjamins, Amsterdam.

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.

Goodman, N. and Frank, M. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, 20(11):818–829.

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.

Potts, C., Lassiter, D., Levy, R., and Frank, M. C. (2015). Embedded implicatures as pragmatic inferences under compositional lexical uncertainty. *Journal of Semantics*, 33(4):755–802.

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*.

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.

Stevens, S. S. (1957). On the psychophysical law. *Psychological review*, 64(3):153.

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.

Wilson, D. and Carston, R. (2007). Metaphor and the 'Emergent Property' Problem: A Relevance-Theoretic Approach. *Baltic International Yearbook of Cognition, Logic and Communication*, 3(1).