The Shape of Modified Numerals

Fausto Carcassi (fausto.carcassi@gmail.com)
Institute for Logic, Language and Computation
Department of Linguistics, University of Amsterdam
Spuistraat 134, 1012 VB Amsterdam

Jakub Szymanik (jakub.szymanik@gmail.com)
Institute for Logic, Language and Computation
Department of Linguistics, University of Amsterdam
Spuistraat 134, 1012 VB Amsterdam

Abstract

The pattern of implicatures of modified numeral 'more than n' depends on the roundness of n. Cummins, Sauerland, and Solt (2012) present experimental evidence for the relation between roundness and implicature patterns, and propose a pragmatic account of the phenomenon. More recently, Hesse and Benz (2020) present more extensive evidence showing that implicatures also depend on the magnitude of n and propose a novel explanation based on the Approximate Number System (Dehaene, 1999). Despite the wealth of experimental data, no formal account has yet been proposed to characterize the full posterior distribution over numbers of a listener after hearing 'more than n'. We develop one such account within the Rational Speech Act framework, quantitatively reconstructing the pragmatic reasoning of a rational listener. We show that our pragmatic account correctly predicts various features of the experimental data.

Keywords: modified numerals; more; rational speech act;

Introduction

Traditional pragmatics mostly limited itself to qualitative accounts of implicatures, in which an utterance in a context implicates some propositions, excluding or including some possible world states. For instance, 'most cats knit' implicates that it is not the case that all cats knit. In the last twenty years, new experimental and statistical methods have been applied to capture subtler patterns in speaker behaviour (see e.g. Cummins and Katsos (2019) for an overview). In particular, the development of Bayesian cognitive models of pragmatic language use and sophisticated experimental designs have allowed researchers to test more fine-grained hypotheses about graded notions of implicature (Franke & Bergen, 2020).[1] In this picture, rather than a qualitative difference between states pragmatically compatible or incompatible with an utterance, a pragmatic listener has a full prior over states which is updated after receiving the utterance. The semantic and pragmatic content of the utterance contributes to the listener's estimated probability of each possible state, allowing for a graded and quantitative notion of compatibility between an utterance and a possible state.

As a case study in this approach to pragmatics, in this paper we look at modified numerals, i.e. expressions such as 'more than 3'. Modified numerals usually convey information about the cardinality of the intersection of two sets. For instance, 'about 4 Frenchmen yawn' conveys that the cardinality of the intersection of the set of Frenchmen and the set of yawning things is not far from 4. Examples of modified numerals are 'at least 4', 'exactly 1', 'more than 3'. In this paper, we focus on the latter expression: 'more than n' (for some integer n). We develop an account of the shape of the posterior distribution over numbers of a language user upon hearing an expression containing a modified numeral.

While the meaning of 'more than n' might at first appear straightforward, the usage of the expression is in some respects puzzling. First, as noticed already in Krifka (1999) and experimentally confirmed in Geurts (2010), the standard Horn account is at odds with the behaviour of modified numerals. Specifically, modified numerals do not seem to elicit some of the predicted scalar implicatures, e.g. 'more than 3' does not seem to implicate 'not more than 4'. Second, as discussed in Cummins et al. (2012) the implicatures drawn from 'more than n' are influenced by the *roundness* of n. For instance, 'more than 10' seems to implicate 'not more than 20', although not 'not more than 13'. This might be a consequence of '20' being more round than '13'. Third, as discussed in Hesse and Benz (2020), the range of numbers for which 'more than n' is used is influenced by the *magnitude* of n. Specifically, the greater the magnitude of n, the greater the range of numbers above n to which 'more than n' still probably applies. For instance, 'm is more than 10' prima facie would not be used when $m = 1010$, but 'm is more than 25000' seems appropriate for $m = 26000$, although the difference between m and n is the same in the two cases.

The three patterns discussed above constitute qualitative differences between the implicatures induced by various modified numerals. However, the three discussed factors, among others, follow from the full posterior distribution over numbers induced in a listener after hearing 'more than n': $p(\cdot|\text{Speaker uttered 'more than } n\text{'})$, which we abbreviate to $p(\cdot|MTn)$. In other words, $p(\cdot|MTn)$ describes the probability that the listener attributes to m being each number after hearing 'm is more than n' (with n known and m unknown).

$p(\cdot|MTn)$ could prima facie take various shapes. For instance, $p(\cdot|MTn)$ could approximate a discrete approximation of a (lower-truncated) normal distribution with mean m, where $m > n$ and the distance $m - n$ could depend on the

[1] For lack of conventional and apt terminology, in the following we will use slightly abuse the term 'implicature', using it for phenomena that might not traditionally count as such. We will for instance speak of an 'accumulation of implicatures'. The context and model will clarify what is meant in each case.

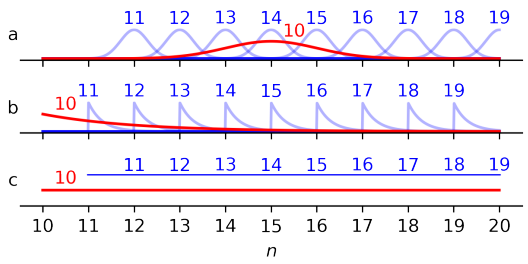

Figure 1: Three possible usage distributions patterns for modified numerals 'more than 10' to 'more than 19' (the distributions are over integers albeit showed as continuous for ease of visualization). Each line corresponds to a different utterance of the form 'more than n', and shows the posterior distribution of a listener after hearing the signal. The n is shown above each distribution's peak. Blue lines are used for the numerals of lowest roundness level, red line is used for 'more than 10', which is more round.

roundness of n and its magnitude (option a in Figure 1). A second option is that $p(\cdot|MTn)$ could resemble an discretized exponential distribution shifted to start at n, with a variance that again depends on granularity and magnitude of n (option b in Figure 1). A third option is that $p(\cdot|MTn)$ looks like a discrete uniform distribution from n to m, with m depending on granularity and magnitude of n (option c in Figure 1).[2]

The three categorical implicature effects we discussed are compatible with all three options from Figure 1. First, for all three options 'more than n' can in general cover $n+2$, implying that 'not more than $n+1$' is not implicated. Second, in all options 'more than 10' can behave differently from 'more than n', with n less round than 10. Lastly, in all three options the relevant parameters can be sensitive to the n's magnitude. This demonstrates that accounting for categorical implicatures is not enough to fully characterize the information conveyed by modified numerals.

Little works has been done to predict, characterize, or describe this distribution as a result of pragmatic reasoning.[3] The main aim of this paper is to propose a quantitative model of $p(\cdot|MTn)$ which can account for the previously discussed qualitative patterns as well as experimental data in previous literature, and that can generate novel empirical predictions. We start by discussing previous literature more in detail, and we identify how previous accounts fall short of a full characterization of a listener's understanding of 'more than n'. Then, we present a model in the Rational Speech Act framework, and show how it can account for previously puzzling phenomena.

[2] As we discuss below in more detail, experimental work in Hesse and Benz (2020) shows that option b, with some caveats, is in fact closest to how 'more than n' is actually interpreted.

[3] Hesse and Benz (2020) give a partial characterization which we discuss below. Moreover, Benz (2015) gives an account of implicature patterns of 'more than n' for round ns which is some ways similar to the one presented in this paper, but stops short of characterizing the whole listener's posterior.

Previous literature and data

Granularity-based approaches

Early work on 'more than n' focused on roundness as a possible factor to explain the unusual implicature patterns discussed above. The concept of roundness can be analysed in terms of the concept of scale. Scales consist of the set of multiples of certain numbers, e.g. 5, 10, 50, 100, which are particularly cognitively simple.[4] One scale is more granular than another if it divides the number scale in points that are closer together. The roundness of a numeral can then be thought of as the level of the least granular scale that the numeral belongs to. As an example, consider 30 and 200. 200 belongs to many scales—e.g. the ones containing the multiples of 1, of 2, of 10—but the least granular scale it belongs to is arguably the scale of multiples of 100. On the other hand, the least granular scale 30 belongs to is that of the multiples of 10. 200 is therefore more granular than 30 because the former belongs to a scale that is less granular than the most granular scale where the latter figures.

Cummins et al. (2012) argue that roundness plays a role in the pattern of implicatures of modified numerals.[5] For instance, 'more than 1000' lacks the implicature 'not more than 1001'. In order for the implicature to be calculated, the listener would have to assume that, had the speaker observed e.g. 1002, they would have said 'more than 1001'. However, 1000 is rounder than 1001, and therefore uttering 1001 comes at an additional cognitive cost compared to 1000. For the speaker, the additional cognitive cost is too great for the little additional information conveyed by uttering 'more than 1001'. The listener cannot therefore infer that the speaker would have said 'more than 1001' had they observed a state (e.g. 1002) for which 'more than 1001' would have been only slightly more informative than 'more than 1000'.

Crucially, Cummins et al. (2012) note that the same argument does not apply when the implicated sentence contains a numeral at the same roundness level as n. For instance, after observing 2300 a speaker would rather utter 'more than 2000' than 'more than 1000', everything else being equal, because the two involved numerals are cognitively equally costly, but the former utterance is more informative. Therefore, the listener would have reason to infer that if the speaker uttered 'more than 1000', they did not observe a state for which 'more than 2000' applies. More generally, 'more than n' should generate a scalar implicature to 'not more than m',

[4] Note that not all possible scales are used to determine roundness. For instance, 202 is less round than 200, despite being divisible by a number, 101, which is greater than the greatest divisor of 200, namely 100. This indicates that the scale of multiples of 101 does not play a role in determining roundness. We do not develop an account of which scales influence roundness, but rather rely on previous characterizations from the literature.

[5] Cummins (2013) give a more theoretically grounded and precise account of the implicature patterns for modified numerals based on Optimality Theory (Prince & Smolensky, 2008). While these previous accounts include more phenomena than are discussed here, they do not provide a full characterization of $p(\cdot|MTn)$. Therefore, here we focus on the experimental results.

where m is the next numeral at the same roundness level of n. Moreover, Cummins et al. (2012) point out that, for similar reasons, 'more than n' should also implicate 'not more than m' for any m at a higher roundness level than n. For instance, 'more than 90' is predicted to implicate 'not more than 100'.

Based on these arguments, Cummins et al. (2012) makes two experimental predictions. First, the rounder the n, the higher responders' estimates will be compared to n. Second, in the range condition typical estimates will be of the form '$n+1$ to m', where m is the value after n with the same granularity as n or higher.

Cummins et al. (2012) then present an experiment to test the two prediction. Participants ($n = 1200$) were presented with 16 contexts. The following is one such context (varying by condition as indicated):

> **Information** A newspaper reported the following. "[Numerical expression] people attended the public meeting about the new highway construction project."
>
> **Question** Based on reading this, how many people do you think attended the meeting?
>
> Between _____ and _____ people attended [range condition].
>
> _____ people attended [single number condition].

The numerical expression consisted of a quantifier (either 'more than' or 'at least n') and a numeral belonging to one of three levels of granularity (multiples of 100, multiples of 10 but not 100, and non-round such as 93).

Overall, the results confirmed the two experimental predictions. The range of interpretation increases with the roundness of the numeral. Moreover, most responses in the range condition were as predicted. For instance, 'more than 100' typically conveyed an upper bound at 150.

Hesse & Benz (2020)

Hesse and Benz (2020) consider two empirical predictions from Cummins et al. (2012). First, the rounder the n in 'more than n', the wider the range of potential values. Second, the rounder the n, the further away from n will be the single most likely value. In a series of experiments, Hesse and Benz (2020) test these two predictions for a wider range of numerals than Cummins et al. (2012), and find that they are not borne out.

The first experiment identifies some domains for which participants do not have strong prior beliefs as to the size of the involved numerals. In such contexts, prior belief does not play a strong role in the estimation of the numerals, and therefore the effect of roundness and magnitude can be isolated from other prior factors. Four such domains are identified: petition signatures, audience size at a music concert, votes in an election, spectators at a sporting event.

The second experiment is a replication of the experiment in Cummins et al. (2012) for a wider range of numerals at four different roundness levels (50, 90, 93, 100, 110, 130, 150, 200). Results do not corroborate the two predictions based on Cummins et al. (2012): the (median) distance between n and the guessed number in the single number condition does not increase the rounder the n is, and neither does the (median) range in the range condition.

In the third experiment, numerals of a wider range of magnitudes are tested (20, 30, 40, 60, 70, 80, 120, 140, 160, 170, 180, and 190) and only the four contexts identified in experiment 1 are used. In the combined data from the second and third experiments (as well as in the data from the second experiment alone), magnitude is a stronger predictor of the range of produced values than roundness. The second and third experiments in Hesse and Benz (2020) fail to find evidence for two the effects predicted by Cummins et al. (2012). However, two different patterns emerge. First, the median numbers in the single number condition are a constant distance of 10 above the modified numeral. Second, participants tend to guess numbers with an upper bound located at the next round number above the modified numeral. For instance, when presented with 'more than 120', 'more than 130', or 'more than 140' participants tend to guess numbers up to 150 (a round numeral). This produces a 'squeezing' effects for modified numerals immediately below a round number. Both patterns can be observed in the left plot in Figure 2.

In the fourth and last experiment, Hesse and Benz (2020) focus on larger numerals. They administer the same task, with 6 contexts (number of signatures on a petition, size of the audience at a music concert, turnout at an election, number of spectators at a sporting event, size of a meeting, and budget for a reception) and larger numerals (1k, 1.1k, 1.4k, 15k, 16k, 19k, 20k, 21k, 24k, 25k). Like in the previous experiments, greater roundness does not *per se* cause a greater range of guessed numbers or wider ranges in the range condition. Moreover, the two new patterns noticed in the previous experiments persist, but scale proportionally to the magnitude of the involved numerals. While in the 1-100 range the median guessed number was around 10 above the modified numeral, in the range of thousands it is 100 above, and in the range of tends of thousands it is 1000 above. While the upper boundaries participants tend to select are at the roundness level of multiples of 50 or 100 in the 100 interval, they are multiples of e.g. 500 in the 1000 range and multiples of 5k in the tens of thousands range. Both effects can be seen in the right plot in figure 2.

Hesse and Benz (2020) also give a characterization of their data in terms of a boundary function, and they propose an explanation for the fact that larger ns lead to guesses with a proportionally greater variance. The explanation relies on the Approximate Number System (ANS), namely the cognitive mechanism that underlies the approximate perception of magnitudes. When using the ANS, numbers are not encoded precisely, but rather as distributions over numbers. Moreover, the variance of this distribution increases with the magnitude of the number. Hesse and Benz (2020) argue that, as the modified numeral gets larger, participants will associate the numeral with increasingly wide distributions, and therefore the

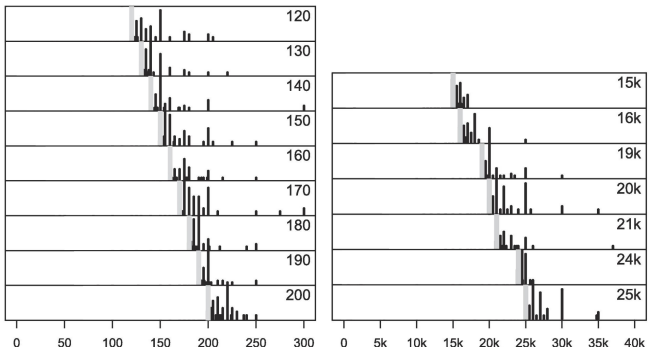

Figure 2: Some participants' responses from the single numbers condition in experiment 2 (left, low magnitude) and 3 (right, large magnitude) in Hesse and Benz (2020). Responses for 'more than n' are shown on the right side of the grey line, and n is shown on the top right of each subplots.

spread of their guessed numbers will also increase.

The account developed in Hesse and Benz (2020) paints a clear picture of the patterns in production for modified numerals of the form 'more than n' and 'less than n'. However, the paper does not give a full model of the way a listener calculates a posterior over numbers. In order to evaluate their proposal quantitatively, more detail would be needed for the implementation, specifically concerning the relation between the ANS component of their account and roundness. For instance, a bare ANS account alone leaves unexplained why participants tend to produce signals at higher levels of roundness when dealing with larger numbers, rather than simply producing from a distribution with greater variance. Among the one thousand numerals guessed in the fourth experiments for $n \geq 15000$, only 88 were at a level of roundness lower than 500. As we argue in the next section, a simple model of recursive mindreading can explain various patterns in the data.

A Unified Model

The Rational Speech Act Framework

The RSA framework is meant to model the process of recursive mindreading that lies behind the pragmatic interpretation or production of utterances (Frank & Goodman, 2012; Goodman & Frank, 2016; Franke & Jäger, 2016). RSA models usually start with a pragmatic listener who interprets utterances based on the simulated behaviour of a pragmatic speaker. The pragmatic speaker in turn given an observation tends to choose the most useful utterance for a literal listener who interprets it based solely on its literal meaning. We will first explain the simplest type of RSA model, and then a modification that will be useful to model modified numerals.

The simplest RSA model starts with a set of utterances u and a set of possible states s. The meaning of each utterance can be encoded as the set of those states that verify the utterance. The pragmatic listener L_1 receives an utterance u and

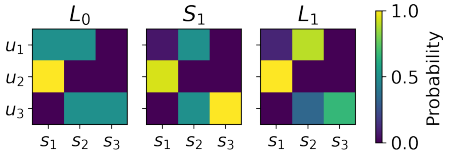

Figure 3: Simple RSA model with three possible utterances u (y-axis) and three states s (x-axis). L_1 calculates a scalar implicature for utterances u_1 and u_2 ($\alpha = 4$). The left, central, and right plots correspond to L_0, S_1, and L_1 respectively. The color indicates the probability of guessing a state given a signal for L_0 and L_1, and the probability of producing a signal given a state for S_1.

calculates a posterior over states by Bayesian update, combining their prior over states with the probability that the pragmatic speaker S_1 would have produced the utterance given each state:

$$p_{L_1}(s|u) \propto p_{L_1}(s)p_{S_1}(u|s) \tag{1}$$

The pragmatic speaker in turn observes a state and produces an utterance with a probability that depends on a cost-related salience utterance prior (see Chapter 3 in Scontras, Tessler, and Franke (2018)):

$$p(u;C) \propto \exp(-c(u)) \tag{2}$$

and on the utility $U(u|s)$ for a literal listener L_0 given the state:

$$p_{S_1}(u|s) \propto \exp(\alpha U(u|s))p(u;C) \tag{3}$$

where α is the speaker's rationality parameter: the higher the value of α, the more the speaker's distribution will be peaked at the most useful utterances. The utility $U(u|s)$ is the negative surprisal of the state given the utterance, so that the speaker favours utterances that make the state less surprising for the literal listener:

$$U(u|s) = \log(p_{L_0}(s|u)) \tag{4}$$

Finally, the probability that literal listener L_0 attributes to each state given an utterance is simply 0 if the utterance is not verified by the state, and proportional to the prior for the state otherwise:

$$p_{L_0}(s|u) \propto \begin{cases} p_{L_0}(s) & \text{if } s \text{ verifies } u \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

Figure 3 shows L_0, S_1, and L_1 in this simple RSA model. The crucial phenomenon that can be observed in figure 3 is that L_1 calculates a scalar implicature: although utterance u_1 is, in its literal sense, compatible with both s_1 and s_2, S_1 tends to produce u_1 mostly for s_2, because when s_1 is observed S_1 tends to use the more useful signal u_1. Therefore, when hearing u_1 L_1 is more likely to guess s_2.

A Simple Model of Modified Numerals

We make a few simple changes to the basic RSA model above. First, rather than the 3 states in the toy model above, the states in our model are the integers up to 10000, and the signals all the expressions 'more than n' with $0 \leq n \leq 10000$.

Second, we let cost depend on the roundness level in a way consistent with Hesse and Benz (2020)'s measure of roundness, itself based on the measure in Cummins et al. (2012) and inspired by Jansen and Pollmann (2001). Specifically, we calculate cost as the inverse rank of roundness, in the following way:

Least Granular Scale	Cost $c(n)$	Least Granular Scale	Cost $c(n)$
1000	0	50	4
500	1	10	5
200	2	5	6
100	3	1	7

For instance, according to this measure 3000—whose greatest divisor in the table is 1000—gets cost 0, while 350—whose greatest divisor in the table is 50—gets cost 4. The relation between roundness level and cognitive cost is confirmed (for a different but related construal of roundness) by Solt, Cummins, and Palmović (2017), which gives experimental evidence that the level of granularity influences cognitive complexity of the expressions.

Third, we specify a prior for both the literal listener, p_{L_0} and the pragmatic listener, p_{L_1}. For simplicity, we assume that the two priors are identical. This corresponds to the default assumption that the literal listener assumes that the pragmatic speaker has an accurate representation of the listener's prior. Moreover, we assume for simplicity that this prior has a geometric distribution.

Lastly, we assume that when higher numerals are mentioned, the distribution over the true state covers higher magnitudes. For instance, upon hearing 'more than 4 people came to the concert', I attribute small variance to the number of people that possibly came, and upon hearing 'more than 10k people signed the petition', I attribute large variance. Note that this is independent of the roundness effect: upon hearing 'more than 10004 people came to the concern', I still attribute a priori high variance to audience size. Rather, this could be an effect of ANS, as argued in Hesse and Benz (2020), or an effect of knowledge about the underlying data-generating process. This latter explanation is particularly plausible for the items used in Hesse and Benz (2020), and provides an alternative to their ANS account. The number in most of the contexts shown to participants, e.g. the size of the audience at a music concert, are approximately Poisson distributed. This implies that the larger the magnitude of the number, the greater the plausible range of the number. The ANS account and the generative-model account make different predictions for cases where the variance does not increase with the mean of a random variable, which could be tested experimentally.

Practically, in the model the listeners' expectation about n depends on the modified numeral n itself. Specifically, listeners in the model distinguish two types t of events, $T = \{\text{small}, \text{large}\}$, e.g. small and large concerts.[6] Listeners have one prior distribution over the true state for small events—with the geometric distribution's $k = 0.007$—and another distribution over states for large events—$k = 0.0002$ (Figure 4). We assume that when n is small, the listener concludes that the event is small, and when the n is large, the listener concludes the event is large, and use the respective prior over s. Formally, $p(t = \text{large}|u) = 1$ iff u contains a number in the hundreds or smaller, and $p(t = \text{small}|u) = 1 - p(t = \text{large}|u)$. Equation 1 then becomes:

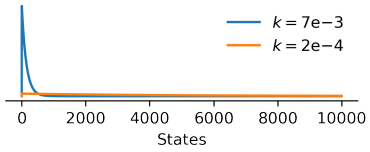

Figure 4: Prior for small (blue) and large (red) numerals.

$$p_{L_1}(s|u) \propto \sum_{t \in T} p_{L_1}(t|u) p_{L_1}(s|t) p_{S_1}(u|s) \tag{6}$$

The results of these modifications for small n can be seen in figure 5.[7] Some crucial features of the data are predicted. First, as observed in the data in Hesse and Benz (2020), the distribution for 'more than n' resembles option b in Figure 1. The reason for this, which as far as we are aware has not been discussed in the literature, is an accumulation of very weak scalar implicatures. If the listener hears 'more than n' and consider whether the true state is m, then they reason that for all j such that $n < j < m$, the speaker had to choose to not utter j, since all expressions 'more than j' would also be true. Therefore, the greater the number of js, the less plausible it is that the true state is m.

A second feature in the data that our model correctly models is the spread of the distribution as a function of the numeral's roundness. In partial agreement with Cummins et al. (2012), some round numbers have a greater variance than less round numbers. For instance, 'more than 0' is predicted to have greater variance than 'more than 20'. However, contra Cummins et al. (2012) and consistently with Hesse and Benz (2020), the effect is small for all numbers except 0 in the model.

A third feature in the data captured by our model is the squeezing effect. As the signal approaches a round number from below, the probability mass becomes more concentrated between n and the round number. This is clear in the right plot in Figure 5 for numerals approaching 150 and 200. The effect also exists, but to a lesser extent, for numerals of lower round-

[6] In a more sophisticated and realistic version of the model, rather than few event types there would more fine grained variation.

[7] `https://github.com/thelogicalgrammar/modifiedNumerals` contains all the code needed to reproduce the results.

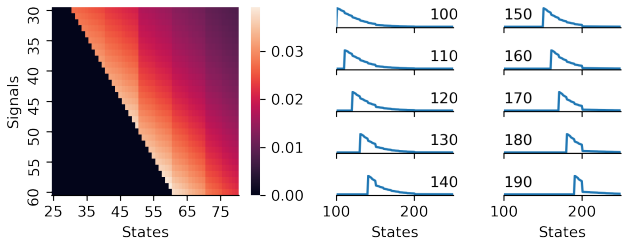

Figure 5: Left: listener's posterior probability $p(\cdot|MTn)$ over numbers, given each signal. The squeezing effect can be visualized as the increasing concentration of posterior mass approaching round numbers from below ($\alpha = 7$). Right: $p(\cdot|MTn)$ for n at intervals of 10. The squeezing effect—the posterior distribution concentrating between n and the closest round number above—is particularly clear for $n = 140, 190$.

ness levels. As already described in Cummins et al. (2012), the effect is a consequence of an implicature. If the true number had been higher than a round numeral higher than the one the speaker chose, the speaker would have chosen the higher round numeral instead. Therefore, the number has to be lower than any round numeral above the one actually chosen by the speaker.

While the simple pragmatic listener model can account for some feature in the experimental data, it differs in a crucial way from the participants in the experiment. Namely, the RSA pragmatic listener is a pure listener, while the participants were asked to produce a guess, and therefore are also in a certain way speakers. In order to make predictions comparable to the experimental data presented in Hesse and Benz (2020), we also propose a simple way that the listener might use their posterior over states given a modified numeral to give a response in the experiment. In our participant model, listeners tend to produce states that have a high probability, with an additional utterance prior against producing signals with high cost:

$$p(\text{Producing } m \mid u) \propto \exp\left(\rho \log(p_{L_1}(m \mid u)) - c(m)\right)$$

where u is the utterance shown to the participant and ρ is the softmax (inverse temperature) parameter, encoding a tendency of the listener to select the signal with the highest posterior probability. This model of participant production reflects the speaker model of pragmatic RSA agents, for a listener who is asked to make a guess as to the world state. Figure 6 shows the predicted production probabilities for a participant in the single number condition. The predictions resemble the observed data shown in the left plot of Figure 2, both in terms of which numerals are produced and more generally in terms of the shape of the produced numerals.

Figure 7 shows the results for greater magnitudes (4k-5k interval), using the prior for large numbers described above. With a higher variance prior, the listener infers a greater possible range for each numeral. The figure shows that the assumption of a prior which increases with the magnitude of

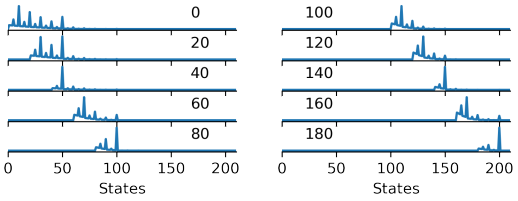

Figure 6: Predicted production probabilities for a simple production model based on an RSA pragmatic listener L_1 ($\alpha = 7, \rho = 3$). The production model correctly describes the participants' tendency to guess a round number above the observed number.

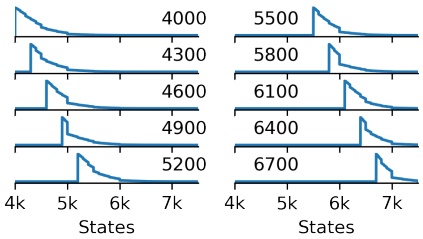

Figure 7: Results for large numerals ($\alpha = 10$). The squeezing effect is stronger for numerals of higher roundness (e.g. 5000) than lower roundness (e.g. 4500).

the numeral can explain the increasing ranges of guesses for greater numerals observed in the data (right plot in Figure 2).

Conclusions

Overall, our account provided a quantitative account of some aspects of the way participants understand the modified numeral 'more than n'. Various feature of the data are predicted by our model. First, and most importantly, the general shape of the guesses are predicted to be similar to option b in Figure 1, which is consistent with the data in Hesse and Benz (2020). Second, the model captures the patterns of implicature discussed at the beginning of the paper: the prima facie surprising lack of strong implicatures for successive numerals and the dependency of implicatures on roundness and magnitude. Third, a simple extension of the model captures some crucial features of the production behaviour of participants in experimental data. The model predicts the squeezing effect below round numerals (Figure 5), an addition of a preference for producing round numerals predicts the observed production patterns (Figure 6), and a prior with a larger variance for the listeners produces the observed change in production range for greater numerals 7).

The work in this paper could be extended in various ways. First, a Bayesian statistical model can be developed to fit the data in Hesse and Benz (2020), and Bayesian model comparison can be used to compare our account to the ANS account. The account proposed in Hesse and Benz (2020) is not a quantitative account of production, and therefore it would

have to be extended to directly predict experimental data.

Second, the model could be extended to include more modified numerals. For instance, Hesse and Benz (2020) also includes modified numeral 'less than'. Other modified numerals that could be modelled are 'at most n' and 'at least n', which has been shown to differ from 'less than $n+1$' and 'more than $n-1$' in interesting ways (Spector, 2020).

Third, a simplifying assumption in the model above is that we only used two priors, one for small numerals and one for large numerals. However, it is more plausible that the prior changes continuously as a function of the numeral. Future work can explore the functional relation between the prior parameters and the magnitude of the modified numeral.

The model also makes some assumptions and predictions beyond the data in Hesse and Benz (2020) that could be tested experimentally. For instance, it assumes that listeners end up with a full posterior distribution over numbers after hearing a modified numeral. However, it is not obvious that language users would possess representations of this kind, especially over a infinite set of numbers. Moreover, a listener would in principle need to calculate implicatures over an infinite set of possible utterances, which is implausible from a processing point of view. While we see the RSA model as a computational level model, these considerations should be taken into account if the gap to the algorithmic level is to be bridged.[8] We leave all these possibilities to future work.

References

Benz, A. (2015). Causal Bayesian Networks, Signalling Games and Implicature of 'More Than n'. In H. Zeevat & H.-C. Schmitz (Eds.), *Bayesian Natural Language Semantics and Pragmatics* (pp. 25–42). Cham: Springer International Publishing. doi: 10.1007/978-3-319-17064-0_2

Cummins, C. (2013). Modelling implicatures from modified numerals. *Lingua*, *132*, 103–114. doi: 10/ghspkr

Cummins, C., & Katsos, N. (Eds.). (2019). *The Oxford Handbook of Experimental Semantics and Pragmatics*. Oxford, New York: Oxford University Press.

Cummins, C., Sauerland, U., & Solt, S. (2012). Granularity and scalar implicature in numerical expressions. *Linguistics and Philosophy*, *35*(2), 135–169. doi: 10/ghspkx

Dehaene, S. (1999). *The Number Sense: How the Mind Creates Mathematics*. New York: Oxford University Press.

Frank, M. C., & Goodman, N. D. (2012, May). Predicting Pragmatic Reasoning in Language Games. *Science*, *336*(6084), 998–998. doi: 10/gffdh3

Franke, M., & Bergen, L. (2020). Theory-driven statistical modeling for semantics and pragmatics: A case study on grammatically generated implicature readings. *Language*, *96*(2), e77-e96. doi: 10/ghwsgg

Franke, M., & Jäger, G. (2016). Probabilistic pragmatics, or why Bayes' rule is probably important for pragmatics. *Zeitschrift für Sprachwissenschaft*, *35*(1), 3–44. doi: 10/ghs2t2

Geurts, B. (2010). *Quantity Implicatures*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511975158

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, *20*(11), 818–829. doi: 10/f89z39

Hesse, C., & Benz, A. (2020). Scalar bounds and expected values of comparatively modified numerals. *Journal of Memory and Language*, *111*, 104068. doi: 10/ghbtnz

Jansen, C. J. M., & Pollmann, M. M. W. (2001, December). On Round Numbers: Pragmatic Aspects of Numerical Expressions. *Journal of Quantitative Linguistics*, *8*(3), 187–201. doi: 10/d84dgx

Krifka, M. (1999). At least some determiners aren't determiners. In K. Turner (Ed.), *The Semantics/Pragmatics Interface from Different Points of View (Current Research in the Semantics/Pragmatics Interface)* (Vol. 1, pp. 257–291). Elsevier.

Prince, A., & Smolensky, P. (2008). *Optimality Theory: Constraint Interaction in Generative Grammar*. Wiley-Blackwell.

Scontras, G., Tessler, M., & Franke, M. (2018). *Probabilistic language understanding.* http://www.problang.org/.

Solt, S., Cummins, C., & Palmović, M. (2017). The preference for approximation. *International Review of Pragmatics*, *9*(2), 248–268. doi: 10/ghspkn

Spector, B. (2020). Modified Numerals. In *The Wiley Blackwell Companion to Semantics* (pp. 1–28). American Cancer Society. doi: 10.1002/9781118788516.sem120

[8] We thank an anonymous reviewer for raising these points.