Causal Bayesian Networks, Signalling Games and Implicature of '*More than n*'

Anton Benz

Centre for General Linguistics, ZAS, Berlin

Abstract. We use causal Bayesian networks for modelling experimental data on scalar implicatures of numerals modified by the comparative quantifier '*more than*'. The data are taken from the study of Cummins, Sauerland and Solt [1]. They showed that subjects infer upper bounds for the true number n when '*more than n*' has been used. They also showed that the difference between upper and lower bounds differs with the roundness of n. We show how the data can be explained from a production model of the speaker. We are interested in the architecture of such a model, and argue that causal Bayesian networks provide a natural and promising approach.

1 Introduction

Causal Bayesian networks [2] provide a natural framework for high–level descriptions of the interaction of different (cognitive) modules involved in communication. In the simplest case, a Bayesian network is a sequence of random variables in which the value of each variable conditionally depends on the value of the preceding variable. For example, the speaker's intended meaning depends on his world knowledge, the signal which he produces depends on the intended meaning, the hearer's interpretation depends on the signal she receives, and subsequent actions she chooses may depend on the information gained from the signal. The state of the world, the intended message, the signal, its interpretation, and the subsequent action can all be seen as values of random variables which are causally dependent on each other in a linear order. The Bayesian representation of the communication process can be turned into a game by assuming that the random variables are independent players which have to solve the coordination problem of successful communication. Games of this type have been introduced under the name of *error models* in [3]. This model provides the background for the causal models which will be applied to experimental data on the comparative quantifier '*more than n*' [1].

A key prerequisite for drawing the quantity implicature from '*three*' to '*not more than three*' in e.g. '*Nigel has three children*' is the assumption that the speaker knows the exact number of Nigel's children. This led to the assumption that '*more than n*' constructions generate no implicatures as the comparative quantifier '*more than*' signals that the speaker lacks sufficient knowledge for making a more precise statement [4]. However, experimental results from Cummins, Sauerland & Solt [1] show that scalar implicatures are available from these constructions. For example, the size of the interval defined by the estimated lower and upper bounds for the number of people getting married is much higher in (1b) than in (1a):

(1) a) More than 90 people got married today.
    b) More than 100 people got married today.

Cummins [5] proposes an optimality theoretic model of the experimental findings. It is a one–sided production model with free constraint ranking. The input–output pairs are the speaker's intended meaning and its linguistic realisation. Implicatures are calculated by inferring the speaker's meaning from output. The model has no theory of information states. We show that this, together with the '*the winner takes it all*' rule of optimality theory, means that it cannot fully account for '*more than*' implicatures with round numerals.

As did Cummins, we derive implicatures from a production model. However, we represent it in the form of a causal Bayesian network: The first random variable represents the state of the world, in example (1) the true number of people getting married, the second variable represents the speaker's information state, in our example an interval in which the speaker expects the true number to lie, then the third variable describes the possible linguistic output, and the fourth variable the hearer's interpretation of the output. The key step is the definition of the possible speaker information states. We apply here the unmodified sequence rule of [6] such that an open interval (n, m) is a possible information state if m immediately follows n in a sequence of natural numbers defined by Jansen and Pollmann's rule.[1] Although this use of the sequence rule is not unproblematic from a cognitive perspective, we show that the model predict the experimental results of [1] about implicatures from '*more than* n' with round n.

In the experiments of Cummins, Sauerland & Solt, subjects were asked to estimate the lower and upper bounds of the interval in which the true number of people has to lie. As the use of a modified numeral '*more than* n' does not allow for an unambiguous reconstruction of the upper bound of the speaker's information state, the game theoretic model proposed in [3] implies that the upper bound cannot be part of the intended speaker meaning. This raises the question, how these implicatures about upper bounds are related to scalar implicature. Based on our production model, we argue that inferences about upper bounds are not communicated non–natural information in the sense of Grice [7, Ch. 14].

2 Scalar implicature

Normally, an utterance of (2) will communicate that *some but not all* of the students failed:

[1] It defines sequences of *round* numbers in different scales, e.g. '$5, 10, 15, \cdots$', '$10, 20, 30, \cdots$' or '$50, 100, 150, \cdots$.' The rule says that numbers appearing in approximation contexts like '*between* n *or* m ' are always adjacent elements of such sequences, e.g. $(20, 30)$, $(50, 100)$, or $(100, 150)$. See the Definition in (10) and examples in (11).

(2) Some of the students failed

Grice [7, Ch. 2] distinguished between *what is said* (*some, possibly all*), i.e. the semantic meaning of the utterance, and what is *implicated* (*not all*). Both aspects are part of the intended speaker meaning. The implicated meaning is inferred from the assumption that the speaker adheres to the so–called cooperative principle, and a number of *maxims*. In short, they ask the speaker to be cooperative, and to say as much (Quantity) as he knows (Quality) as long as it is relevant (Relation). In addition, his presentation should be short, orderly, perspicuous, etc. (Manner). For Grice, an implicature is a proposition which one can infer the speaker to believe in order to maintain the assumption of cooperativity and adherence to the maxims. Implicatures are, therefore, inferred after semantic meaning, hence pragmatic reasoning follows semantic reasoning, except in cases of disambiguation and saturation of anaphors and indexicals, which may involve pragmatic reasoning and precede semantic interpretation. That implicatures are not part of semantic meaning can be seen from their *cancellability*, i.e. they can be explicitly negated by the speaker as in '*Some, in fact, all of the students failed*'.

The implicature from '*some*' to '*not all*' is explained by adherence to the maxim of quantity: if the speaker had known that all failed, he would have said so (Quantity). So, he cannot know that all failed. If it is assumed that he knows whether all failed (Expert), then the implicature that not all failed follows.

This reasoning has been systematised by so–called *neo–Griceans*.[2] In the previous example, we find a *scale* $\langle$*all, some*$\rangle$ and a sentence frame $A(x) \equiv$ 'x *of the students failed*' such that $A(all)$ semantically implies $A(some)$ but not the other way round. As we have seen, an utterance of the weaker $A(some)$ implicates that $\neg A(all)$. This generalises as follows: given a *scale* $\langle e_1, \ldots, e_n \rangle$ of expressions of similar complexity which can be filled into a sentence frame $A(\,.\,)$ such that for $i < j$ $A(e_i) \rightarrow A(e_j)$ and $A(e_j) \not\rightarrow A(e_i)$, then an utterance of a weak alternative $A(e_j)$ *implicates* that all stronger alternatives $A(e_i)$, $i < j$, are false. In the following, we write $A(e_j) +\!> \neg A(e_i)$ if $A(e_j)$ implicates $\neg A(e_i)$.

According to the standard theory, numerals also define a scale $\langle \ldots,$ *five, four, three, two, one*$\rangle$. This predicts that a use of '*three*' will normally implicate that the same sentence with '*three*' replaced by '*four*' is false:

(3) A: How many children does John have?
B: John has three children.
+> John does not have four children.

This assumes that a numeral as '*three*' is semantically equivalent to '*three or more*'. Hence, the sentence '*John has three children*' is true semantically even if John has four, five or more children. This assumption is supported by examples as the following:

(4) a) A: Does John have three children?
B: Yes, he has even four!

[2] See [8, Ch. 3] for a summary.

b) All of the students solved three problems, some even five or six.

If '*three*' would mean '*exatly three*', then B could not answer with '*yes*' in (4a), and (4b) would be a contradiction.

One of the major issues of the standard theory is the question what exactly constitutes a scale? Not just any set of phrases $\langle e_1, \ldots, e_n \rangle$ ordered by logical strength can form a valid scale. It is often assumed that the e_is have to be of equal complexity. But even this condition is no guarantee for the e_is forming a scale. Modified numerals are a case in point:

(5) a) John has *more than two* children.
b) John has *more than three* children.
c) John has *more than four* children.

As '*more than* $n + 1$' implies '*more than* n', then, assuming that ⟨*more than four, more than three, more than two*⟩ forms a scale, an utterance of (5b) would implicate that John does not have more than four children, and, hence, that he has exactly four. Intuitively, the use of the modifier '*more than*' indicates that the speaker does not know the exact number. This was taken as evidence that comparatively modified numerals do not form scales and, hence, do not give rise to implicatures [4, 9].

3 Inferences from comparatively modified numerals

Although comparatively modified numerals do not give rise to the neo–Gricean scalar implicature, they allow for some conversational inferences. The following examples are taken from either [10], [1], or [5]. Consider:

(6) London has more than 1000 inhabitants.

As an answer to the question '*How many inhabitants has London?*', (6) is obviously underinformative. It conveys the impression that London is a small city, and a naïve addressee would clearly exclude the possibility that London has a million, let alone 10 million inhabitants. A similar effect is observable in the following example:

(7) a) More than 80 people got married today.
b) More than 90 people got married today.
c) More than 100 people got married today.

The sentences give rise to different expectations about the size of the interval that must contain the true number of people married that day. It should be larger for (7c) than for (7b). Whereas (7c) seems to be compatible with the real number being 120, (7b) suggests that less than 100 got married. Also (7a) seems to suggest a smaller interval than (7c). In (8), the expected interval becomes even smaller. One would expect that the true number is either 97, 98, or 99.

(8) More than 96 people got married today.

In addition to the inferences about the interval that must contain the true value, comparatively modified numerals also give rise to certain *ignorance* implicatures. For example, (8) conveys that the speaker does not know whether more than 97 people got married that day. It has been claimed that, in general, '*more than* n' gives rise to the clausal implicatures '*possibly* n' and '*possibly more than* $n+1$' [5].

Central for the following discussion are the experiments by Cummins, Sauerland and Solt [1], who provide support for the claim that the use of '*more than* n' invites additional inferences about the possible range of values. In these experiments, test subjects had to answer a questionnaire on the Amazon Mechanical Turk platform. The items consisted of a short statement containing a modified numeral followed by a question which asked subjects to estimate the true number. An example is shown in Figure 1.

> *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]

Fig. 1. Example item from Cummins, Sauerland and Solt [1, p. 146].

Cummins et al. considered the modifiers '*more than*' and '*at least*'. We are only concerned with the cases in which the numerical expression of the test item was '*more than* n'. Cummins et al. further distinguished two types of estimates: in the *single number* condition, subjects were required to provide a natural number; in the *range* condition, they had to provide a lower and an upper bound (*between* m *and* n) for the true number. In the following, we only consider the results for the range condition, which are shown in Figure 2.

The '*Low*' column shows the different lower bounds for '*more than* n', followed in brackets by the number of subjects providing the lower bound. For example, 40 subjects gave a lower bound of 100 for the '*more than* 100' condition, and 28 subjects the lower bound 101. The '*High*' column shows the different upper bounds.[3] The experimental data overall support the introspective data described before. However, they also show that lower and upper bounds tend to be *round* numbers. They are always round if n is round. Even for $n = 93$,

[3] The table provides no indication of how lower and upper bounds were paired. So, the estimated intervals cannot be reconstructed.

	Low	High
$n = 100$	100 (40)	150 (24)
	101 (28)	125 (12)
		120 (8)
		200 (7)
		1000 (5)
$n = 110$	110 (46)	120 (28)
	111 (31)	150 (24)
	100 (9)	200 (7)
		119 (6)
		130 (5)
		115 (5)

	Low	High
$n = 93$	93 (32)	100 (34)
	94 (29)	95 (14)
	90 (14)	99 (7)
		150 (6)
		125 (5)

Fig. 2. Experimental results from Cummins, Sauerland and Solt [1].

which is not round, the bounds are round numbers, or clearly chosen due to their proximity to a round number (99); the only exceptions are 93 and 94.

To sum up, there is evidence that comparatively modified numerals invite inferences which bear some similarity to standard scalar implicature. In particular, subjects feel entitled to make guesses about the intervals which contain the true values.

4 An Optimality theoretic model

In [1], the implicature from modified numerals were explained as a kind of weaker implicature without being too explicit about the exact status of these weaker implicature. In a theoretical article, Chris Cummins's proposed an optimality theoretic model of the experimental results [5].

Optimality Theory (OT) was introduced to phonology in the early 90ies [11]. It describes grammar as a symbolic system in which surface forms are considered solutions to a constraint optimisation problem. The constraints are violable and hierarchically ranked according to their strength such that the stronger constraints strictly dominate the weaker constraints. In a production model, the linguistic surface forms are the optimal output forms for an underlying input form. Cummins [5] proposes an OT production model in which the constraints are special instances of Gricean maxims. The output forms are the modified numerals '*more than n*', for example, '*more than 50*', '*more than 60*', '*more than 70*', etc. The input from which these surface forms are generated are the speaker's information states.

In the experiments of [1], subjects first read a text containing the modified numeral '*more than n*', and had then to estimate lower and upper bounds for the true number. For the experimental item shown in Figure 1, the text may have said, for example, that '*More than 100 people attended the public meeting about the new highway construction project.*' Cummins's OT model is a model of the author producing the text. It tells us for which speaker information states the text is an optimal output. The assumption is that, as the OT model is common knowledge, the subjects can use this knowledge for inferring the underlying information state, and thereby for inferring the lower and upper bounds.

The optimal output for an information state is calculated with the help of a tableau as that shown in (9). The left upper corner of (9) shows the information state of knowing that between 101 and 109 people came to the meeting. In the column under the information state all possible outputs are listed. Cummins assumes that only modified numerals enter the competition. The constraints are listed in the top row.[4] First, there is the QUALITY constraint, which asks the speaker to say only what he believes to be true. This constraint is assumed never to be violated. It is therefore always to be ranked highest. Next come the QUANTITY constraint, which asks the speaker to choose the most informative output, and the SALIENCE constraint, which asks the speaker to choose round numbers. It is assumed that different speakers can come with different rankings of QUANTITY and SALIENCE. Hence, from the hearer perspective, these constraints can be arbitrarily ordered. In (9), this is indicated by the dashed vertical line between the last two columns.

(9)

101–109	QUALITY	QUANTITY	SALIENCE
more than 110	*!		**
☞ more than 100			
more than 90		*!	**
more than 80		*!*	*

As the speaker believes that the true number of people coming to the meeting does not exceed 109, it would violate the QUALITY constraint to say that '*More than 110 came*'. This is indicated by the * in the column below QUALITY. All other numerals satisfy the constraint. The numerals 90 and 80 are less informative than 100. This is indicated by the * in the column below QUANTITY. The violation is stronger for 80 than for 90, hence, 80 receives two stars and 90 only one star. The SALIENCE constraint measures the *roundness* of the chosen numeral. In the given example, 100 is rounder than 80, and 80 is rounder than 90 and 110. The violation of SALIENCE, shown in the last column, grows stronger with decreasing roundness.

An output F wins over another output G, if the first constraint for which they differ is violated more strongly by G than by F. '*More than 100*' is the only output which satisfies all constraints. It therefore wins the competition. The '!' in other rows marks the first violation which makes the corresponding output loose the competition with the winning form.

[4] The full model also contains a constraint preferring the re-use of previously mentioned (*primed*) numerals. For the special problem of implicatures from '*more than n*' for *round* numerals n, this constraint can be left out of consideration. It becomes relevant for non–round numerals, as in '*more than 93*'. We also changed constraint naming: Cummins's tableaux do not include (Quality), and we call the quantity maxim by its conventional name of QUANTITY instead of INFORMATIVITY, which Cummins chose.

Cummins's measure of roundedness is based on the corpus studies of Jansen & Pollmann [6]. As an indicator of the degree of roundness Jansen & Pollmann considered the frequencies of numerals in approximation contexts created by, for example, Dutch '*ongeveer*', German '*etwa*', French '*environ*', and English '*about*'. The more frequent a numeral is in such contexts, the *rounder* it is assumed to be.

Cummins's OT model makes predictions about which numerals n are optimal in '*more than n*' contexts. He provides an explanation for why round numerals allow for a larger range of upper bounds. He also claims that his model explains the ignorance implicature from '*more than n*' to *the speaker believes it possible that* $n + 1$ ($\Diamond n + 1$).

The wider range of upper bounds for round numerals is explained as a consequence of the free ranking of the QUANTITY and SALIENCE constraints. For example, if the speaker knows that the true number of people attending some event lies in the open interval $(150, 200)$, then, assuming that 100 is rounder than 150, it follows that a speaker ranking SALIENCE over QUANTITY will prefer the less precise '*more than 100*' over '*more than 150*'. If, in contrast, the speaker knows that the true number lies in some interval above 100, for example in the interval $(110, 120)$, then there is no constraint ranking which could make the choice of '*more than 90*' optimal. This restricts the allowed range for the true number to $(90, 100)$ for '*more than 90*', and allows a wide range for '*more than 100*'. However, if the belief interval is $(150, 200)$ and SALIENCE is ranked higher than QUANTITY, then the model not only predicts that '*more than 100*' wins over '*more than 150*', but also that '*more than 10*' wins over '*more than 100*', as 10 is rounder than 100.[5] Hence, the speaker should choose '*more than 10*'. The model also predicts, as there is no higher number rounder than 100, that '*more than 100*' wins over all alternatives '*more than n*' for numbers n greater than 100. This is a consequence of the '*the winner takes it all*' property of OT models. It entails that the most salient number N, whichever it is, has no upper bound, and all other numbers greater than N can only be used if QUANTITY is ranked over SALIENCE. The model entails, therefore, that the ignorance implicature only exist for numbers which are preceded by a more salient number. For the maximally salient numbers, the implicature from '*more than N*' to $\Diamond N + 1$ does not hold true. If someone who believes the true number to lie in $(150, 200)$ can utter '*more than 100*', then the addressee cannot infer from this utterance that the speaker believes 101 to be possible.[6]

As the model imposes no restrictions on the possible speaker information states, no inferences about upper bounds can be drawn if QUANTITY is ranked over SALIENCE. For example, if all intervals $(110, L_u)$ are admissible information states of the speaker, then from an utterance of '*more than 110*' no restriction

[5] In fact, this depends on which number is considered to be roundest. For example, the frequency graph in [6, p. 193] shows an even higher peak for 20 than for 10. However, pure frequencies may not be the best criteria of roundness.

[6] This reveals a (minor) inconsistency in Cummins's argumentation [5], as he claims that the ignorance implicature always holds true.

on the upper limit L_u can be inferred. It could be arbitrarily large. Cummins obviously assumes some restrictions on possible information states, but they are not made explicit in his model.[7]

We sympathise with Cummins's approach of explaining the observed implicatures from a production model. However, the model has shortcomings which warrant considering alternatives. In particular, it seems necessary to have a model that explicitly describes the set of admissible information states. Without it no inferences can be made about upper bounds. We also will drop the optimality theoretic framework.[8] We prefer a representation which describes the communication process and the experimental situation more directly. As we are mainly interested in the general architecture of such a representation, we consider only a limited subset of the experimental results. For example, we leave out of consideration non–round numerals, as, for example, 93.[9] The main question in the next section is how to account for the influence of roundedness on the estimated difference between lower and upper bounds even if it is assumed that the speaker tries to be as informative as possible, i.e. even if, in OT terms, the speaker ranks QUANTITY over SALIENCE.

5 A causal Bayesian network model

Our aim is to explain the experimental data shown in Figure 2 on '*more than 100*'. The main effort will be to set up a model of the author producing the text appearing in the questionnaire shown in Figure 1, for example, the text '*More than 100 people attended the public meeting about the new highway construction project.*'

In Figure 3, a model of the experimental situation which starts with a reported (imaginary) event and concludes with the test subject's answer is shown. Each node in this sequence corresponds to a random variable in a causal Bayesian network. The values of the first variable named WORLD are natural numbers N, those of the third variable named UTTERANCE are sentences containing '*more than n*' for some natural number n, and the values of the last variable named ESTIMATE are pairs $\langle lb, ub \rangle$ of lower and upper bounds for the true number N. In order to complete this model, it has to be known which values the variable named KNOWLEDGE can take, and it has to be known what the conditional probabilities of the various values are.

We will identify the knowledge states with open intervals (n, m) using Jansen & Pollmann's unmodified *sequence rule* [6]. It is obviously not possible to deter-

[7] Excluding examples with anaphoric uses, Cummins always considers information states (n, m) for which n and m are *adjacent* round numbers. Making this aspect explicit would presumably reveal the same application of Jansen & Pollmann's sequence rule as in in our own model.

[8] It may also be debated that an OT model with free constraint ranking is still in the spirit of optimality theory.

[9] For these numerals, Cummins has also to apply a different model with an additional constraint.

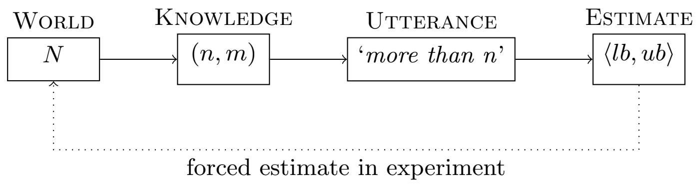

Fig. 3. A model of the Cummins, Sauerland, and Solt experiment [1].

mine the conditional probabilities. Fortunately, this is not necessary as we will see once we know how to define implicature. We address this issue next.

5.1 Implicature in Bayesian networks

In a Bayesian network, the different random variables carry information about each other. For our purposes, we can assume that communication is describable by a linear sequence of random variables $(\mathscr{X}_i, P_i)_{i=1}^n$ with conditional probabilities $P_i(x_i|x_{i-1})$ defined for $x_i \in \mathscr{X}_i$ and $x_{i-1} \in \mathscr{X}_{i-1}$. With the product probability $P(x_1, \ldots, x_n) = P_1(x_1) \times \ldots \times P_n(x_n|x_{n-1})$, we can define the probability that the i^{th} variable has a value in some set X given that the j^{th} variable has a value in Y as follows:[10]

$$P_{i|j}(X|Y) := \frac{P(\pi_i^{-1}(X) \cap \pi_j^{-1}(Y))}{P(\pi_j^{-1}(Y))} \text{ with } \pi_k(x_1, \ldots, x_n) = x_k. \tag{1.1}$$

In a linear Bayesian network describing communication, one variable will describe the utterances the speaker produces. Let us assume it is the j^{th} variable, and that the utterance is F. Then we can say that the utterance of F *implicates* that some other variable $\mathscr{X}_i$ has a value in X iff $P_{i|j}(X|F) = 1$.

It follows that, in order to determine implicature, it is only necessary to know which values x_i of a variable $\mathscr{X}_i$ are *consistent* with an utterance of F. Hence, it suffices to know for which x_i $P_{i|j}(x_i|F) > 0$.

This definition is justified by Grice's contention [7, p. 86] that '*what is implicated is what it [sic!] is required that one assume a speaker to think in order to preserve the assumption that he is observing the Cooperative Principle (and perhaps some conversational maxims as well), ...*' The most obvious interpretation is that what is implicated by an utterance of F is the information that is carried by F about values of variables representing the speaker's beliefs.[11]

---

[10] Whenever a probability appears in the denominator, we implicitly assume that it is greater than zero. If X or Y is a singleton set, for example $Y = \{y\}$, the brackets around y are omitted in argument positions of probability distributions.

[11] This definition deviates from common usage in so far as it includes literal meaning in implicated meaning, whereas what is implicated is generally understood to be the meaning that is communicated but not literally expressed. For simplicity, we use this inclusive definition. It can easily be modified so that it excludes literal meaning.

5.2 Knowledge states and the sequence rule

All estimated lower and upper bounds shown in Figure 2 are round numbers, or differ only by a value of one from a round number. As we have seen, the speaker's preference for round numbers is also a key element of Cummins's OT model. In the Bayesian network in Figure 3, the (imaginary) speaker's first task is to estimate the true number of people taking part in some event. We assume with Cummins's that this estimate is some interval. More specifically, we assume that it is an open interval (n, m) with at least two elements defined by two round numbers n and m. To further restrict the admissible pairs of numbers n, m, we make use of the corpus study by Jansen & Pollmann [6] who found that pairs of numbers n, m in the approximative construction '*about n or m*' follow the following rule:

(10) *Sequence rule:*[12] For context '*about n or m*', the numbers n, m follow each other directly in a sequence:

$$L_{k,l} = \{i\,(k \times 10^l) \in \mathbb{N} \mid i \in \mathbb{N}\} \text{ with } k = 1, 2, {}^1/_2, {}^1/_4, \text{ and } l \geq 0. \quad (1.2)$$

The following list in (11) shows all sequences which are relevant to us.

(11)

$L_{1,0} = 1,\ 2,\ 3,\ 4,\ 5,\ \ldots$	$L_{2,1} = 20,\ 40,\ 60,\ 80,\ \ldots$
$L_{2,0} = 2,\ 4,\ 6,\ 8,\ 10,\ \ldots$	$L_{1/4,2} = 25,\ 50,\ 75,\ \ldots$
$L_{1/2,1} = 5,\ 10,\ 15,\ 20,\ \ldots$	$L_{1/2,2} = 50,\ 100,\ 150,\ 200,\ \ldots$
$L_{1,1} = 10,\ 20,\ 30,\ 40,\ 50,\ \ldots$	$L_{1,2} = 100,\ 200,\ 300,\ 400,\ 500,\ \ldots$

Let us write $(n, m) \sqsubset L_{k,l}$ if n and m are consecutive numbers in sequence $L_{k,l}$. We assume that (n, m) is a possible information state if n and m are consecutive numbers in some sequence $L_{k,l}$. In addition, we assume that the speaker's estimate is correct, i.e. that the true number N is an element of (n, m). Translated into conditional probabilities $P((n, m)|N)$, this means:

$$P((n, m)|N) > 0 \textit{ iff } \exists k \exists l > 0 : (n, m) \sqsubset L_{k,l} \wedge N \in (n, m). \quad (1.3)$$

As mentioned before, it suffices for implicature calculation to know when probabilities are greater zero. In order to rule out intervals with one or zero elements, we assume that $l > 0$.

However, Grice's original characterisation of conversational implicature also does not exclude literal meaning [7, pp. 30f].

[12] This is the *unmodified* sequence rule. Jannsen & Pollmann [6] see no justification for the inclusion of $k = {}^1/_4$. However, we can ignore this modification for our purposes. For $k = {}^1/_4$, l must be greater than 1 to make sure that all elements of $L_{1/4,l}$ are natural.

5.3 Speaker's strategy and experimental data

The last parameter to be fixed for the model in Figure 3 is the probability with which the speaker chooses '*more than n*' given that his knowledge state is (n', m'). In game theoretic terminology, this conditional probability can be called the speaker's *strategy*. We make the assumption that the speaker can only produce F_n = '*more than n*' if $n' = n$:

$$S(F_n|(n', m')) > 0 \text{ iff } n = n'. \tag{1.4}$$

The situations represented by condition in (1.4) correspond to those situations in which QUANTITY is ranked higher than SALIENCE in Cummins's optimality theoretic model. As we have seen, for these cases, the optimality theoretic model cannot make predictions about upper bounds. In Section 5.4, we will shortly discuss cases in which the speaker could have been more informative, for example, because he knew that exactly 93 people came to an event, or because he knew that at least as many came, but chooses a less informative lower bound, for example '*more than 90*'.

This completes the set-up of our model. The probabilities with which the test subjects choose their estimates and the nature of the estimates are provided by the experimental data.

The table in (12) lists the belief intervals which have probability greater zero given that the speaker uttered F_n for the numbers $n = 80, 90, 100$. To each interval, the sequence to which the interval belongs is added. More precisely, if $\mathcal{F}$ is the name of the random variable representing utterances, and Θ the random variable for the speaker's belief, then the intervals (n, m) listed in table (12) are those for which $P_{\Theta|\mathcal{F}}((n', m')|F_n) > 0$.

(12)

F_n	belief intervals			
'*more than 80*'	$L_{1/2,1}$	$(80, 85)$	$L_{1,1}$	$(80, 90)$
	$L_{2,1}$	$(80, 100)$		
'*more than 90*'	$L_{1/2,1}$	$(90, 95)$	$L_{1,1}$	$(90, 100)$
'*more than 100*'	$L_{1/2,1}$	$(100, 105)$	$L_{1,1}$	$(100, 110)$
	$L_{2,1}$	$(100, 120)$	$L_{1/4,2}$	$(100, 125)$
	$L_{1/2,2}$	$(100, 150)$	$L_{1,2}$	$(100, 200)$

Given our definition of implicature, this table says that an utterance of '*more than 80*' implicates that the speaker believes that the true number is an element of $(80, 85)$, $(80, 90)$, or $(80, 100)$. Likewise, an utterance of '*more than 100*' implicates that the speaker estimates the true number to lie in one of the intervals with lower bound 100 and upper bounds 105, 110, 120, 125, 150, and 200. Accordingly, following from the assumption that the speaker's beliefs are true, an utterance of '*more than 100*' implicates that the true number lies in the interval $(100, 200)$, and an utterance of '*more than 80*' that it lies in the interval $(80, 100)$. The model accounts for the different ranges of estimated intervals without having to rank the preference for round numbers over the preference for information.

Equation (1.4) entails that the speaker chooses the most informative version of '*more than n*'. The model also accounts for the ignorance implicature '*more than n*' implicating *possibly n*.

It is difficult to say how test subjects construed the task in the questionnaire. They were not explicitly asked for *implicatures*, only for estimated lower and upper limits of the true number N. It is plausible to assume that they provided probable and safe limits within which the speaker's belief intervals lie. The experimental results for '*more than 100*' are repeated in (13). We find a good match between the predicted set of possible upper bounds and elicited upper bounds. The only predicted upper bounds which cannot be found in the experimental data are 105 and 110, and the only upper bound not predicted is 1000.

(13)

	Low	High
$n = 100$	100 (40)	150 (24)
	101 (28)	125 (12)
		120 (8)
		200 (7)
		1000 (5)

A plausible reason for the missing 105 and 110 is that the resulting intervals would be too small. As test subjects had no information about speakers, the belief intervals $(100, 105)$ and $(100, 110)$ would have been too unlikely to be the only speaker types. On the other side, the estimated upper bound of 1000 may be the result of a special safety strategy of test subjects.[13]

5.4 Extensions of the basic model

In the basic model we only considered the case in which the assumed speaker of the test sentence believes that the true number N can lie anywhere in an open interval (n, m). When he produces an utterance with '*more than*', then he always chooses n as the lower bound. This was encoded in Equation (1.4). In this section, we consider other strategies which may lead to utterances containing '*more than n*'. The reasoning is admittedly even more speculative than in the previous sections. The aim is to show that the constraints over possible estimates for the interval containing the true number N that may be produced by the *test subjects* remain valid also when considering other speaker strategies which might have produced the test sentences. A second aim is to show how the sequence rule *might* be used to capture a variety of other reasons leading to utterances with numerals modified by '*more than*'.

We first consider the following possibility: If the speaker says that '*More than 90 people attended the public meeting about the new highway construction*

[13] In the cognitive literature it is also argued that the distances on the mental number line are rescaled according to some logarithmic rule [12]. This may have an effect on choosing upper bounds which is not accounted for in our model but may explain why some people estimate the true number to lie between '*100 and 1000.*'

project', it is conceivable that he knows that exactly 93 people attended but chose the round number 90 as lower bound for '*more than n*'. In Cummins's model one has to assume here that SALIENCE outranks QUANTITY. If the speaker knows that exactly N people attended, the dominance of SALIENCE over QUANTITY means that the speaker must choose the roundest number n smaller or equal to N. However, as we have seen, the restriction to the *roundest* number smaller than N seems to be too strong a requirement. We will now show how a prediction about possible lower bounds n can be derived from the sequence rule. It can be assumed that the different sequences $L_{k,l}$ resulting from different choices of numbers $k = 1, 2, 1/2, 1/4$ and $l \in \mathbb{N}$ represent different levels of precision with which the speaker wants to describe the true situation. In the following model, it is assumed that the speaker, who knows that exactly N people attended, first chooses a level of precision $L_{k,l}$ with which he wants to describe the situation, and then the largest number n in $L_{k,l}$ which is smaller than N as a lower bound for '*more than n*'. The following constraint, which modifies the conditions stated in equations (1.3) and (1.4), captures this idea. It says that F_n = '*more than n*' is a possible utterance if there is an interval (n, m) of adjacent numbers in some sequence L_{kl} which contains N:

$$S(F_n|N) > 0 \text{ \textit{iff} } \exists k \exists l > 0 \, \exists m > n : (n, m) \sqsubset L_{k,l} \wedge N \in (n, m). \qquad (1.5)$$

For $N = 93$, this rule predicts that the speaker can produce the following sentences:

(14) a) More than 90 people attended. ($(n, m) = (90, 95), (90, 100)$)
b) More than 80 people attended. ($(n, m) = (80, 100)$)
c) More than 75 people attended. ($(n, m) = (75, 100)$)
d) More than 50 people attended. ($(n, m) = (50, 100)$)

These are all possibilities. Numbers like $n = 20$ or $n = 10$ are ruled out because there is no sequence $L_{k,l}$ for which there exist intervals $(20, m)$ or $10, m)$ that would contain 93 as an element. What is important with respect to the experimental data of Cummins, Sauerland & Solt [1] is that the strategy defined in (1.5) does not introduce expected intervals (n, m) different from the intervals introduced by the basic rule in (1.4): If the addressee hears that e.g. '*more than 80 people attended*', then she can infer that the true number is an element of $(80, 100)$. So, even when the addressee does not know whether the speaker's information state is an interval or an exact number, she should arrive at the same estimates.

Let us consider one further scenario. It is conceivable that the speaker counted the people attending the meeting and stopped counting at, say, 91. He may be certain that at least two more attended, and, maybe, up to a hundred. When asked, how many were there, he may well answer '*more than 90*'. Also for such a scenario we can derive a plausible mode from Jansen & Pollmann's sequence rule. In our example, the speaker knows that at least 93 attended. So let N be any number, not necessarily round, for which the speaker verified by counting that at least N attended. Then the following strategy produces plausible '*more than n*' utterances with round lower bounds n:

$$S(F_n|(N,m)) > 0 \text{ iff } \exists k \exists l > 0 : (n,m) \sqsubset L_{k,l} \wedge N \in (n,m). \quad (1.6)$$

The predicted intervals for $N = 93$ are the same as in (14). This also means that the hearer, when confronted with a '*more than n*' message, has no reason to change her estimates for the interval containing the true number.

As mentioned before, the modifications presented in this section are increasingly speculative. However, they demonstrate how the model can be extended to cover more scenarios in which the speaker can produce a comparatively modified numeral. In terms of linguistic depth, the achievements may seem modest. Even with the extensions presented, the model does not account for vague interpretations of unmodified numerals, as, for example, in '*A hundred people came to the talk*', which can mean that *about 100* people came. It also does not account for an utterance of '*more than 93*'. For these examples, the model has to take into account additional parameters, e.g. previously mentioned numerals, and contextually required precision. This leads to nontrivial extensions, which we leave to future research.

Before we end this section, we provide an overview of how roundness can influence the experimental results in the setting of Cummins, Sauerland & Solt [1]. In our basic model, roundness entered the definition of conditional probabilities of belief intervals. However, roundness can also influence the choice of utterances and the interval estimation by the test subject. These possibilities are shown in Figure 4.

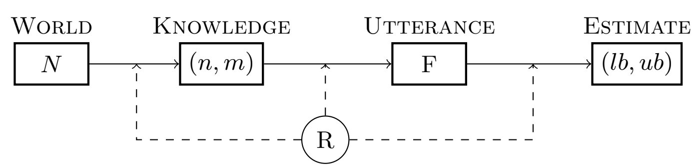

Fig. 4. Influence of roundness R on probabilities.

If roundness determines the formation of belief intervals, then the (imaginary) speaker of the '*more than n*' sentence is restricted to belief states (n, m) with round neighbours in a Jansen & Pollmann sequence, as assumed in our basic model. If roundness enters the choice of utterance, then the speaker may produce '*more than 100*' although he knows that there are more than 110. This is the case we have considered in this section. There is also the possibility that roundness enters the choice of the test subject's estimate of the interval containing the true number. This case we have not considered so far. For example, the test subject may estimate that the true number is between 103 and 150, but when asked for lower and upper bounds, she may respond between 100 and 150, simply because $(100, 150)$ is a particularly salient interval. However, to disentangle how roundness exactly influences which module in Figure 4 needs further experimental investigations.

6 Implicature in which sense?

In this section we consider the question what type of implicature are inferences from '*more than n*' to upper bounds for the true n. Krifka [4] maintained that '*more than n*' does not generate implicatures. Cummins et al. [1] considered '*more than n*' implicature to be *weak* implicatures compared to standard scalar implicatures. Cummins [5] explained them as inferences about the speaker's beliefs which are inferred from an optimality theoretic production model. In this model, there seems to be no principled difference between implicature from '*more than n*' and those from unmodified numerals, for example, from '*John has three children*' to John has exactly three children.

We quoted before a passage from Grice which says that an implicature is what is required that one assume a speaker to think in order to preserve the assumption that he is following the conversational maxims. This suggests that an implicature is an inference about some belief of the speaker. However, in [7, Ch. 14], Grice defined communicated meaning as meaning that the speaker *intended* to communicate. This makes an implicature an inference about intentions, and not just about some belief he happens to entertain. Accordingly, we can distinguish between two types of implicature:

(15) a) (Grice–B): A belief of the speaker. (*Logic of Conversation* [7, Ch. 2])
b) (Grice–I): The part of the intended meaning which is not literally expressed. (*Meaning* [7, Ch. 14])

These implicature can be distinguished in a causal model if distinct nodes for beliefs and intentions are introduced, as shown in Figure 5.

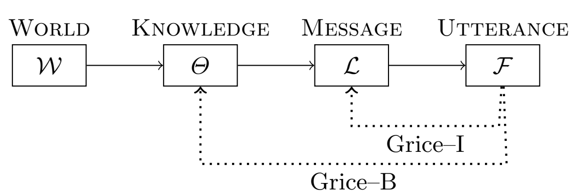

Fig. 5. Grice–B and Grice–I implicature.

The standard examples of scalar implicature in (16) are clearly part of what has been communicated. Hence, they constitute Grice–I implicature.

(16) a) Some of the students failed the exam. +> not all failed.
b) John has three children. +> He has not more than three.

However, the implicatures about upper bounds in (7), repeated here as (17), are normally not part of what the speaker intended to communicate. In this sense they are *weaker* implicature as Cummins et al. [1] maintained.

(17) a) More than 80 people got married today. +> between 80 and 100
b) More than 90 people got married today. +> between 90 and 100
c) More than 100 people got married today. +> between 100 and 200

This becomes also apparent from the causal model we considered in the previous section. Implicature from '*more than n*' were there considered inferences about the speaker's knowledge.

This also reconciles Krifka's [4] view that '*more than n*' does not generate implicature with the experimental results of [1], if one assumes that Krifka's claim relates to Grice–I implicature, and not to Grice–B implicature.

7 Conclusion

The main aim of this paper was to show the usefulness of causal networks for analysing the experimental setting in experiments of the type performed by Cummins, Sauerland & Solt [1]. Causal networks provide a high level description of communication by showing the causal dependencies between different modules in language production and interpretation. They are a very natural framework to use. We have seen how to explicate the notion of *implicature* in these networks as inferences from one special variable representing utterances to properties of other variables. This provided us with a general definition, and we could show that, in spite of its abstractness, it helps to make fine distinctions within the group of quantity implicature.

We have seen that the model provides a good match for upper bounds of '*more than 100*'. It explains why differences between lower and upper bounds are larger for round numbers. It also shows that this pattern is consistent with the assumption that the speaker tries to be as informative as possible. Hence, varying differences are consistent with the ignorance implicature from '*more than* n' to '*it is possible that* $n+1$'. The model also predicts that '*more than* n' can only be used with round numbers.

We have argued that pragmatic inferences from, for example, '*more than* 100' to '*less than 200*' are not Grice-I implicatures, hence, that they are not part of the speaker's intended message. This makes them different from implicatures triggered by unmodified numerals, as, for example, by '*John has three children*' implicating that '*John has exactly three children*'. In this sense, we agree with [4, 9] that comparatively modified numerals do not give rise to standard scalar implicatures. The experimental results of [1] seem to be the consequence of general inferences about the speaker's probable knowledge state.

References

1. Cummins, C., Sauerland, U., Solt, S.: Granularity and scalar implicature in numerical expressions. Linguistics and Philosophy **35** (2012) 135–169

2. Pearl, J.: Causality: Models, Reasoning, and Inference. Cambridge University Press, Cambridge (2000)
3. Benz, A.: Errors in pragmatics. Journal of Logic, Language, and Information **21** (2012) 97–116
4. Krifka, M.: At least some determiners aren't determiners. In Turner, K., ed.: The semantics/pragmatics interface from different points of view. Volume 1 of Current research in the semantics/pragmatics interface. Elsevier, Oxford (1999) 257–292
5. Cummins, C.: Modelling implicatures from modified numerals. Lingua **132** (2013) 103–114
6. Jansen, C.J., Pollmann, M.M.: On round numbers: Pragmatic aspects of numerical expressions. Journal of Quantitative Linguistics **8(3)** (2001) 187–201
7. Grice, H.P.: Studies in the Way of Words. Harvard University Press, Cambridge MA (1989)
8. Levinson, S.C.: Pragmatics. Cambridge University Press, Cambridge (1983)
9. Fox, D., Hackl, M.: The universal density of measurement. Linguistics and Philosophy **29** (2006) 537–586
10. Cummins, C.: The interpretation and use of numerically-quantified expressions. PhD thesis, University of Cambridge (2011)
11. Prince, A., Smolensky, P.: Optimality theory: Constraint interaction in generative grammar. Technical Report 2, Rutgers University, Centre for Cognitive Science (1993)
12. Dehaene, S.: The Number Sense: How the Mind Creates Mathematics, Revised and Updated Edition. Oxford University Press, Oxford (2011)