

Relevance Theory, Grice and the neo-Griceans: a response to Laurence Horn's 'Current issues in neo-Gricean pragmatics'*

In this short forum piece, I respond to three of the main points Laurence Horn made in his replies to questions from Istvan Kecskes in this journal (Horn 2005a). My response will naturally (and as requested) reflect my relevance-theoretic allegiances, but I am looking as much for points of agreement and/or complementarity between Relevance Theory and Horn's neo-Gricean framework as for divergences.

1. Grice, neo-Griceans and Relevance theorists: different projects with different goals?

The first section of Horn's paper is a response to Kecskes' request to elaborate on the view that there is a fundamental difference in perspective and goals between the neo-Gricean and the relevance-theoretic approaches to pragmatics. The crux of Horn's response is as follows: 'Grice's goal of developing an account of speaker meaning (of which implicature constitutes a proper subpart) is distinct from Relevance theorists' goal of developing a cognitive psychological model of utterance interpretation, which does not address the question of how and why the speaker, given what she wants to convey, utters what she utters.' Horn (2005a: 194). While it is true that the Gricean project and that of Relevance Theory (RT) have largely distinct goals,¹ the quoted statement is too stark as it stands. RT is very much concerned with speaker meaning, both what it is and how an addressee attempts to recover it. As regards the complex question of why the speaker, given her communicative intention, utters what she utters (that is, how and why she chooses a particular linguistic form), I believe that RT does make some predictions about this (as does Horn's system) and will discuss the issue in the third section of the paper. A point to note about the Kecskes/Horn exchange is that it seems to carry a presupposition that Grice's project and the neo-Gricean project are at one in their goals, as distinct from RT. This is highly debatable. First, there are considerable differences within the broad camp of neo-Griceans; for instance, Horn (2005a, 2005b) explicitly distances himself from the other best-known neo-Gricean, Stephen Levinson (in particular, Levinson (2000)), pointing out that Levinson's perspective is more akin in important respects to that of RT. More pertinent here, though, Horn's own project diverges in major ways from Grice's, as I shall try to show.

Before looking a little more closely at these issues, I will take the opportunity to endorse and develop Horn's point that a primary goal of RT pragmatics is to provide an account of utterance comprehension. The RT framework starts with a general view of human cognitive processing, of what motivates us to attend to certain information sources but not others, and to process certain inputs while filtering out others. Omitting details, the claim is that human cognition is oriented to selecting and processing *relevant* inputs, that is, inputs that are potentially beneficial to the ongoing functioning of the cognitive system and whose processing costs to the system are relatively low (see Sperber and Wilson 1995, 2002). One very rich source of information is linguistic utterances, which are usually attention pre-empting. According to the RT account, this is because they, unlike most other information sources, carry a presumption that they will achieve a certain level of relevance, that is, of satisfactory cognitive benefit for no gratuitous processing effort. This is simply intrinsic to the kind of stimulus that they are: 'ostensive', that is, produced by an intentional agent who in so doing makes fully overt (mutually manifest, in RT terms) an intention to provide relevant information. Correlatively, the cognitive system of an addressee is attuned to looking for a certain degree and quality of relevance from ostensive stimuli and processes them on this basis.

Clearly, this construal of pragmatics as a cognitive capacity has a very wide reach: it touches on questions about cognitive architecture (what sort of a system is responsible for pragmatic processing? is it task-dedicated or a more general problem-solving system? what other cognitive systems does it interact with?); it is receptive to considerations from evolutionary psychology concerning the kinds of cognitive systems that have been naturally selected to solve particular adaptive problems (is pragmatics one of these?); it must answer to experimental findings about the nature and time-course of utterance comprehension (for example, which elements of conceptual information are activated or deactivated at which points in the processing of ambiguous words, or metaphorical uses, or cases of implicature?).

According to the current state of play in RT, the pragmatic processes of utterance comprehension are performed by an evolved task-specific (modular) system, which, although closely related to a more general 'mind-reading' system whose domain is all intentional human behaviour, has its own idiosyncratic principles and procedures dedicated to interpreting communicative behaviour (Sperber & Wilson 2002; Wilson & Sperber 2004). Central here is the Communicative Principle of Relevance, according to which utterances (ostensive stimuli generally) convey a presumption of their own *optimal relevance*. What this

amounts to is that any given utterance can be presumed to be: (a) at least relevant enough to warrant the addressee's processing effort, and (b) the most relevant one compatible with the speaker's current state of knowledge and her personal preferences and goals. The general procedure (which can take several more specific forms) that the system follows in comprehending an utterance (that is, in recovering the speaker's meaning) is to test possible interpretations in their order of accessibility, stopping once the expectation of (optimal) relevance is satisfied. This relevance-based comprehension strategy is often ignored by critics of RT, but it has been a central component of the theory since at least 1995 (its first published appearance was in Sperber et al. (1995: 51)).

Grice wouldn't have been much interested in this cognitively-oriented approach and nor is Horn. Neither has been concerned to provide an account of the processes of on-line utterance comprehension. However, since both offer principles/maxims that, in some sense at least, are taken to bear on the speaker's choice of utterance to communicate her meaning and on how it is understood by the hearer, it seems reasonable that initial attempts to account for utterance comprehension should turn to such principles and see whether they might serve adequately in such an account. This has been done with some of Grice's maxims. For instance, early psycholinguistic work on the processes involved in comprehending non-literal utterances (metaphor and irony) developed and tested a model based on the idea that these involve a flouting of the first maxim of Quality (truthfulness). According to such a processing employment of Grice's framework, a literal interpretation is always retrieved first, tested for conformity with the maxims, then, if found unsatisfactory (blatantly false or trivially true), a non-literal interpretation is accessed. A vast swathe of experimental work indicates that this sort of three-stage, 'literal priority' model is incorrect (see, for instance, Gibbs 1994 and Glucksberg 2001). More recently, hypotheses based on descendants of Grice's first maxim of Quantity (informativeness) have been tested in experiments on the on-line derivation of scalar implicatures (see, for instance, Noveck 2004). And there have been attempts to employ Gricean maxims in formulating computational accounts of aspects of interpretation and of implicature generation (for instance, Dale and Reiter 1996, Green and Carberry 1999). These efforts at building psychologically plausible or computationally explicit accounts of aspects of utterance understanding are all very recent endeavours. They had to start somewhere and, for the psychologist in particular, who better to start with than Grice, whose emphasis on speakers' addressee-directed intentions, on hearers' expectations about speakers' conversational behaviour and on the role of non-demonstrative inference in

implicature derivation seemed to hold out a reasonable possibility of bearing on the psychology of comprehension.

Horn says the goal of Gricean theory is to explicate speaker meaning, or, more specifically, ‘to characterize implicature as part of speaker meaning’ (Horn 2005a: 192). This is a rather narrow conception of Grice’s project. As Neale (1992: 509) lucidly outlines, the aims of Grice’s philosophy of language included at least the following: (1) to distinguish what an expression *e* means, what a speaker *S* *said* on a given occasion by uttering *e*, and what *S* *meant* by uttering *e* on that occasion, (2) to elucidate the nature of non-natural meaning by providing analyses of speaker’s meaning, sentence meaning, and what is said, (3) to provide an account of how it is possible for what *S* says and what *S* means to diverge, (4) to distinguish between semantic content and pragmatic implications and so clarify the relationship between classical logic, natural language semantics and ordinary language use. Of particular importance in this project was the role of ‘what is said’ or ‘the statement made’, since this is where linguistic expression meaning and speaker meaning interface, and it is the truth or falsity of this central aspect of a speaker’s meaning on which the validity of particular philosophical theses rests (Grice 1967/1989b; lecture 1).²

Horn’s perspective on pragmatics seems rather different. His focus is almost entirely on the idea that there are two kinds of fairly regularly occurring conversational implicature, one kind resulting from his R-Principle, which minimizes speaker’s effort, the other from his Q-Principle, which minimizes hearer’s effort, (discussed more in section 3). While his system is not intended to provide an account of utterance processing like RT, nor does it have Grice’s philosophical analytical concerns; in fact, as I will argue later, due to his quite different outlook, Horn’s conception of ‘what is said’ and of the maxims/principles departs significantly from Grice’s. Horn’s orientation is linguistic; he is primarily interested in the interaction of pragmatics with the grammar (for instance, the distribution of different kinds of anaphor) and the way in which pragmatic factors affect lexicalisation by, for instance, promoting abbreviation of frequently used forms (so saving speaker effort) and blocking other forms (that might increase hearer effort) (see Horn 1984, 2004). Thus, he has concentrated on the most systematic and least context-sensitive aspects of pragmatics, that is, short-circuited and generalized conversational implicature, as opposed to particularized implicature.³ Horn’s linguistic pragmatics has very much the flavour of a kind of competence system, a component of the language faculty, interacting with the components of grammatical and lexical knowledge.⁴

In line with this competence orientation, his system has been recently recruited by scholars interested in extending the approach of Optimality Theory, previously applied successfully in generative phonology and syntax, to semantics and pragmatics (see, in particular, Blutner 1998, 2004; Blutner and Zeevat 2004). The following bidirectional optimality-theoretic pragmatics (adapted from Blutner and Zeevat 2004: 12-13) gives a sense of the way in which Horn's principles can be used on this approach:

1. A form-meaning pair (f, m) is optimal iff it is realized by *Gen* and it satisfies both the R- and the Q-principles, where:
 - a. (f, m) satisfies the R-principle iff there is no other pair (f, m') realized by *Gen* such that $(f, m') < (f, m)$
 - b. (f, m) satisfies the Q-principle iff there is no other pair (f', m) realized by *Gen* such that $(f', m) < (f, m)$

Gen is an abstract generative mechanism that provides all the possible pairings of (syntactic) forms f with utterance meanings (interpretations) m , and $<$ is an ordering relation meaning 'less costly than' (all detail of the underlying metric omitted)⁵. This is clearly not intended as any kind of performance system, whether for production or comprehension, but as an abstract characterisation of the best pairings of form and interpretation, using the two perspectives on optimization provided by Horn's principles.

To conclude this section, we have three rather distinct projects, the Gricean, the neo-Gricean (specifically Hornian) and the relevance-theoretic, each with its own goals and orientation, but all intersecting with each other at certain points. Their differences can, at least to some extent, be laid at the door of the disciplines they each ally with: Grice with philosophical analysis, Horn with linguistics, in particular lexis, and RT with cognitive processing. However, even given their different perspectives, it seems unlikely that they are simply complementary in all respects. Consider the issue of the 'right' conversational/communicative maxims/principles. It would be odd, though not impossible, to find that there is one set of such principles that are useful for making an analytical semantics/pragmatics distinction, another set that interacts with the grammar, shaping it to some extent, and yet another that guides the hearer's processes of utterance comprehension. Certainly, both neo-Griceans (Horn and Levinson), on the one hand, and relevance theorists, on the other, argue in favour of their own principles, as opposed both to each other's and to Grice's.

2. The concepts ‘what is said’ and ‘explicature’ (or ‘implicature’), with particular reference to the question: ‘So what are we then to make of pragmatic enrichment or “intrusion”?’ (Horn 2005a: 192)

Most pragmatists nowadays agree that there is a central component of speaker meaning (communicated content), which is distinct both from conversational implicature and from a minimalist conception of ‘what is said’. Pragmatic processes of completion, saturation, expansion, free enrichment, or meaning modulation are required to develop the schematic linguistically-encoded meaning into this component of communicated truth-conditional content. (The literature abounds with examples, which I won’t repeat here in the interests of space.) Recognition of the ‘linguistic underspecification’ of utterance content is one of the most important developments since Grice’s work, but, although it is widely accepted across different frameworks, there is much less consensus on how it is to be accommodated.

In fact, the main bone of contention concerns not so much the level of what is variously labelled ‘explicature’ (RT), ‘implicature’ (Bach), ‘i-content’ or ‘what is said_{max}’ (Recanati), but more whether or not there is need for a minimal level of ‘what is said’ in addition to this pragmatically-imbued level of truth-conditional content. In characterizing the Gricean (and neo-Gricean) project, Horn points out that for Grice what someone has said is (a) part of what she means (in the sense of being backed by the complex audience-directed intentions that accompany utterer’s meaning quite generally) so that ‘nothing can be said that isn’t meant’, and is (b) closely related to the conventional meaning of the linguistic forms used (Horn 2005a: 192). The problem with this is that it seems to be impossible to find a level that consistently has both properties, that is, is both speaker meant and minimalist. Bach (1994: 144) sees this problem and explicitly opts to drop the ‘speaker meant’ requirement in favour of a wholly semantic notion of ‘what is said’ which comports with his ‘Syntactic correlation constraint’, according to which ‘what is said’ must correspond to the elements of the sentence used, their order and their syntactic character. Recanati (2001, 2004) also sees the problem, but takes a crucial property of ‘what is said’ to be that it is speaker meant and so goes for a pragmatic notion, ‘what is said_{max}’. Relevance theorists do the same, but drop the problematic term ‘what is said’ and adopt the label ‘explicature’ for that communicated proposition which is built on the linguistically-encoded meaning (as distinct from implicatures, which are wholly pragmatically inferred).⁶

What is Horn's view on 'what is said'? He is not entirely explicit about this, but seems to be drawn to Kent Bach's approach: 'As stressed by Bach (1994, 2001), some aspects of speaker meaning need not be considered either part of what is implicated or of what is said' (Horn 2004: 21); 'Bach retains a neo-classically Gricean semantic characterization of what is said, along with a post-semantic understanding of conversational implicature: it is implicatures, not implicatures, that can determine the relevant truth conditions ...' (Horn 2004: 22). Both here and in Horn (2005b), he extols the virtues of the concept of 'implicature', which is in fact pretty much coextensive with the RT concept of 'explicature', but which comes with the ultra minimalist Bachian concept of 'what is said'. It is this package that appeals to Horn, I think, since it provides a way of accommodating the various 'pragmatic intrusions' into truth-conditional content while maintaining the kind of 'what is said' which he needs as the basis for his Q and R principles (discussed in the next section). So it seems pretty clear that he too favours the minimal linguistic notion of 'what is said', which is not necessarily speaker-meant.

In this respect, Horn (like Bach) departs in a major way from Grice. It was essential to Grice that what is said is an aspect of speaker meaning. In his various attempts to find a satisfactory definition of 'what is said', the one component of the definition that was always present was the following: "'U (utterer) said that p'" entails "'U did something x by which U meant that p'" (Grice 1969/89b: 87)⁷. To drop this is to pull out a pivotal plank in the edifice of the Gricean project to provide a complete account of meaning which is 'relativized to the users of words or expressions' (Grice 1989a: 340). It is this non-negotiable aspect of saying that forced Grice to resort to the locution 'what the speaker made as if to say' in his discussion of nonliteral language use (Grice 1975/89b: 34). In such cases, it is clear that the proposition literally expressed is not something the speaker could possibly mean (e.g. 'You are the cream in my coffee' or an ironical use of 'Bill is an honorable man'), hence it cannot, strictly speaking, be something the speaker said.

As Neale (1992: 523) points out, 'on Grice's account, *one cannot unintentionally say something* (a fact that has interesting consequences for, e.g., slips of the tongue and misused expressions).' One of the interesting consequences alluded to is frequently missed by people who discuss slips and misuses. For instance, Korta and Perry (forthcoming): 'if I intend to say that Jesus Mari knows Latin, and I say "Jesus Christ knows Latin", I will not have said that Jesus Mari knows Latin, even if my hearers all rightly interpret my intention, and I will have said that Jesus Christ knows Latin, even though no one thinks I intended to do so.' The

latter conjunct is simply untrue, at least if it is meant to represent the Gricean view. Rather, the speaker has failed to say anything: being speaker meant is a necessary, though clearly not a sufficient, condition on being said.

The cost of the Bach/Horn account is an extra level of meaning: as well as linguistic expression type meaning (LET) and the two kinds of communicated content (explicature/ implicature and implicature), there is a level of ‘what is said’, which is minimally distinct from LET, in that it involves values for a subset of indexicals, just the so-called pure ones, such as ‘I’, ‘you’, ‘here’, ‘today’, which are supplied directly by the narrow (semantic) context of utterance, requiring no guidance from pragmatic maxims/principles geared to the recovery of speaker’s intentions. I have discussed this position in detail elsewhere (Carston 2002: section 2.5; Carston forthcoming), so will be very brief here. My basic point is that it is hard to see what this conception of ‘what is said’ buys one. It does not include values for demonstrative indexicals (‘he’, ‘they’, ‘we’, ‘this’, ‘those’, etc), it’s entirely unclear how it deals with ambiguity (which can usually only be resolved by considerations of wide, pragmatic context), and it often does not provide a full proposition but just a ‘propositional radical’ (as in cases like ‘He is tall enough’, ‘These are ready’). Bach says that it is needed ‘to account for the linguistically determined input to the hearer’s inference to what, if anything, the speaker intends to be conveyed in uttering the sentence’ (Bach 2001: 15). But this is surely what LET gives us – what is gained by the addition of, at best, one or two indexical values?

Horn (2005a: 192) maintains that there is an important role for what he characterises as a Gricean speaker-based notion of ‘what is said’. First, it is quite unclear what he means by this, given the problem of finding a notion which both falls within speaker meaning and is as pragmatically minimal as he seems to want (indeed, entirely semantic, if he is following Bach here). Setting that aside, the claim (following Saul 2002) is that ‘it is precisely when the hearer misunderstands the speaker’s utterance intentions, and comes to recognize this misunderstanding that we see the need for a speaker-based notion of both what is said and what is meant’ (Horn 2005a: 192). Bach, on the other hand, when discussing cases of speakers’ slips of the tongue and other misuses, says: ‘the speaker is not the final authority on what is said’ (Bach 1994: 144). Since hearers may misperceive or misconstrue, and speakers may mispronounce or misuse, who does have the ultimate say about what has been said in any given instance? As far as I can see, although a good case can be made in many instances and

agreement reached, there just is no objective fact of the matter about what is said (any more than there is about what is explicated or implicated).

Let's turn now to the pragmatically augmented level of truth-conditional content, with a few relatively uncontroversial examples by way of reminder. The material in curly brackets has been pragmatically supplied:

2. a. I haven't eaten {supper} {tonight}.
- b. She handed him the key and {then} he opened the door {with that key}.
- c. Everybody {in our pragmatics class} passed the exam.

There is a silly but persistent terminological issue here. Following Bach, Horn objects to the RT term 'explicature' because 'the strengthened proposition is not explicitly communicated' (Horn 2005a: 193). Well, it is communicated (speaker-meant) and it is not implicated, and, since there is no other level of communicated content, it must follow that, for Horn (and Bach), there is nothing that is explicitly communicated. This runs contrary to the intuition that one component of what is communicated is at least much more explicit than other components. And, even if one insists on equating 'explicit' with 'linguistically encoded', the content at issue is, in fact, partly explicit (linguistically encoded) and partly implicit (pragmatically inferred). In fact, the RT concept of 'explicature' is defined so as to accommodate different degrees of explicitness and implicitness, with the crucial core in all cases being some element of linguistically-given meaning (Sperber and Wilson 1986: 182; Carston 2002: section 2.3). What I like about the term 'explicature' (aside from its clarity and its transparent duality with 'implicature') is that it nicely captures an apparent paradox of human linguistic communication, which is that this partially pragmatically-determined proposition is about as explicit as we ever are, or can be.

Finally, in this niggly vein, the minimal orthographic difference between the terms "implicature" and "implicature" together with the longevity and much greater currency of the latter term mean that, more often than not, that pesky "i" in the first term is just not noticed and "implicature" occurs where "implicature" is intended (the reference list of the Horn paper I am here responding to is just one of many such instances). Given human perceptual frailty, we seem to have a term which intrinsically fails to conform to any maxim of communicative clarity (or speaker economy, for that matter).

3. Pragmatic principles: informativeness, relevance, and least effort (for whom?)

Horn (2005a) reiterates the position he has maintained since his important and influential 1984 paper, according to which, apart from the Quality (truthfulness) maxim(s), which he considers essential and unreducible (Horn 1984: 12; 2004: 13), the Gricean maxims should be reduced to two general principles. These are the Q Principle and the R Principle, the first of which is oriented to the interests of the hearer and the second to the interests of the speaker.

He invokes a number of forerunners to this position, including in particular Zipf (1949) and Martinet (1962). In developing an ecological account of human behaviour quite generally, Zipf emphasised the fundamental role played by a Principle of Least Effort: ‘the primary principle that governs our entire individual and collective behaviour of all sorts, including the behaviour of our language ...’ (Zipf 1949: vii), and distinguished speaker’s economy of effort from hearer’s. Martinet (1962) takes it that the two factors, whose interaction is responsible for language change, are ‘first, the requirements of communication, the need for the speaker to convey his message, and second, the principle of least effort, which makes him restrict his output of energy, both mental and physical, to the minimum compatible with achieving his ends.’ (Martinet 1962: 139).

According to Horn: ‘These same two antinomic forces – and the interaction between them – are largely responsible for generating Grice’s conversational maxims and the schema for pragmatic inference derived therefrom’ (Horn 1984: 11-12), and he sets up his Q and R Principles so as to more clearly reflect these ‘two functional economies of Zipf and Martinet.’ Even before looking into the details, this might strike one as a surprising endeavour, since, as the Martinet quote most clearly indicates, what is needed to make communication (conversation) work is pitched against the speaker’s natural inclination to conserve his energy. *Prima facie*, then, we would not expect communicative principles (or conversational maxims) to somehow include the very force they are designed to overcome. I am not disputing the role of speakers’ articulatory (and other) economies in language change, for which there is considerable evidence (and which is the focus of both Zipf and Martinet’s observations), but it is a very different matter to draw this into the precepts that govern effective communicative conduct.

Here are the two principles as they are standardly presented (Horn 1984, 1989, 2004). Note the neat symmetry: hearer-oriented vs. speaker-oriented, lower-bounding vs. upper-

bounding, Gricean Quantity 1 vs. Gricean Quantity 2, and the split between the four Manner maxims.

The Q Principle (Hearer-based):

MAKE YOUR CONTRIBUTION SUFFICIENT

SAY AS MUCH AS YOU CAN (modulo R)

Lower-bounding principle, inducing upper-bounding implicata

[Grice's first Quantity maxim and the first two Manner maxims]

The R Principle (Speaker-based):

MAKE YOUR CONTRIBUTION NECESSARY

SAY NO MORE THAN YOU MUST (modulo Q)

Upper-bounding principle, inducing lower-bounding implicata

[Grice's second Quantity maxim, Relation maxim and the second two Manner maxims]⁸

Although the one is hearer-based and the other speaker-based, they are both presented (like the Gricean maxims) as injunctions to the speaker, and each presumably has its hearer's corollary since 'The speaker and hearer are aware of their own and each other's desiderata [that is, the economies of both parties], and this awareness generates a variety of effects ...' (Horn 2005a: 19). I assume that the two formulations given for each principle are intended to have roughly the same import and that, therefore, the difference between 'your contribution', in the one, and what you 'say', in the other, is not to be taken as significant. When he's being more summary, Horn uses the second formulations (Horn 2005a: 196), and those are the versions that matter in discussions of the implicatures arising from these principles (applied to what was said and what was not said).

The Q principle has received a vast amount of attention (due mainly to the veritable industry that has developed around the phenomenon of scalar implicature). Here I'll focus instead on the R Principle. This is taken to correspond with Zipf's 'speaker's economy', that is, a force pressing in the direction of a single word or sound that would express all possible meanings, thereby sparing the speaker the mental effort that is involved in selecting the appropriate linguistic forms for the meanings she wants to convey (Zipf 1949: 20) and the physical effort of articulating them. At the same time, it is claimed to mop up a number of the Gricean maxims, including the relevance maxim. I find this quite puzzling. What need is

there for a conversational maxim or principle that enjoins a speaker to minimise her effort? First, she's going to do it anyway (modulo her goals, of course), as Zipf and Martinet tell us. Second, what the Gricean maxims are for (and so also the neo-Gricean maxims/principles derived from them) is to account for conversational implicature derivation. Conversational implicatures are assumptions that a hearer has to attribute to a speaker in order to preserve the presumption that the speaker is observing the maxims. But while a hearer has reason to be concerned that a speaker is being relevant, informative, truthful, and orderly, why would he care to preserve an assumption that a speaker is minimising her effort? The opposition of forces that interests Horn seems to be between a general non-communicative principle of effort minimisation and whatever principles/maxims are responsible for successful communication.

Let's consider the various Gricean maxims which the R Principle is claimed to subsume. First, the second Quantity maxim: 'Do not make your contribution more informative than is required'. Grice was unsure about this maxim, making a number of points against it, including the likelihood that it would be replaced by a properly developed maxim of Relation (relevance). It is clear from his discussion that he did not think of either of these maxims as being concerned with speaker effort. Rather, they were both intended to avoid 'confusing [the hearer] in that it [overinformativeness] is liable to raise side issues; and there may be an indirect effect, in that the hearers may be misled as a result of thinking that there is some particular *point* in the provision of the excess of information' Grice (1975/89b: 26-27, 34). In other words, these maxims are oriented towards sparing the *hearer* gratuitous processing effort. Now, briefly, let's consider the two Manner maxims that fall under the R Principle: 'Be brief', 'Be orderly'. These are surely not concerned with the speaker's effort (be brief so as to save yourself the physical energy of exercising your articulatory organs? be orderly in the way you present information because that will save you mental effort?). Like all the other Gricean maxims, these are aimed at achieving the speaker's intention of affecting the hearer in certain ways (getting him to entertain certain thoughts or perform certain actions) (Grice 1975/89b: 28).

The Q and R principles are claimed to have complementary bounding properties. The talk of bounds is reasonably clear in the Q cases: what is encoded or said sets a lower bound, e.g. 'at least three/some/possible' and the Q-based implicature provides an upper bound, e.g. 'at most three/some/possible'. It's considerably less clear how the alleged reverse pattern of bounds pans out in the R cases. What does the R Principle place an upper bound on – is it

speaker effort, or information given, or (somehow) both? Consider a couple of examples standardly cited as giving rise to R-based implicatures (or, in some cases perhaps, implicatures):

3. a. Utterance: I broke a finger yesterday.
 b. Implicature: The finger is mine.

4. a. Utterance: Can you reach the salt?
 b. Implicature: I request you to pass the salt to me.

In what sense are the implicatures here lower-bounding? The idea seems to be that on the basis of the proposition p that is said, there is an R-inference to ‘more than p ’ and the hearer has to go on to figure out what this more is. In the case of ‘and’-conjunctions (see (2b) above), it is a stronger relation between the conjuncts, in (3) it’s a narrowing to a specific domain of fingers, and in cases of indirect speech acts, such as (4), it is simply a different speech act from the (irrelevant) one directly expressed. But one could say much the same of the Q cases: something is explicitly said and what is communicated is stronger (more informative/relevant). In both kinds of case, there is a strengthening of communicated content: from ‘at least some’ to ‘just some’, from ‘a finger’ to ‘my finger’, etc. I find little support for the claim (Horn 1984: 14) that the effects of the two principles are mirror images of each other. (For fuller discussion of this point, see Carston 1998: section 3).

Another interesting feature of the examples governed by the R principle is that, in each case, it seems that whatever utterance apparently conforms to the principle of least speaker effort is also the low cost option for the hearer. He has immediate mental access to stereotypical scripts about the temporal sequence and consequence relations between events (for the ‘and’-conjunctions) and to the most frequently encountered finger-breaking scenario. Can we find cases where speaker’s effort and accessibility to the hearer might come apart? Suppose you ask me if I know what has become of our old college friends, Ann and Bob, whom you lost touch with a long time ago, and suppose the first thought to occur to me is that they recently had a baby. I next recall that they got married a while ago. Although this is the order of my thoughts, it is very likely (I think) that I will choose to give you this information via an utterance of ‘They got married and have had a baby’, which reflects the order in which

the events occurred, because this is the order from which you will most easily derive the appropriate (temporal, etc) implications.

Consider the following pairs of possible utterances:

5. a. Please pass the salt and pepper.
 b. Please pass the condiments.

6. a. Do you have any brothers or sisters?
 b. Do you have any siblings?

7. a. There's a lot of interesting animal life in this region.
 b. There's a lot of interesting fauna in this region.

The second member of each of the pairs should cost the speaker less articulatory effort than the first while, most likely, the first in each case is the more frequent, the less marked, and hence the more accessible (other things being equal) to the hearer. Note that it's not obvious that the members of these pairs are each pragmatically associated with complementary denotations, as is the case for the examples usually cited by Horn to illustrate the 'division of pragmatic labor', such as the 'pink'/'pale red', 'kill'/'cause to die', 'mother'/'father's wife' pairs (Horn 1984: 22-31; 2004: 16-17). The pairs in (5)-(7) seem to be synonymous. So which will a speaker choose? Other things being equal (as ever), the intuition is that she is likely to choose the more frequently used phrase over the briefer, less articulatorily demanding word. Furthermore, if she did choose the lexical form her utterance would tend to have a layer of effects (of a largely social sort: distance, formality) not present when the longer, more syntactically complex expression is used. So we may have a split here between speaker's and hearer's effort. I would suggest, though, that this sort of case really falls together with those that Horn presents ('kill'/'cause to die', etc), which have the apparently opposite properties in that it is the 'more complex, less lexicalised expression' that gives rise to some sort of special effect (Horn 2004: 17). However, this unification is only obviously possible if the effort at issue is that of the hearer rather than that of the speaker. Then, as predicted by the RT account, in all instances, the extra processing effort required is offset by particular cognitive effects which the less effort demanding variant would not have had.

As Horn often points out, Zipf wrote not only of the speaker's economy, but also of the 'auditor's economy' which, he claimed, tends towards the establishment of as many different linguistic expressions as there are messages to communicate (Zipf 1949: 21; Horn 1984: 11). I don't think Zipf is correct about this, at least if the implication is that for any meaning *m*, the least effortful, hence most economical, utterance from the hearer's point of view would be the one that fully encoded *m*. Even if this were possible, there is just no evidence that hearers prefer a given message or meaning to be as fully encoded as possible. A good speaker, as opposed to one who is found to be tedious, pedantic, even misleading, is one who does not over-encode, judging more or less correctly what a hearer can easily infer in a specific context: when a particular referent is highly salient, a pronoun is preferable to a name or a description; when a topic is established, a phrasal utterance ('On the table') may be preferable to a sentential one ('Mary put the book on the table'), etc. On the RT account, it is considerations such as these that inform a speaker's choice of form to utter.

Finally and very briefly, what about speaker effort within the RT account? Horn and others often say it is an entirely hearer-based account. In fact, considerations of speaker effort do play a role, although nothing like as central a role as Horn favours. Recall the definition of "optimal relevance" given in section 1: what a hearer is entitled to expect from an utterance is that it will be the most relevant one (have the most cognitive effects for the least cognitive expenses), modulo the speaker's abilities and *preferences* or goals. No doubt, speakers do often prefer to save themselves mental and physical effort, but this preference will usually be in interplay with other goals. A speaker may prefer not to risk offending her hearer and so speak in a roundabout way on a particular matter, she may prefer not to disclose some highly relevant but perhaps incriminating information, she may want to impress with her learned vocabulary, she may want her speaking to distract the hearer's attention from something, and so on. All of these and many other possible preferences (distinct from the fundamental communicative goal) affect the degree of effort a speaker is willing to expend in encoding her thoughts and articulating linguistic forms. So for the hearer, following the relevance-based comprehension procedure, consideration of the speaker's effort is but one of numerous factors that may affect his specific expectation of relevance on a given occasion of utterance.

References:

- Bach, Kent. 1994. Conversational implicature. *Mind and Language* 9:124-162.
- Bach, Kent. 2001. You don't say? *Synthese* 127: 11-31.
- Blutner, Reinhard. 1998. Lexical pragmatics. *Journal of Semantics* 15:115-62.
- Blutner, Reinhard. 2004. Pragmatics and the lexicon. In Horn, Laurence and Gregory Ward (eds.), *Handbook of Pragmatics*, 488-514. Oxford: Blackwell.
- Blutner, Reinhard and Henk Zeevat. 2004. Editor's introduction: pragmatics in optimality theory. In Blutner, Reinhard and Henk Zeevat (eds.), *Optimality Theory and Pragmatics*, 1-24. Basingstoke: Palgrave.
- Carston, Robyn. 1998. Informativeness, relevance and scalar implicature. In Carston, Robyn and Seiji Uchida (eds.), *Relevance Theory: Applications and Implications*, 179-236. Amsterdam: John Benjamins.
- Carston, Robyn. 2000. The relationship between generative grammar and (relevance-theoretic) pragmatics. *Language and Communication* 20: 87-103.
- Carston, Robyn. 2002. *Thoughts and Utterances: the Pragmatics of Explicit Communication*. Oxford: Blackwell.
- Carston, Robyn. 2004. Relevance theory and the saying/implicating distinction. In Horn, Laurence and Gregory Ward (eds.), *Handbook of Pragmatics*. Oxford: Blackwell. 633-656.
- Carston, Robyn. forthcoming. Pragmatics, semantics and semantics/pragmatics distinctions. *Synthese*.
- Chierchia, G. 2004. Scalar implicatures, polarity phenomena and the syntax/pragmatics interface. In Belletti, Adriana (ed.), *Structures and Beyond*. Oxford University Press.
- Dale, Robert and Ehud Reiter. 1995. Computational interpretations of the Gricean maxims in the generation of referring expressions. *Cognitive Science* 19: 233-263.
- Fodor, Jerry. 1983. *Modularity of Mind*. Cambridge Mass.: MIT Press.
- Gibbs, Ray. 1994. *The Poetics of Mind*. Cambridge University Press.
- Glucksberg, Sam. 2001. *Understanding Figurative Language*. Oxford: Oxford University Press.
- Green, Nancy and Sandra Carberry. 1999. Interpreting and generating indirect answers. *Computational Linguistics* 25(3): 389-435.

- Grice, H. Paul. 1967. Prolegomena. *Logic and Conversation*. William James lectures. Harvard University. Reprinted in Grice, H. Paul. 1989b, 3-21.
- Grice, H. Paul. 1969. Utterer's meaning and intentions. *Philosophical Review* 78, 147-177. Reprinted in Grice, H. Paul. 1989b, 86-116.
- Grice, H. Paul. 1975. Logic and conversation. In Cole, Peter and Jerry L. Morgan (eds.), *Syntax and Semantics 3: Speech Acts*, 41-58. New York: Academic Press. Reprinted in Grice, H.P. 1989b, 22-40.
- Grice, H. Paul. 1989a. Retrospective epilogue. In Grice, H. Paul. 1989b, 339-385.
- Grice, H. Paul. 1989b. *Studies in the Way of Words*. Cambridge: Harvard University Press.
- Horn, Laurence. 1984. Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In Schiffrin, Deborah (ed.), *Meaning, Form and Use in Context (GURT '84)*, 11-42. Washington: Georgetown University Press.
- Horn, Laurence. 1989. *A Natural History of Negation*. University of Chicago Press.
- Horn, Laurence. 2004. Implicature. In Horn, Laurence and Gregory Ward (eds.), *Handbook of Pragmatics*. Oxford: Blackwell. 3-28.
- Horn, Laurence. 2005a. Current issues in neo-Gricean pragmatics. *Intercultural Pragmatics* 2 (2): 191-204.
- Horn, Laurence. 2005b. The border wars: a neo-Gricean perspective. In Turner, Ken and Klaus von Heusinger (eds.), *Where Semantics meets Pragmatics*. London: Elsevier.
- Korta, Kepa and John Perry. forthcoming. Three demonstrations and a funeral. *Mind and Language*.
- Landman, F. 2000. *Events and Plurality*. Dordrecht: Kluwer.
- Levinson, Stephen. 2000. *Presumptive Meanings: The theory of Generalized Conversational Implicatures*. Cambridge: MIT Press.
- Martinet, Andre. 1962. *A Functional View of Language*. Oxford: Clarendon Press.
- Neale, Stephen. 1992. Paul Grice and the Philosophy of Language. *Linguistics and Philosophy* 15: 509-559.
- Noveck, Ira. 2004. Pragmatic inferences related to logical terms. In Noveck, Ira and Dan Sperber (eds.), *Experimental Pragmatics*, 301-321. Basingstoke: Palgrave.
- Recanati, Francois. 2001. What is said. *Synthese* 128: 75-91.
- Recanati, Francois. 2004. *Literal Meaning*. Cambridge: Cambridge University Press.
- Saul, Jennifer. 2002. What is said and psychological reality: Grice's project and relevance theorists' criticisms. *Linguistics and Philosophy* 25: 347-372.

- Sperber, Dan, Francisco Cara, and Vittorio Girotto. 1995. Relevance theory explains the selection task. *Cognition* 57: 31-95.
- Sperber, Dan and Deirdre Wilson. 1986. *Relevance: Communication and Cognition*. Oxford: Blackwell
- Sperber, Dan and Deirdre Wilson. 1995. Postface. *Relevance: Communication and Cognition*. Second edition. Oxford: Blackwell; Cambridge, Mass.: Harvard University Press.
- Sperber, Dan and Deirdre Wilson, 2002. Pragmatics, modularity and mind-reading. *Mind and Language* 17: 3-23.
- Wharton, Tim. 2002. Paul Grice, saying and meaning. *UCL Working Papers in Linguistics* 14: 207-248.
- Wharton, Tim. 2003. Natural pragmatics and natural codes. *Mind and Language* 18: 447-477.
- Wilson, Deirdre and Dan Sperber. 2004. Relevance theory. In Horn, Laurence and Gregory Ward (eds.), *Handbook of Pragmatics*. Oxford: Blackwell. 607-632.
- Zipf, George K. 1949. *Human Behaviour and the Principle of Least Effort*. Cambridge, MA: Addison-Wesley Press.

* Many thanks to Tim Wharton, Deirdre Wilson and Vladimir Zegarac for interesting and relevant discussion.

¹ What prompted Istvan Kecskes' question and part of what informs Horn's response is Saul's (2002) claim that the Gricean enterprise and the goals of Relevance Theory are different and, therefore, complementary. I note, however, that Saul's focus is entirely on Grice and not at all on current neo-Gricean theorising. The only theorist she mentions who could possibly be thought of as a neo-Gricean is Kent Bach and he departs from Grice in a number of important ways, especially with regard to 'what is said' (to be discussed in section 2).

² The nature of speaker meaning and its relation to ostensive communication has been addressed in some detail in the RT framework, by Sperber and Wilson (1986: chapter 1) and, more recently, by Tim Wharton, who argues that the domain of pragmatic principles or maxims extends beyond Grice's speaker's meaning (meaning-*nn*) to include cases of natural behaviours (hence natural meaning) which are deliberately shown (Wharton 2003).

³ There are several current attempts to draw these phenomena (especially the Q-based scalar implicature) more directly into the domain of generative grammar (Landman 2000, Chierchia 2004). On these accounts, an algorithmic process of implicature derivation is postulated to accompany the bottom-up computation of truth-conditional content, starting with the most deeply embedded clause and proceeding up the syntactic tree to the top clausal node. This is patently not a matter of what goes on in actual (left to right) comprehension but a clear instance of characterising an aspect of linguistic competence. It should be noted, however, that Horn himself does not favour this degree of integration with the grammar (see Horn 2005b).

⁴ See Fodor (1983: 1-10) on the distinction between knowledge systems and mental mechanisms. For further discussion of the competence/performance distinction and its bearing on pragmatic theorising, see Carston (2000).

⁵ This is not the version ultimately favoured by Blutner and Zeevat (2004), but for my purposes here absolutely nothing hangs on the detail and this version is more immediately transparent than the final (weaker) version.

⁶ How Grice himself would have resolved this issue is unclear to me. However, his insistence that 'saying' entails 'meaning' together with some suggestive comments on 'dictiveness without formality' in his *Retrospective Epilogue* (Grice 1989a: 361) might indicate a preference for the more pragmatically-enriched conception of what is said (see Wharton 2002 for arguments to this effect). For fuller discussion of the relation between semantics, 'what is said' and explicature, see Carston 2004.

⁷ All page numbers given for quotes from Grice's work refer to the reprints in Grice (1989b).

⁸ Originally, all four of Grice's manner maxims were collected under the R Principle (Horn 1984:13) and this is (tentatively) maintained by Blutner and Zeevat (2004: 12).