Reference and Relevance

DEIRDRE WILSON

Abstract

In this paper, I compare three accounts of reference assignment which have the same overall structure but differ in the pragmatic criteria they propose for evaluating alternative interpretations. I argue that the criteria proposed by Lewis (1979) and Clark (1977) are inadequate to deal with a certain range of examples which are more insightfully handled by the relevance-theoretic criterion proposed by Sperber & Wilson (1986).

1 Introduction

The assignment of reference to definite noun-phrases such as 'the man', 'the door', has been approached from several perspectives: semantic, pragmatic and psycholinguistic. In this paper, I will take a pragmatic perspective, considering reference assignment as part of an overall process of utterance interpretation whose goal is to recognise what propositions and propositional attitudes the speaker intended to convey. My theoretical framework will be broadly Gricean: I will assume without argument that utterance interpretation involves a substantial element of non-demonstrative inference, and that the inference process consists in a search, among candidate interpretations, for one that satisfies a pragmatic criterion which it is the task of pragmatic theory to describe. Grice's proposed criterion was based, of course, on his Co-Operative Principle and associated maxims of truthfulness, informativeness, relevance and clarity. In a number of recent works, Dan Sperber and I have argued for an alternative criterion of consistency with a principle of optimal relevance (see for example Sperber & Wilson 1986, 1987). The aim of this paper is to consider a range of broadly Gricean approaches to reference assignment, examine their proposed criteria for evaluating interpretations, and show that the relevance-theoretic criterion is preferable on both empirical and theoretical grounds.

2 The problem

The pragmatic literature on reference assignment has concentrated on two main problems:

(a) what happens when there is more than one previously established candidate referent and the hearer has to make a choice?
(b) what happens when there is no previously mentioned candidate and an appropriate referent has to be inferred?

These problems are illustrated in (1) and (2):

(1) a. The room had three doors, one of which was open. I closed the door.
    b. Sean Penn attacked a photographer. The man was quite badly hurt.
(2) a. I walked into the room. Both windows were open.
    b. Jane dropped linguistics. The lectures were boring.

In (1a), though the NP 'the door' could in principle refer to any of the three doors previously mentioned, the natural assumption is that it was intended to refer to the open one; in (1b), though the NP 'the man' could refer to either Sean Penn or the photographer, the natural assumption is that it was intended to refer to the photographer. The question raised by utterances like (1) is how the choice between the two explicitly mentioned candidates is made.

In (2) the problem is slightly different. Instead of having several candidate referents to choose from, the hearer has to go through an inference process to establish even one. In (2a), for example, it is nowhere explicitly stated that the room the speaker went into had two windows: in order to assign reference to the definite NP 'both windows', however, the hearer has to assume that the room had two windows and that these were the open ones. Similarly, in assigning reference to the definite NP 'the lectures' in (2b), the hearer has to assume that studying linguistics involves going to lectures and that the linguistics lectures were the boring ones.

Examples of this type have been investigated by Herb Clark and his associates under the name 'bridging'. Intuitively, bridging involves the retrieval or construction of a contextual assumption introducing a candidate referent, which is then used to assign reference to a subsequent definite NP. I will call the referent-introducing assumption the bridging assumption, or bridge. Thus, for (2a) the bridging assumption would be (3a), and retrieval of this assumption would make (2a) essentially equivalent to (3b):

(3) a. The room had two windows.
b. I walked into the room. The room had two windows. Both windows were open.

I will also be looking at a third set of examples, which has been relatively little studied, and which presents the hearer with a combination of the problems raised by (1) and (2): that is, it involves cases of bridging where there is not merely one possible bridging assumption but two. Consider (4a) and (4b):

(4) a. I switched from linguistics to geography. The lectures were less boring.
b. I ran from the classroom into the playground. The children were making too much noise.

In (4a) there are two potential bridges, (5a) and (5b), yielding the interpretations in (6a) and (6b), respectively:

(5) a. Studying linguistics involves going to lectures.
b. Studying geography involves going to lectures.

(6) a. The linguistics lectures were less boring than the geography ones.
b. The geography lectures were less boring than the linguistics ones.

Notice that whichever interpretation is chosen, both bridges will be used; the issue is, which of them will be used to assign reference to the NP 'the lectures' in (4a)? I take it that in normal circumstances the preferred interpretation would be the one based on (5b) and (6b).

In (4b) there are again two potential bridges and two potential interpretations, as in (7) and (8):

(7) a. There were children in the classroom.
b. There were children in the playground.

(8) a. The children in the classroom were making too much noise.
b. The children in the playground were making too much noise.

In this case, unlike the previous one, there is no need to use both bridges. I take it that in this case, without further information, either interpretation would be possible, though there may be a preference for the one based on (7b) and (8b).

The general problem that interests me, then, is this: given a definite description such as 'the man', 'the door', how does the hearer choose the intended referent from a range of candidate referents whose existence may not have been explicitly stipulated, and may therefore have to be inferred?

3 Some previous accounts

Examples such as (1) and (2) refute one traditional semantic analysis of definite descriptions, on which a definite NP such as 'the door' is taken to refer to the unique object in the universe of discourse which meets the description of being a door. In fact, the inadequacy of this analysis was already recognised in David Lewis's 'General Semantics' (1970/83), a classic sketch of a formal semantics for English. In the appendix to that paper, Lewis wrote:

> 'consider the sentence 'The door is open'. This does not mean that the one and only door that now exists is open; nor does it mean that the one and only door near the place of utterance, or pointed at, or mentioned in previous discourse, is open. Rather, it means that the one and only door among the objects that are somehow prominent on the occasion is open.'
> (Lewis 1970/83: 227)

With the introduction of the notion of prominence, cognitive factors explicitly entered the domain of semantics. An object might be made prominent, said Lewis, by being nearby, or pointed at, or mentioned, but the crucial consideration was more general than these. Contextual prominence, he said, was determined by 'mental factors such as the speaker's expectations regarding the things he is likely to bring to the attention of his audience' (Lewis 1970/83: 227-8). Thus the role of cognitive factors in reference assignment was explicitly acknowledged.

This insight was developed in Lewis's paper 'Scorekeeping in a language game' (1979/83). Here, the earlier notion of prominence was replaced by a notion of salience, where an object could become salient, for example, by being easily noticeable, or by being mentioned in discourse. Lewis then proposed a general rule for assigning reference to definite descriptions: a definite description refers to the most salient object in the domain of discourse meeting that description. Consider (9), for example:

(9) The dog got in a fight with another dog.

What this means, says Lewis, is that the most salient dog got in a fight with some less salient dog.

Finally, Lewis invoked a notion of pragmatic acceptability, and argued that if the rule of reference assignment yielded pragmatically unacceptable results, a reinterpretation must take place. To motivate this addition, he used the following scenario. At a certain point in a conversation, there are two salient cats: mine, which is with us in the room, and yours, which you've left at home but have just been telling me about. At a point when your cat is the most salient one, mine gets under your feet, and I say:

(10) Watch out, you're about to tread on the cat.

Question: which cat should you take me to have meant? Although your cat is the most salient one, it can't be the one I meant to refer to, says Lewis, because on that interpretation my utterance will be pragmatically unacceptable. Accordingly, he proposed a 'rule of accommodation for comparative salience', which runs as follows:

> *Rule of accommodation for comparative salience*
> If at time t something is said that requires, if it is to be acceptable, that x be more salient than y; and if, just before t, x is no more salient than y; then - ceteris paribus and within certain limits - at t, x becomes more salient than y.

In other words, for pragmatic reasons, the relative salience of objects can be shifted so that the output of the reference assignment rule is pragmatically acceptable.

Lewis's remarks on the notion of pragmatic acceptability are extremely brief. He merely mentions three contributory factors. First, truth: an utterance is pragmatically unacceptable if it is untrue; second, non-triviality, by which he means, essentially, informativeness: an utterance is pragmatically unacceptable if it is uninformative; and third, what he calls warranted assertion, by which he means, essentially, that an utterance is pragmatically unacceptable if it is unevidenced. These factors have obvious similarities to Gricean maxims, though no maxim of relevance is mentioned. His claim is, then, that reference is assigned to referring expressions depending, on the one hand, on the comparative salience of candidate referents, and on the other hand, on the desire to obtain an overall interpretation that satisfies a pragmatic criterion requiring an utterance to be true, informative and evidenced.

Herb Clark's analysis of the bridging inferences illustrated in (2) and (4) above has the same overall structure. In his best-known paper, 'Bridging' (1977), Clark discusses examples such as (11):

(11) I looked into the room. The ceiling was high.

Taking a resolutely cognitive approach, he argues that reference assignment involves not so much the location of an appropriate object as the retrieval of an appropriate mental representation from memory. Sometimes, as in (11), an appropriate representation will have to be constructed ad hoc, and it is this ad hoc construction process that he calls bridging.

Which bridge should the hearer construct? Why is it that in (11), for example, the appropriate bridge is the one that links the NP 'the ceiling' to the ceiling in the room the speaker looked into, rather than, say, the ceiling in the speaker's bedroom at home? Clark's answer is that the appropriate bridge will be the shortest plausible one, i.e. the shortest bridge which contains only assumptions that in the circumstances are likely to be true.

To get an intuitive understanding of how bridges may vary in length, consider the following examples:

(12) a. John and Mary went out last night. The three witches were excellent.
b. John and Mary went to the theatre last night. The three witches were excellent.
c. John and Mary went to a play last night. The three witches were excellent.
d. John and Mary went to *Macbeth* last night. The three witches were excellent.

The examples are given in increasing order of acceptability, with (12a) being the least acceptable and (12d) the most. These differences in acceptability might be explained in terms of the length of the bridge needed to introduce an appropriate referent for the NP 'the three witches'. In (12a), for example, in accessing an appropriate referent, the hearer has to construct or retrieve the following assumptions:

(13) a. John and Mary went to the theatre.
b. At the theatre they saw a play.
c. The play they saw was *Macbeth*.
d. *Macbeth* has parts for three witches.

In (12d), by contrast, he has to retrieve only the information in (13d). Notice that, like Lewis's notion of salience, Clark's notion of length of the bridge is a factor affecting the accessibility of referents: the shorter the bridge, the more accessible the referent, and in this case the more acceptable the overall result.

On pragmatic acceptability, Clark is no less vague than Lewis. He claims, first, that the bridge must be short, and, second, that the bridge must be plausible: it is the one, he says, 'that requires the fewest assumptions, yet whose assumptions are all plausible given the listener's knowledge of the speaker, the situation, and facts about the world' (Clark 1977: 420). Finally, the resulting interpretation must satisfy a pragmatic criterion called the Given-New Contract, which is introduced as an extension of Grice's Co-Operative Principle (see Clark & Haviland 1977), and applies to all utterances containing referring expressions. To satisfy the Given-New Contract, an utterance must meet two conditions: (a) the intended referent must be uniquely identifiable from memory, and (b) the utterance must make some informative assertion about that referent. In other words, for a bridging utterance such as (2), (4) or (11) to be pragmatically acceptable, the bridge must be short, the bridge must be plausible, the referent must be uniquely identifiable, and the resulting overall interpretation must be informative.

Clearly, the proposals of Clark and Lewis have a common structure, and clearly this structure is essentially correct. Reference assignment is affected, on the one hand, by the accessibility of potential referents, and on the other hand, by the pragmatic acceptability of the resulting overall interpretation. What is curious, though, is how cursory the analysis of pragmatic acceptability is on these accounts. Neither Clark nor Lewis makes any reference to relevance;[1] both appeal to commonsense notions of truth, degree of evidence, and informativeness. It might seem, indeed, that though commonsensical, the accounts are so vague as to be unfalsifiable. I will argue that, vague though they are, they make definite predictions, and that these predictions are false. Before doing so, however, I want to sketch a relevance-theoretic account of reference assignment and to show that it has much the same structure as those

[1]Clark & Haviland (1977) introduce the Given-New Contract as an extension of Grice's Co-Operative Principle. Clark may therefore intend to invoke the Co-Operative Principle and maxims as additional constraints on interpretation - a possibility that is considered below.

of Clark and Lewis, differing only in the criterion of pragmatic acceptability proposed.

4 Relevance theory

Relevance theory, unlike Grice's pragmatic theory, is not a rule-based or maxim-based system. It is based on a few very simple ideas. First, that every utterance has a variety of possible interpretations, all compatible with the information that is linguistically encoded. Second, that not all these interpretations are equally accessible on a given occasion. Third, that hearers are equipped with a single, very general criterion for evaluating interpretations as they occur to them. And fourth, that this criterion is strong enough to exclude all but at most a single possible interpretation, so that the hearer is entitled to assume that the first interpretation that satisfies the criterion is the only one.

The criterion that Dan Sperber and I proposed in our book *Relevance* develops from a basic assumption about human cognition. The assumption is that human cognition is relevance-oriented: we pay attention to information that seems relevant to us. Now every act of communication starts out as a request for attention. As a consequence, it creates an expectation of relevance. It is around that expectation of relevance that our criterion is built.

Relevance is defined in terms of contextual effect and processing effort. Contextual effects are achieved when newly presented information interacts with a context of existing assumptions in one of three ways: by strengthening an existing assumption, by contradicting and eliminating an existing assumption, or by combining with an existing assumption to yield a contextual implication: that is, a logical implication derivable neither from the new information alone, nor from the context alone, but from the new information and the context combined. We claim that newly-presented information is relevant in a context when and only when it achieves contextual effects in that context, and the greater the contextual effects, the greater the relevance.

Contextual effects, however, do not come free: they cost some mental effort to derive. The effort needed to compute the contextual effects of an utterance depends on three main factors: first, the linguistic complexity of the utterance; second, the accessibility of the context; and third, the inferential effort needed to compute the contextual effects of the utterance in the chosen context. Each increment in processing effort detracts from overall relevance. So: the greater the contextual effects, the greater the relevance; and the smaller the effort needed to achieve those effects, the greater the relevance.

Every utterance, we said, creates expectations of relevance, and around that expectation our criterion for evaluating possible interpretations is built. In our book, we define a notion of optimal relevance which is meant to spell out what the hearer is looking for in terms of effort and effect:

> *Optimal relevance*
> An utterance, on a given interpretation, is optimally relevant iff:
> (a) it achieves enough effects to be worth the hearer's attention;
> (b) it puts the hearer to no unjustifiable effort in achieving those effects.

A word of explanation about each of these clauses.

On the effect side, what the hearer is entitled to look for is enough effects to make the utterance worth his while to process. In general, what that means is that he's entitled to expect more effects than he would have got from any other information that he could have been processing at the time. How much that is depends on what is going on elsewhere in his cognitive environment. For example, suppose that someone walks into an important lecture and says (14):

(14) Ladies and gentlemen, I have to tell you that the building's on fire.

'The building' is a referential expression, and different assignments of reference will lead to different levels of contextual effect. In the circumstances, the first hypothesis to come to the audience's mind will be that 'the building' means the building where the lecture is taking place. Clearly the utterance, on this interpretation, would have enough effects to be worth the audience's attention: their minds would be immediately filled with thoughts of how to get out. Given that at a formal lecture the audience is supposed to be entirely absorbed in what the lecturer is saying, it is hard to see what other interpretation would have enough effects to justify the interruption; and in these circumstances, the interpretation just suggested is basically the only possible one.

It might be thought that in other circumstances the intended interpretation would be harder to pin down. Surely there might be several radically different combinations of content and context, all of which would yield enough contextual effects to make the utterance worth the audience's attention? That is where clause (b) of the definition of optimal relevance comes in. Clearly, a speaker who wants to avoid any risk of misunderstanding should make sure that there is no interpretation more accessible to the hearer than the intended one, which has enough effects to be worth the hearer's attention, and which might therefore lead him astray. Clause (b) of the definition of optimal relevance, which excludes the unjustifiable expenditure of effort in recovering the intended effects, covers this type of case: that is, it excludes the possibility that the hearer will be expected to recover, process and accept the wrong interpretation before lighting on the intended one. From clause (b), it follows that a speaker aiming at optimal relevance should try to formulate her utterance in such a way that the first acceptable interpretation to occur to the hearer is the one she intended to convey. From the hearer's point of view, this clause has an immediate practical consequence. Having found an interpretation which satisfies his expectations of relevance in a way the speaker might manifestly have foreseen, he need look no further. The first such interpretation is the only such interpretation, and is the one the hearer should choose.[2]

I should note here in passing that in order to be acceptable and comprehensible, an utterance does not actually have to *be* optimally relevant. Suppose that, as you are about to leave the house, I warn you that it is raining; as it happens, this is something you already know. In the circumstances, the proposition I have expressed will have no contextual effects and hence be irrelevant to you; nonetheless, my utterance will be both comprehensible and acceptable as long as you can see how I might reasonably have expected it to be relevant. The actual pragmatic criterion we propose, then, is a criterion of consistency with the principle of relevance which takes account of this type of case:

> *Criterion of consistency with the principle of relevance*
> An utterance, on a given interpretation, is consistent with the principle of relevance if and only if the speaker might rationally have expected it to be optimally relevant to the hearer on that interpretation.

Vague as this criterion may sound, it makes one important prediction that I have not seen matched elsewhere. This prediction follows from clause

[2]We are not claiming, of course, that the interpretation selected by this criterion is invariably correct; merely that it is the only one the hearer is rationally justified in choosing. In a broadly Gricean framework, utterance interpretation is a non-demonstrative inference process which takes place at a risk; there is no guarantee that its outcome will be correct.

(b) of the definition of optimal relevance and its consequence that the first interpretation tested and found consistent with the principle of relevance is the only interpretation consistent with the principle of relevance. Let us assume that, in interpreting an utterance, the hearer starts with a small initial context left over, say, from his processing of the previous utterance; he computes the contextual effects of the utterance in that initial context; if these are not enough to make the utterance worth his attention, he expands the context, obtaining further effects, and repeats the process until he has enough effects to make the utterance optimally relevant in a way the speaker could manifestly have foreseen. At that point, he has an interpretation consistent with the principle of relevance, and it follows that he should stop; or, at least, he is entitled to go on on his own account but is not entitled to assume that the speaker intended to communicate anything more. In other words, all the hearer is entitled to impute as part of the intended interpretation is the *minimal* context and set of contextual effects that would be enough to make the utterance worth his attention. I will make use of that fact below.

5 Relevance theory and reference

How, within this framework, might the hearer identify the intended referent of a definite NP such as 'the door'? In *Relevance*, we assume, as does Herb Clark, that reference assignment involves the retrieval or construction of an appropriate conceptual representation, one that uniquely identifies the intended referent. We assume, furthermore, that this representation is incorporated into a representation of the proposition expressed by the utterance, so that in interpreting (15a), for example, the hearer might construct a conceptual representation such as (15b), which will act as input to further inference processes:

(15) a. The door is open.
b. Door d is open at time t.

These are not, incidentally, assumptions that a semanticist like Lewis would necessarily share. Lewis is interested in assigning the correct truth conditions to an utterance such as (15a). On his analysis, the utterance would simply be evaluated against a package of contextual information containing, among other things, a salience ranking of the candidate referents. There is no suggestion that a representation of the selected candidate should be integrated into a conceptual representation such as (15b).

Any examination of the pragmatics of reference assignment is complicated by the fact that tentative assignments are apparently made while the utterance is still in progress. What strategy should hearers pursue in making these tentative choices? There is clear experimental evidence about what they actually do. In this, as in every other aspect of utterance interpretation, they follow the path of least effort; or rather, they follow this path unless it already seems likely to them that the resulting interpretation will be pragmatically unacceptable.

On the face of it, this behaviour is rather surprising. What right do hearers have to assume that the interpretation they find easiest to construct will be the one the speaker intended? Relevance theory suggests an explanation. As we have seen, a speaker aiming at optimal relevance must put her hearer to no unjustifiable effort in achieving the intended effects. Clearly, a speaker who allows her hearer, even momentarily, to recover, process and accept the wrong interpretation would put him to considerable unjustifiable effort. Relevance theory thus sheds light on the least-effort strategy automatically adopted by hearers in utterance interpretation. I know of no other pragmatic theory which does the same.[3]

Returning now to our original examples, consider how (1a) and (2a) might be analysed within this framework. On hearing the definite NP 'the door' in (1a), the hearer will have immediately accessible (from his processing of the first part of the utterance) the information that the room the speaker was talking about had three doors, one of which was open. Suppose, then, that the open door (strictly speaking, the hearer's mental representation of the open door) is the most accessible one, and therefore the first to be tested for pragmatic acceptability. It is easy to see how the utterance might be acceptable on this interpretation. We can all think of circumstances in which a speaker might find it relevant to mention that she had closed an open door, and of contextual effects that the utterance, on this interpretation, might be intended to achieve: it might imply, for instance, that the speaker wanted to shut off noise from outside, to signal that the lecture was about to begin, to deter latecomers, and so on. By the criterion of consistency with the principle of relevance, this first acceptable interpretation will be the only acceptable interpretation, and is the one the hearer should choose.

---

[3]Clark & Haviland (1977), for example, build into their analysis of utterance interpretation the idea that the first acceptable candidate should be selected, but their framework offers no explanation of why this should be so. Perhaps an explanation could be constructed around Grice's Manner maxim 'Avoid obscurity'; but I know of no serious attempt within the Gricean framework to develop such an account.

Suppose, alternatively - a hypothesis that is implausible in this case, but might be worth considering in others - that representations of all three doors are equally accessible. Then, in order to choose between them, the hearer must look for a context which combines with one of them to yield an interpretation consistent with the principle of relevance. As noted above, it is easy to think of effects that a speaker might have intended to achieve by mentioning that she had closed an open door; it is much harder to think of effects that she might have intended to achieve by mentioning that she had closed a door that was already shut. By the criterion of consistency with the principle of relevance, the first accessible context to yield an acceptable overall interpretation is the only acceptable context, and is the one the hearer should choose.

Whichever supposition we make about the relative accessibility of candidate referents, the result will be the same. Either the open door will be the most accessible candidate, and will combine with an easily accessible context to yield an overall interpretation consistent with the principle of relevance; or all three doors will be equally accessible, and the choice among them will be determined by the accessibility of potential contexts with which they might combine to yield an overall interpretation consistent with the principle of relevance; here again the choice will fall on the open door. These two factors - accessibility of referents and accessibility of contexts - are crucial to the relevance-theoretic account of reference assignment.

Notice, incidentally, that we are now in a position to say something about the stylistic infelicity of examples such as (16):

(16) The room had three doors, all but one of which were closed. I closed the door.

Although in (16), as in (1a), there is only a single open door, which the speaker must be presumed to have closed, (16), on this interpretation, could not be optimally relevant. It could not be optimally relevant because, although it might achieve adequate effects, it would put the hearer to some unjustifiable effort in deriving them. In the first place, the existence of an open door has to be inferred in (16), whereas it was explicitly stipulated in (1a); yet this extra inferential effort is not offset by a corresponding decrease in linguistic complexity. In the second place, intuitively at least, (16) focuses the hearer's attention on the doors which were closed, thus making these the first to be tested and adding to the hearer's effort. More generally, it follows from the definition of optimal relevance that an utterance, on a given interpretation, can not be optimally relevant if some other utterance that would have been equally easy for the speaker to produce would manifestly have achieved the intended effects more economically. This sheds some light on the stylistic infelicity of examples (12a)-(12c) above.[4]

In analysing the bridging example (2a) above, the only additional factor that has to be taken into account is that there is no ready-made referent waiting to be retrieved and the hearer must engage in some form of inference to derive one. As always, the most accessible (here, easily inferable) referent will be the first to be tested, and the result will be accepted if it meets the overall criterion of consistency with the principle of relevance.

Here we must make some tentative assumptions about how inferred referents are accessed. In *Relevance*, as mentioned above, we assume that the hearer starts out with an initial context consisting of the information he has most recently processed, and that this can be expanded in a variety of directions: for instance, by adding encyclopaedic information about objects mentioned in the utterance or the initial context. Thus, in (2a), the hearer might expand the context by adding to it the information that rooms may have one or more windows; he could then make the hypothesis that the room the speaker went into had two windows, and use these in assigning reference to the definite noun-phrase 'both windows'. It is easy to think of effects that a speaker might expect her utterance, on this interpretation, to achieve: for example, that the room was cold, that the windows should be closed; that the room was well-aired and comfortable; that the occupants of the house had not gone out, and so on. Hence, inferred referents work in just the same way as explicitly mentioned referents, and are affected by the same two factors: accessibility of referents and accessibility of potential contexts with which these might interact to yield an overall interpretation consistent with the principle of relevance.

---

[4]What we are claiming goes beyond the idea that there is some absolute limit on the amount of effort the interpretation of an utterance may require. To satisfy himself that an utterance, on a given interpretation, is consistent with the principle of relevance, the hearer has take into account, not whether some absolute threshold of effort has been passed, but whether some other utterance could have achieved the same effects more economically. Consider (i), for example:

(i) John and Mary went to see the Scottish play last night.
The three witches were excellent.

Whether a hearer finds this utterance acceptable may depend on his familiarity with the superstition in certain acting circles that bans the use of the name *Macbeth* on pain of bad luck. In such circles, (i) may well be the most economical way available to the speaker of achieving the intended effects.

6 A problem with Clark and Lewis's criterion

A detailed comparison of these three approaches to reference assignment is beyond the scope of this paper. Such a comparison might explore the relations between Lewis's notion of salience (a property of objects) and Sperber & Wilson's notion of accessibility (a property of mental representations), or between Clark's contract-based approach to pragmatics and Sperber & Wilson's principle-based approach. Here my ambition is more limited. As we have seen, the three approaches to reference assignment have a common overall structure, differing only in the pragmatic criteria they propose for evaluating alternative interpretations. I want to compare these pragmatic criteria by looking at the predictions they make across a certain range of examples, and to argue that the relevance-theoretic criterion sheds light on these examples in a way that Clark and Lewis's criteria do not.

Comparisons are not easy, since the proposed analyses are quite complex. In particular, although all are sensitive to the accessibility of candidate referents, none of them claims straightforwardly that the most accessible referent is invariably selected. That claim could be straightforwardly refuted by comparing examples such as (1b) above and (17):

(1) b. Sean Penn attacked a photographer. The man was quite badly hurt.
(17) Sean Penn attacked a photographer. The man must be deranged.

In each of these examples there are two candidate referents for the NP 'the man': in (1b) the intended referent is the photographer; in (17) it is Sean Penn. An account based on accessibility of referents alone can not explain these results. It might deal with (1b) by assuming that the most recently mentioned man is the most salient/accessible one; but this would make the wrong predictions about (17). It might deal with (17) by assuming that the referent of a preceding subject NP is the most salient/accessible one; but this would make the wrong predictions about (1b). Since the first parts of these two utterances are identical, it seems that whatever man is made most accessible by one of them must be made most accessible by the other, and an account based on accessibility of referents alone would make the wrong predictions about at least one of these examples.[5]

Lewis could, however, appeal to his rule of accommodation for comparative salience to explain these results. According to Lewis, although

[5]See Kleiber (1990, forthcoming) for further important arguments against the view that accessibility of referents is the only factor affecting reference assignment.

reference is invariably assigned to the most salient candidate, initial salience rankings must be adjusted so that their output is pragmatically acceptable. By Lewis's criterion, to be pragmatically acceptable, an utterance must be true, informative and evidenced. As we have seen, in (1b) 'the man' should be interpreted as referring to the photographer, but this is also the interpretation on which (1b) is most likely to be true. Similarly, in (17) 'the man' should be interpreted as referring to Sean Penn, but this is also the interpretation on which (17) is most likely to be true. Hence, the rule of accommodation, interacting with a criterion of overall pragmatic acceptability, should have no trouble with examples such as (1b) and (17).

Or consider (4a) above, where there are two candidate referents for the NP 'the lectures' and the intended referent is the geography lectures rather than the linguistics ones:

(4) a. I switched from linguistics to geography. The lectures were less boring.

Here the potential bridges are (5a) and (5b), with resulting interpretations (6a) and (6b), respectively, of which (6b) will be preferred:

(5) a. Studying linguistics involves going to lectures.
 b. Studying geography involves going to lectures.

(6) a. The linguistics lectures were less boring than the geography lectures.
 b. The geography lectures were less boring than the linguistics lectures.

As it stands, Clark's analysis fails to explain why (6b) is preferred. He claims, on the one hand, that the selected bridge must be both short and plausible: this would not distinguish between (5a) and (5b), which are equally short and equally likely to be true. He also claims that the resulting overall interpretation must be informative; but this would not distinguish (6a) from (6b), which would both, quite clearly, be informative.

Let us assume, then, that Clark, like Lewis, would be prepared to evaluate not only the bridge but also the resulting overall interpretation for likelihood of being true. This would enable him to explain the preference for (6b). Suppose you were told that Mary had switched from linguistics to geography, and then asked whether it was more likely that Mary found the linguistics lectures less boring than the geography lectures, or the geography lectures less boring than the linguistics lectures. You would no doubt reply that it is more likely that she found the geography lectures less boring than the linguistics lectures. Thus Clark, like Lewis, could claim that the choice of an appropriate referent is affected by the resulting interpretation's likelihood of being true.

It is more plausible to suppose, however, that in line with his general commitment to a Gricean approach, Clark intends to supplement his Given-New Contract with some version of Grice's maxims, in which a criterion of truthfulness would replace Lewis's criterion of truth. On this approach, overall interpretations would be evaluated not for likelihood of being true but for likelihood that the speaker *believes* that they are true. An approach based on such a criterion would have no difficulty in distinguishing interpretations (6a) and (6b). Suppose you were told that Mary had switched from linguistics to geography and then asked whether it was more likely that Mary believed that the linguistics lectures were less boring than the geography lectures, or that the geography lectures were less boring than the linguistics lectures. You would no doubt reply that it is more likely that she believed that the geography lectures were less boring than the linguistics lectures. Thus Clark could claim that the choice of an appropriate referent is affected by the resulting interpretation's likelihood of being, if not true, at least truthful, i.e. believed by the speaker to be true.

To refute these accounts, we need an utterance with the following structure: first, it has two possible overall interpretations, which are equally informative, equally evidenced, and equally likely to be true (or truthful); second, hearers systematically prefer one interpretation to the other; and third, this choice cannot be explained by appealing to differences in the inherent accessibility of referents. If Clark and Lewis are right, utterances of this type should not exist. If they do, something is missing from their accounts.

Example (18) seems to me a case in point:

(18) I switched from linguistics to geography. The lectures were just as boring.

Here there are two potential bridges, (19a) and (19b), and two possible overall interpretations, (20a) and (20b):

(19) a. Studying linguistics involves going to lectures.
b. Studying geography involves going to lectures.

(20) a. The linguistics lectures were just as boring as the geography lectures.
b. The geography lectures were just as boring as the linguistics lectures.

The natural interpretation of (18), I take it, is (20b), with bridging assumption (19b) being used to assign reference to the NP 'the lectures'. How might this be explained on Clark's (or Lewis's) account? The choice is not determined by the length or plausibility of the bridge, since the two potential bridges are equally short and equally plausible. It must therefore depend on the resulting interpretation's likelihood of being true (or truthful), informative or evidenced. But as far as I can tell, interpretations (20a) and (20b) are logically equivalent, and hence equally likely to be true (or truthful), equally informative, and equally evidenced. Thus, the pragmatic criteria proposed by Clark (and Lewis) can not explain the preference for (20b) over (20a).

On Clark and Lewis's approaches, the choice of (20b) would therefore have to be explained by appeal to variations in the accessibility of candidate referents. It might be claimed, for example, that though both bridges are equally short and plausible, one of them is inherently more accessible than the other. This is a reasonable claim, since bridging assumptions are contextual assumptions, which quite generally vary in accessibility. Let us suppose that the inherent accessibility of a bridge is determined by the syntactic position of its antecedent NP. We might then maintain that the bridge based on the most recently mentioned antecedent NP is the most accessible one, and thus explain why the bridge in (19b), based on the most recently mentioned NP, 'geography', is preferred to the bridge in (19a).

This hypothesis will not do. If it did, then by changing the order of the NPs in (18) we should be able to change the interpretation. But compare (18) and (21):

(18) I switched from linguistics to geography. The lectures were just as boring.
(21) I switched to geography from linguistics. The lectures were just as boring.

Clearly, the change in the order of NPs does not lead to a change in interpretation. In both these utterances, the preferred bridge for the NP 'the lectures' is (19b), giving rise to the overall interpretation (20b).

Indeed, it is hard to conceive of any accessibility ranking based on surface syntactic position alone that would explain the full range of choices that hearers apparently make in examples of this type. In each of (22a-e), for instance, there is a clear preference, in interpreting the NP 'the lectures', for the bridge based on the antecedent NP 'geography', despite the variations in its surface position:

(22) a. Geography followed linguistics. The lectures were just as boring.
b. Linguistics was followed by geography. The lectures were just as boring.
c. Linguistics gave way to geography. The lectures were just as boring.
d. After linguistics came geography. The lectures were just as boring.
e. Geography came after linguistics. The lectures were just as boring.

Nor can these choices be explained by invoking a rule of accommodation which would adjust inherent salience rankings to meet some overall criterion of pragmatic acceptability. As we have seen, the resulting overall interpretations of (18) and (21) are logically equivalent, and hence equally likely to be true (or truthful), informative and evidenced; hence the pragmatic criteria proposed by Clark and Lewis do not choose between them.

To save these criteria, one would therefore have to find an alternative accessibility ranking, based on something other than surface syntactic position, that would explain the preference for the bridge in (19b) across the full range of examples (18), (21) and (22). One might appeal, for instance, as do Erku & Gundel (1987), to a notion of topic, and claim that the most accessible bridge is the one based on the topic NP; one might appeal to an accessibility hierarchy based on underlying thematic role (see Keenan & Comrie 1977); or one might explore the notion of focus used in some of the recent AI literature (see for example Asher & Wada 1988). Let me emphasise that I am not against attempts to explore these possibilities. Clearly, reference assignment is affected by accessibility of referents, and it may turn out that surface syntactic position of the antecedent NP is not the only factor affecting accessibility of referents in this type of example. My point here is that Clark and Lewis's criteria force them to appeal to variations in accessibility of referents in analysing these examples; that in the absence of any clear account of how accessibility is determined, the resulting analyses cannot but be unilluminating; and that the relevance-theoretic criterion, as I will shortly argue, offers some immediate insight into how these examples work.

There is, however, a more serious objection to Clark and Lewis's criteria. As we have seen, they claim that overall interpretations are evaluated for truth (or truthfulness), informativeness and degree of evidence. In fact, there are two conceptions of how the evaluation procedure might work. On the first, and weaker, conception, an interpretation would be acceptable as long as it was true (or truthful), informative and evidenced; on the second, and stronger, conception (the one I have tacitly used above), the only acceptable interpretation would be the one that was *most* likely to be true (or truthful), *most* informative and *most* strongly evidenced. I consider each conception in turn.

Suppose we adopt the weaker conception. Then the proposed pragmatic criteria will be unable to distinguish between the possible interpretations of (1b) and (17) above. If Sean Penn attacked a photographer it may be true that the photographer got hurt, but it may also be true that Sean Penn got hurt. If Sean Penn attacked a photographer, it may be true that Sean Penn is deranged, but it may also be true that the photographer is deranged. Although in each case one of these interpretations is more likely to be true, neither can be ruled out on grounds of simple truth (or truthfulness) alone. The same point applies to evaluations based on simple informativeness and degree of evidence: both utterances would be informative and evidenced on either interpretation.

By the same token, an evaluation procedure based on simple truth (or truthfulness), informativeness and degree of evidence would be unable to choose between the two possible interpretations of (4a) above. If Mary switches from linguistics to geography, it may be true that she finds the geography lectures less boring, but it would be quite possible for her to discover, after making the change, that the linguistics lectures were less boring after all; similarly, the utterance might be informative and evidenced on either interpretation.

In other words, the proposed pragmatic criteria, understood in this way, are not powerful enough to choose between the candidate interpretations of (1b), (4a) and (17). But these were just the examples used above to support the view that reference assignment does not depend on accessibility of referents alone, but is also influenced by the need to obtain an overall interpretation that satisfies some criterion of pragmatic acceptability. A criterion as weak as the one we are now considering would be unable to distinguish among the candidate interpretations of (1b), (4a) and (17). If their pragmatic criteria are understood in this way, the whole weight of Clark and Lewis's account will be forced back onto considerations of accessibility of referents, and most of their insights will be lost.

This weak conception of the pragmatic criterion can be further undermined by comparing (4a) with (23):

(4) a. I switched from linguistics to geography. The lectures were less boring.

(23) I switched from linguistics to geography. The lectures were too boring.

In (4a), as we have seen, the correct bridge for the NP 'the lectures' is based on the NP 'geography'; in (23), it is based on the NP 'linguistics'. As we have also seen, neither choice appears to follow from simple considerations of truth (or truthfulness), informativeness or degree of evidence. To account for the interpretation of these examples, given this weak conception of the pragmatic criterion, one would therefore have to appeal to considerations of accessibility of referents. However since the first parts of these utterances are identical, it is hard to see how 'geography' could be the most accessible candidate in (4a) and 'linguistics' in (23). It seems clear that in order to understand these examples, we need a pragmatic criterion that does more work than Clark and Lewis's would on the weak interpretation we are now considering.

In looking above at how Clark and Lewis might deal with the examples (1b), (4a) and (17), I tacitly used a stronger version of their criteria. I took it that these criteria were designed to select not merely an interpretation on which the utterance is true (truthful), informative and evidenced, but the interpretation on which the utterance is *most* likely to be true (truthful), *most* informative and *most* strongly evidenced. Such a criterion, as I have shown above, is powerful enough to distinguish the various candidate interpretations of (1b), (4a) and (17). However, it too encounters a serious problem. Though it may be able to choose among alternative interpretations of a given utterance, the only way it can do so is by considering them all. Recall that what we are looking for is not just a criterion for evaluating alternative reference assignments, but one for evaluating overall interpretations in general. Evaluating overall interpretations means not just considering the two possible assignments of reference I have mentioned for examples like (4a): it means considering all possible parsings, disambiguations, assignments of reference, restorations of ellipsed material and resolutions of vagueness, and all potential sets of contextual assumptions that might have a bearing on these choices. Clearly, an attempt to take all these possibilities into account would create a combinatorial explosion of gigantic proportions, and be quite unlike anything that hearers actually do.

What seems to me one of the virtues of relevance theory is that its pragmatic criterion is explicitly designed to offer a way out of this dilemma. On the relevance-theoretic account, although the hearer may quite often have to make choices, the pragmatic criterion itself does not force him to consider all possible interpretations as a necessary part of the comprehension process. As we have seen, on a relevance-theoretic account, the hearer is justified in accepting the first accessible interpretation that is pragmatically acceptable: he has a criterion, in other words, that enables him to recognise an acceptable interpretation as soon as he sees it. This seems to be a virtue that other pragmatic theories might consider building into their accounts.

7 A relevance-theoretic solution

So what has relevance theory to say about the examples we have been discussing? Consider (4a). Intuitively, there are two ways in which the second part of this utterance can achieve relevance in a context created by the earlier part: first, as an explanation of why I switched from linguistics to geography, and second, as a description of what I found after making the switch. On the preferred interpretation (6b), where 'the lectures' refers to the geography lectures, and against an easily accessible set of contextual assumptions about what might make one switch subjects, the utterance will simultaneously achieve relevance in both these ways. By contrast, on the alternative interpretation (6a), where 'the lectures' refers to the linguistics lectures, the utterance will achieve relevance in neither of these ways. Assuming that hearers are alert to these quite typical ways of achieving relevance, they will easily access interpretation (6b), which will be relevant enough to be worth their attention, making it unnecessary to look further.

Consider (18) and (21):

(18) I switched from linguistics to geography. The lectures were just as boring.
(21) I switched to geography from linguistics. The lectures were just as boring.

Here, one of the two possible lines of interpretation mentioned above will be ruled out: it is hard to see how *either* the fact that the geography lectures were just as boring as the linguistics lectures, *or* the fact that the linguistics lectures were just as boring as the geography lectures, could be an adequate explanation of the switch from one to the other. By contrast, on the assumption that the boring lectures were the geography ones, then both (18) and (21) can achieve relevance as descriptions of what I found after making the switch. Again assuming that hearers are alert to these lines of interpretation, the intended analysis will be easily discovered and all others accordingly disconfirmed.

Now consider (4b):

(4) b. I ran from the classroom into the playground. The children were making too much noise.

Suppose the hearer decides that the second part of the utterance is intended as an explanation of the action described in the first. There are still two possible lines of interpretation: I might have run out of the classroom because the children in the classroom were making too much noise; or I might have run into the playground because the children in the playground were making too much noise. Either interpretation could yield enough effects to make the utterance worth the hearer's attention. The choice between them must therefore depend on the accessibility of contexts capable of combining with one of them to yield adequate effects for no unjustifiable effort in a way the speaker could manifestly have foreseen. In real life, assumptions about the character of the speaker would affect the accessibility of potential contexts: is she the sort of person likely to be frightened by children, is she a stickler for discipline, and so on? In a questionnaire situation, which favours the use of stereotypical assumptions, it would no doubt be easier to imagine her as a teacher running into the playground to stop the noise rather than a teacher fleeing the classroom because of the noise; hence this interpretation should be preferred.

This account, though sketchy, seems to me to give some genuine insight into the examples we have been discussing. It saves Clark and Lewis's claim that while the accessibility of candidate referents plays a role in reference assignment, this role is not decisive - a claim that, as we have seen, is not necessarily compatible with their own pragmatic criteria. It sheds light on the least-effort strategy pursued by hearers - a strategy that is rather surprising on Clark and Lewis's accounts. It maintains, as do Clark and Lewis, that what is decisive in reference assignment is the discovery of an overall interpretation which meets a criterion of pragmatic acceptability. It differs, however, as to the proposed criterion. The criterion proposed by relevance theory is one of consistency with the principle of relevance: to discover an interpretation that meets the criterion, the hearer must be able to access, not only an appropriate referent, but also a context in which the speaker might reasonably have expected her utterance, on that assignment of reference, to be optimally relevant.

Finally, this account sheds light on some stylistic questions that have been to some extent neglected in the philosophical literature on reference assignment. Here is a representative remark from Lewis:

> 'It is not good conversational practice to rely too heavily on rules of accommodation. [In a monologue where] 'the cat' denotes first Bruce then Albert, then Bruce again, what I say is to some extent confusing and hard to follow. But even if my monologue is not perfectly acceptable, its flaws are much less serious than the flaws that are averted by shifts of salience in accordance with our rule of accommodation. Confusing shifts of salience and reference are not as bad as falsity, trivial truth, or unwarranted assertion.'
> [Lewis1979/1983: 243]

That, I think, sums up the current philosophical state of the art. The criteria proposed by philosophers for evaluating alternative interpretations tend to focus on their informational content, and to set aside stylistic questions, questions of effort. If relevance theory is right, economy of effort is just as central to the evaluation process as adequacy of effect. Philosophers have recently begun to wonder about the differential acceptability of examples such as (1a) and (16) above. I would suggest that unless we begin to take seriously the role of processing effort in evaluating overall interpretations, these examples will continue to puzzle us.

References

Asher, N. & H. Wada 1988. A computational account of syntactic, semantic and discourse principles for anaphora resolution. *Journal of Semantics* 6: 309-44.

Clark, H. 1977. Bridging. In P. Johnson-Laird & P. Wason (eds) *Thinking: Readings in Cognitive Science*. CUP, Cambridge:411-20.

Clark, H. & S. Haviland 1977. Comprehension and the Given-New Contract. In R. Freedle (ed.) *Discourse Production and Comprehension*. Ablex, Norwood, N.J.: 1-40.

Erku, F. & J. Gundel 1987. The pragmatics of indirect anaphors. In J. Verschueren & M. Bertuccelli-Papi (eds) *The Pragmatic Perspective*. Benjamins, Amsterdam: 533-46.

Keenan, E.L. & B. Comrie 1977. NP accessibility and universal grammar. *Linguistic Inquiry* 8: 63-100.

Kleiber, G. 1990. *Marqueurs referentiels et processus interpretatifs. Cahiers de linguistique francaise* 11.

Kleiber, G. forthcoming. *Reference pronominale: comment analyser le pronom 'il'*. To appear in B. Munch (ed.) *Proceedings of the International Semiotics Conference on Reference and Indexicalisation.*

Lewis, D. 1970/83. General semantics. *Synthese* 22: 18-67. Reprinted in D. Lewis 1983 *Philosophical Papers*. Oxford University Press, Oxford.

Lewis, D. 1979/83. Scorekeeping in a language game. Reprinted in D. Lewis 1983 *Philosophical Papers*. OUP, Oxford.

Sperber, D. & D. Wilson 1986. *Relevance: Communication and Cognition*. Blackwell, Oxford.

Sperber, D. & D. Wilson 1987. A precis of *Relevance*; Presumptions of relevance. *Behavioral and Brain Sciences* 10: 697-754.