

## Charity and context

Pascal Engel  
Université de Caen  
and CREA, Ecole Polytechnique, Paris

*colloque Interpretation and context, St Andrews 1994, inédit*

«Though truth and falsehood bee  
Neare twins, yet truth a little older is.»

John Donne, Satyre II

### I

Much of the contribution of contemporary philosophy to the problem of interpretation in anthropology is not a contribution to anthropology, but to what might be called *a priori* anthropology (Cooper 1986). This is an armchair speculation, far from the field, about the general requirements of the interpretation of beliefs and other intentional states of creatures understood in a very abstract way, outside their social and psychological context. The philosopher's motto that interpretation (or translation) "begins at home" and that the problems of interpretation are the same for people inside our culture and for people outside this culture serves here as a safeguard against all temptations to tie the ascription of psychological contents to people to the various circumstances of their social life. Thus interpretation, in the philosopher's sense, is from the outset divorced from context, in the sense of the word "context" which presumably interests the anthropologist, especially in this conference. So why bother consider the speculations of philosophers in this domain? Because although interpretation is always interpretation "within" and "relative to" a context, interpretation *is* interpretation nevertheless, and although the claim that interpretation is relative to a context seems to be almost an analytical truth, or a truism, we still do not know exactly what it means. The present paper falls squarely, I am afraid, within the range of a priori anthropology, but I hope that it will be able to raise some of the issues about the context sensitivity of interpretation that might interest anthropologists.

My topic concerns the range of a principle which has been held to be a necessary requirement of the interpretation of psychological contents: the so-

called "principle of charity" (henceforth "PC") , according to which, in one of its most general formulations, other people's beliefs must be, for most of them, presumed true (this is the veridicity version of the principle) and rational by our lights (this the coherence version, and both versions are in general held together— e.g. by Davidson), and thus be charitably understood, as much as possible. The relevance of the principle for theoretical issues in anthropology is generally considered to be the following. If charity is a necessary condition of interpretation, we should from the outset ban the ascription of false or incoherent beliefs to individuals. Thus the PC set limits to (and maybe prohibits) what seems to be one of the most obvious facts of interpretation and understanding in most contexts, especially in the antropological context of understanding people's beliefs within a culture: the natural experience of being unable to understand or to interpret, and the experience of being confronted to beliefs which are obviously false and incoherent. If the PC is correct as a necessary requirement of interpretation, there are limits to misunderstanding and misinterpretation, as well as limits to the claim that other people's beliefs can be different from our own. This is why the PC has been most often invoked by writers who want to reject a form of relativism ("beliefs which are false or incoherent by *our* criteria might be true or coherent by *their* criteria") or a form of hyper-rationalism (" their beliefs display no rationality at all, only *ours* do"). If we were to discover false or incoherent beliefs in other creatures, and if we were tempted to ascribe to them a "prelogical mentality" we should rather question our translation manuals (Quine). Thus the PC seems to lend support to a form of "neo-rationalism" (Cooper 1986) according to which all understanding presupposes unity. I shall not, however, been concerned directly with these implications of the principle of charity for the rationalist/relativism issue. My main concern will be the use of the PC as a maxim of interpretation and with the difficulties that it raises.

I shall proceed as follows. First I shall consider two current criticisms of the PC: first that it has the effect of treating people a much more rational and reliable in their beliefs as they actually are, second that it is too abstract an unspecific do deal with actual interpretations, so that we do not know how to apply it in a given context. In other words it seems that our ordinary and everyday practices of interpretation of beliefs and other intentional states do not in fact conform to the PC. These criticisms, in the contemporary philosophical and psychological literature, are generally made by theorists who claim that everyday interpretation is more the effect of a simulation or of a modelling of other people's mental states than the effect of the imposition of a strong form of rationality. I shall try to point out that the charitable method and the simulation method are not really alternatives, and that there is no obvious ground to separate them. But the fact that we can use either a rationality or a simulation method does not by itself solve the context problem, that is the problem of knowing how such a method can be applied in particular circumstances. In a sense, I shall argue, there is no general solution to this problem. Nevertheless, it does not follow that no rationality rule can be used in actual interpretation. It can, and indeed must be

used, but not globally or in an extensive way; it must be used locally, according to a case by case approach, relative to a particular speaker in a particular circumstance. If such an approach is adopted, the use of the PC is compatible with the attribution of error or of incoherent beliefs. What is true of particular beliefs is also true of our norms of rationality. In some cases, we tend to attribute errors when there are no errors, and in these cases we have better change our normative principles. So my defense of the PC will be only a partial and qualified one.

## II

There are two main criticisms of the PC in contemporary literature. The first is that in so far as it maximizes truth and coherence in the beliefs and other intentional mental states of individuals, it tends to treat them as perfectly veridical and perfectly rational. But humans are not perfectly veridical nor rational agents; they make mistakes, hold many incoherent beliefs, and indeed as many psychologists tell us they tend do systematic mistakes. How could an overall rationalizing principle such as the PC account for these obvious facts? It is not only unrealistic to consider the PC a an actual rule of ordinary interpretation, but also to propose it as a normative rule of understanding.( See for instance Nisbett and Thagard 1983, Cherniak 1986, Goldman 1979, Stich 1990).

The second criticism is that the PC, even if it were correct, is too abstract, too context independent. We do not know how to apply it in a given context. This is presumably to rule out such a criticism and the preceding one that proponents of the PC make the qualifications mentionned above: try to maximize truth and rationality" for the most part", "most beliefs", or "as much as possible". This is meant to imply that *some* beliefs and other states can be false or irrational, in spite of their presumed truth and rationality. Hence the common formulation of the PC (Gauker 1986):

- (1) *Ceteris paribus* we ought to attribute beliefs we regard as true and rational

The problem is: how are we to assess the *ceteris paribus* condition? (1) means that *in some contexts* the principle might be withdrawn. But which contexts? And how could we distinguish contexts in which the principle applies from contexts in which the principle does not apply? How are we to distinguish the cases where the beliefs are true (rational) from the cases where they are not? As a handy example, take the case ( Stjernberg 1991) of a Japanese who starts asking you (an Occidental) something by saying: "Please, Sir, execute me". The charitable interpretation of this is that people in general do not want to be executed by their interlocutors, and that in this particular case the Japanese was probably meaning to be excused rather than executed. The charitable ascription is reached because the speaker's error can be easily explained by the proximity of the two words. But

the hearer might also here take the Japanese's saying at face value, as literally meaning that he wants to be executed, in the context of the hearer's knowledge of the Japanese custom of Hara-Kiri. How is he to choose between the two interpretations? When will he withdraw his charitable ascription?

The two criticisms — unrealism and abstractness— are in fact closely related: it is precisely because the PC is unfit to actual interpretative practice that it does not allow us to isolate a particular mistake or incoherence in a particular circumstance.

The main proponent of interpretative charity, Davidson, answers this kind of criticism by saying that the aim of the PC is not to "maximise" truth and coherence, but to "minimise unexplained error". The idea is that when we can explain errors and mistakes, we can withdraw our overall initial charitable assumptions. As Davidson says, the aim of interpretation is not agreement but understanding (*preface* to Davidson 1984). Davidson is able to maintain the general validity of the PC in spite of the fact that actual interpretation does not really fit its prescriptions, because for him the PC is not simply a methodological maxim of interpretation, "a piece of friendly advice" that one would be free to adopt or not. It is rather a necessary constraint on all possible ascription of beliefs (a "normative" and "constitutive" principle, as one says in Oxford): if a belief is not susceptible to be held true or to be rational, it is not a belief at all; no charity no beliefs, no rationality no thought. Davidson grants that particular applications of the PC will be made in the light of the particular psychological explanations that we have of the individuals we are dealing with. But he leaves the nature of these explanations completely undetermined. On this "normative reading" of the principle, the particular applications of the PC in a given context do not have to be of any concern for us. A normative principle has to be applied in a given context. But is not the job of the principle as such to specify its conditions of applicability.

Again this won't solve the general difficulty that is pointed out by the two criticisms that I have mentioned, since Davidson does not tell us which errors are "explicable" and which are not, and by what criteria. Thus he justifies such criticisms as John Skorupski's, who writes (talking about translation schemes, but what he says can be extended to psychological interpretation in general):

"The degree of probability attaching to the claim that certain incoherent beliefs are held in culture, and thus the degree of probability attaching to the translation scheme which supports the claim, cannot be assessed independently of the psychological and sociological story we may be able to tell as to why people in that culture hold such beliefs, the comparative familiarity we may have with beliefs of the kind in question from our knowledge of other cultures and so forth. The plausibility of the translation scheme is determined in the context of our general sociological and psychological knowledge, and in conjunction with the particular psychological account we may be able to give of the culture concerned." (Skorupski 1978, 104-105)

Here Skorupski is not repeating the Davidsonian point that interpretative charity should be relativised to a given context of behavioural explanation when we have

such an explanation. In fact he is reversing this very point: we can be charitable about other people's beliefs only to the extent that we can explain these beliefs. This is destructive of the PC itself as a principle of interpretation. Skorupski says that "the trouble, with [ the PC] is that, taken as a principle, it is simply wrong." It is wrong because, as Skorupski says quoting Gellner, " it may delude anthropologists into thinking that they have found that no society holds absurd or self-contradictory beliefs." The adjudication [between overall accounts of the meaning of other cultures beliefs] must turn on a different, more complex question: can what people in a given culture mean, on a given theory of meaning/principle of interpretation, by what they say, be linked with a plausible psycho-sociological theory which tells us why they should be given to saying what — so understood — they do say?" (*ibid.*)

On the face of it, such complaints are wholly justified. They call for an understanding of our everyday understanding — and *a fortiori* of our understanding of another culture's beliefs — which would be much more psychologically and sociological than the purely rational method that it involved in the use of interpretative charity. I believe nevertheless that the kind of worry that Skorupski and others expresses can be (partly) resisted.

### III

The most current attempts to deal with the difficulties just adduced consists in trying to analyse our interpretative practices in the light of what appears to be a more plausible theory of these practices, namely one which would be psychologically informed. There are two main views along these lines. Both emphasize that the actual practice of everyday interpretation does not consist in imposing a general and extensive rationality to the beliefs of the individuals interpreted but in trying to model their belief contents by using our own beliefs. In other words, interpretation is, according to this view a form of *simulation* or of projection of the beliefs and other psychological states of the interpreter onto the interpretee. But the two views differ on the nature of the simulation process.

Let us consider the first view, which I shall call the *normative simulation view*. It rests on Grandy's (1973) proposal that we should replace the PC with a "principle of humanity" according to which one must attribute to others the beliefs or inferences which we would have if we were in their place. What this view recommends is that we take the interpreter's actual dispositions and behaviors as providing the standards of rationality. The purpose of this method is to account for the fact that the interpreter can attribute intentional states which fall short of his own standards of rationality. In attributing the intentional states that he would have if he were in a given situation, the interpreter treats others as being as much rational as himself as possible, by his own standards. If the interpreter takes himself as perfectly rational, then the humanity principle of simulation amounts to the PC in the original form. But if, more likely, the

interpreter takes himself as less than optimally rational, the principle of simulation enjoins the interpreter to take others as rational *to a certain degree*, the degree which would be the closest to the degree to which he takes himself to be rational, by his own standards. Sometimes Davidson himself advocates such an idea (e.g Davidson 1984, *preface*). Is the principle of simulation in this sense a genuine alternative to the PC? I do not think so. Remember that the main objection to the PC is that it seems to preclude the possibility that someone could be found *wildly* or massively irrational or mistaken. But it is unclear that the same objection cannot be raised against the principle of simulation. For in order for it not to preclude the possibility of wild irrationality, it would have to be allowed that the interpreter takes himself to be wildly irrational, that is to depart wildly from his own standards. This might happen sometimes: there are cases where we can reflect consciously that we are incoherent. But it is very unlikely that we could adopt this as a general rule or policy of our interpretations. Notice also that the application of the principle of simulation to one's own case, in self-interpretation is vacuous, since it amounts to requiring that the interpreter interprets himself as as rational as himself. Does the principle of simulation fare better than the PC when we have to make room for intelligible error or irrationality? In principle, the simulation method is supposed to allow the interpreter to attribute a false belief or a defective inference whenever (s)he can see that the individual (s)he is interpreting is in a situation where (s)he would be himself liable to make the same kind of mistake. But this does not solve more than the PC the context problem, because there is no reason to think that the interpreter knows when (s)he is liable to be mistaken. If he knew that, (s)he would be much more rational than he in fact takes himself to be, which defeats the very purpose of the humanity or simulation method. We can't presume nor expect the kinds of circumstances when we can be in error. So the simulation method, or any method of sub-optimisation of rationality is not a genuine alternative to the PC.

It does seem, however, to be a genuine alternative when we consider the second view about the simulation method. On this view, the simulation of other people's mental states is not a normative rule that we adopt, but a psychological process. This is why we can call this view the *psychological simulation view*. According to this view, simulation is a basic process of the mind, presumably entrenched in humans in infancy, and possibly innate, whereby we have a general tendency to project our own beliefs onto the beliefs of others. Most proponents of this view, such as Gordon (1986) or the psychologist Paul Harris (1992) claim that this tendency is tied to the emotions. It might even be a module in the mind. So there is nothing "normative" in it. It's a purely factual matter that we do simulate in an unreflective way. This is why, in particular, one proponent of this view, Goldman, insists in saying that interpretation has nothing to do with any principle or norm of rationality, and that interpretation should be "psychologized" (Goldman 1989). But it is not evident that the simulation process, understood in this way, is a genuine alternative to another psychological account of interpretation. According to many psychologists, children's early capacity of

ascribing beliefs and other mental states to other is the product of a "theory of mind". It is often quite unclear what such a "theory" is supposed to be: is it a set of laws of behaviour, a capacity to represent beliefs *about* beliefs in the sense of metarepresentations, or some other conceptual capacity? But whatever the "theory of mind" can be, it is a certain form of *knowledge*, innate or not. And it is quite unclear, in the present state of psychological research, that most of the results about the so-called "Sally Ann test" inaugurated by Wimmer and Perner (1984) (the test whereby one tests the capacity of children for forming false beliefs) really can be accounted in terms of the effects of a simulation-projection capacity instead than in terms of the effects of the knowledge of a certain form of theory. (Stich and Nichols 1992, Perner 1992). The basic point, in a nutshell, was given by Dennett (1987) against the simulation methodology: what can be accounted in terms of a method of psychological simulation can also be accounted in terms of a method of rationalisation, and putting yourself in another person's shoes does seem to be the same as trying to find out what it would be most rational, given your knowledge of what people usually do and think, for them to do and think.

Now this does not solve the problem about the normative status of the simulation method or of the so-called "theory of mind". It depends on what you take norms to be. It is plausible to hold that a norm of interpretation arises out of a conscious reflection on an actual practice or disposition, in the sense of what Peirce called a "leading principle", whereby an habit is turned into a rule. But if there is no genuine rivalry between an account of the child disposition to interpret in terms of a simulation process and an account of this disposition in terms of the possession of a "theory of mind", there is also no genuine rivalry between the different account of norms of interpretation that arise out these psychological capacity.

So it is not obvious than we can solve the problems posed by the rationalisation method in interpretation by switching to an account of interpretation in terms of a principle of humanity or a simulation process.

#### IV

Is there a way of maintaining the use of a principle of charity in interpretation without presupposing a general rationality of the people interpreted and at the same time accounting for how the PC can be applied in particular circumstances? A plausible suggestion is the following. Take the usual, well-known analysis of the context of radical interpretation given by Davidson, whereby one tries both to account for the meaning of a person utterances and for the belief expressed in these utterances (interpretation is "radical" in this sense that we can't presuppose either and must account for both). On this account the

radical interpreter has to impose a semantic theory upon the language than he interprets which yields theorems of the form:

(T) S is true in L iff p

Claims of the form (T) can be tested by deriving a corresponding observational consequence (O) via an auxiliary assumption (C) (Warmbrod 1991):

(C) X holds true S iff S is true in L

(O) P holds true S iff p

Suppose for instance that in a circumstance where a German speaker, Kurt, utters the sentence "Es regnet", the interpreter observes that it rains. The interpreter ascribes to him that belief that it rains, and ascribes this meaning to his sentence. In other cases, the interpreter can in fact confirm a hypothesis that he makes that "Es schneit" is true iff it rains. He uses the general principle of charity to ascribe to Kurt this truthful insight. According to Davidson interpretation in this sense is a form of hypothetico-deductive method, whereby one tests claims of the required (T) form against independent evidence. . But this is much too simple. Kurt could have said "Es regnet" as a lucky guess, without actually believing that it is raining, nor without actually have meant that it rained by uttering these words. One has also to assume semantic uniformity in the language, the fact that Kurt is a reliable speaker of German, and the fact that Kurt is a reliable judge of whether it is raining. If these assumptions are made, Davidson also holds (1984, *preface*) that claims of the form (T) should also take the form of subjunctive conditionals: in similar circumstances, Kurt *would* hold true that it rains. The problem we are dealing with is: why should we make these assumptions? The reason why the PC is so shocking is that it implies, as we said, that *most* of a person's beliefs are true. One reason why Davidson thinks that such ascriptions are reliable in many cases is that they tend to *cohere* together. A belief is reliable if it is consistent with other beliefs. So he seems to be committed both to some form of coherence criterion for knowledge and indeed to the doctrine of semantic holism. In fact we could view the source of Davidson's well-known holism about meaning in his attempt to secure the use of the charitable method (or alternatively his defense of the PC as springing out from his semantic holism). But we do not need to maintain semantic holism nor the extensive charitable method to secure interpretation in particular cases. We only need to assume that a (subjunctive) instance of (C) is true if the semantic theory plus independent information implies that the person interpreted is a reliable judge of whether the sentence is true or false. (Warmbrod 1991, 445). If we do that, there is no need to use a sweeping assumption of general charity. We can be charitable case by case, on local ascriptions. But it does not seem to solve the problem of interpretation. For by what criteria are we going to judge that Kurt, in such and such a context, is a reliable informant about whether it rains? What is reliability in general and what

is reliability in a particular case? Why should we not instead say that in such and such a case Kurt is a reliably judging that water vapor is condensing on dust particles in the atmosphere and falling to the ground? Or that Kurt is judging that the rain god is weeping? The charitable method does not give us any insight into the doxastic and conceptual apparatus that Kurt possesses. Here again the Davidsonian will presumably use holistic considerations. For instance if Kurt asserts that "water falls from the sky when the rain god is weeping", the interpreter will have a reason to ascribe to him the belief that the rain god is weeping when he utters "Es regnet". There is nothing problematic with such holism if it can be maintained on a local basis. For instance in his recent book on *Concepts* (1992, MIT) Christopher Peacocke disclaims a general holistic view about concepts, against the Davidsonian interpretative method, but he holds that it does not prevent concepts from forming "local holisms". But when is holism in concepts and beliefs local, instead of global? Where are we to draw the line? I take it that it is this worry that most of the critics of semantic and psychological holism want to express.

The problem we are dealing with is the problem of distinguishing local consistency and local inconsistency in a belief set from global consistency and global inconsistency. I can only agree here with John Skorupski that any distinction of this kind must be made on the basis of empirical grounds, on the basis of the psychological and sociological story that we can give of why people hold such or such beliefs, and not on a purely *a priori* basis. But it should be noted that even when we want to produce claims about the consistency or the rationality of human subjects on the basis of a psychological theory, we should care not only about the psychological or sociological evidence that we have, but also about the kinds of normative standards that we have at our disposal. This can be illustrated with the famous set of experiments which have led psychologists such as Kahneman and Tversky (1982) to claim that people are not good at estimating the probability of single events, and are not in general good bayesians. Here the psychologists are careful not to conclude from these experiments that people are irrational in their probability judgments, but that they make systematic errors through the biases and heuristics that they use to solve such problems. One of the most famous of Kahneman and Tverky's experiments is the Linda case, which illustrates the so-called "conjunction fallacy" that people are prone to make in their probabilistic estimations.

Linda is a bank teller, single, outspoken and bright. She majored in philosophy, and as a student was deeply concerned in feminist issues in discrimination and social justice, and also participated in antinuclear demonstrations. which is the more probable?

(T) Linda is a bank teller

(T&F) Linda is a bank teller and is active in the feminist movement.

Most subjects answer T & F. In this case a basic principle of probability theory is alleged to be violated, the principle that the probability of a conjunction can never be greater than the probability of one of its conjuncts. But choosing the conjunction is *not* a violation of probability theory if, instead of the single event bayesian version of probability theory, one uses instead the frequency version. When the problem is reformulated in terms of frequencies such as

- There are 100 persons which fits the description above (Linda). How many of them are:
- (a) bank tellers
  - (b) bank tellers active in the feminist movement

the rate of good answers increases, and the "fallacy disappears". (Gigerenzer 1991) As the statistician Barnard (1979) said jokingly (quoted by Gigerenzer), such judgments as those which are supposed to illustrate the fallacy should be treated in the context of psychoanalysis, or indeed, of course, not jokingly, of cognitive or social psychology. (If we remember that the bayesian principles of inference are taken by Davidson to be general principles of rationality in the interpretation of action, we should also doubt that his method of interpretation in this case could locate "explainable errors".)

This shows, it seems to me, that we should care not only for the particular empirical, psychological, sociological, anthropological evidence that we have for interpreting people, but also for the normative standards that we have when we collect such evidence. In this sense, a general methodological rule of charity cannot simply rely on what are taken to be *the* normative standards, but on what we should be prepared to take as *a* normative standard, keeping in mind that there are alternative normative standards. The question then, that we face, is the question of how we are going to evaluate our normative standards, in order to find which ones can be the most general, and which ones we should revise partly in the light of empirical evidence. This is much more general and difficult question. The fact that we must allow that there are a variety of normative standards does not mean that they are irreducibly many and that they have to be estimated on a purely pragmatic basis, as Stich' relativism (1990) implies. On the contrary we can attain what is currently called in moral theory (Rawls) a "reflective equilibrium" (Thagard 1988). I would claim that the same applies to psychology, sociology and anthropology. We should not conclude that interpretation is a purely empirical matter, and we should not conclude either that it is a purely normative matter, based on sweeping assumptions of rationality. But with such a sweeping claim I must end here.

## References

- Cooper, D. (1986) "Anthropology and Translation", *Proceedings of the Aristotelian Society*, LXXXVI, 1985-1986, 51-68
- Davidson, D. (1984) *Inquiries into Truth and Interpretation*, Oxford, Oxford University Press.
- Gauker, C. (1986) "The Principle of Charity", *Synthese*, 69, 1-26
- Gigerenzer, G. 1991 "How to make cognitive illusions disappear: Beyond Heuristics and biases", *European Journal of Social Psychology*, vol.2, 84-115
- Goldman, A. (1989) "Interpretation Psychologized", *Mind and Language*
- Gordon, R. (1986) *The Structure of the emotions*, Cambridge University Press
- Grandy, R. (1973) "Reference, Meaning and Belief", *Journal of Philosophy*, 70, 439-452
- Harris, P. (1992) "From Simulation to Folk Psychology": the Case for Development", *Mind and Language*, 7, 1-2
- Kahneman, D, Slovic, P. and Tversky, A. (1982) eds, *Judgment under Uncertainty, Heuristics and biases*, Cambridge, Cambridge University Press
- Nisbett, R. and Thagard, P. "Rationality and charity", *Philosophy of Science* 50, 250-267
- Skorupski, J. (1978) "The Meaning of Another Culture's Beliefs", in Hookway, C. and Pettit, P. *Action and Interpretation*, Cambridge, Cambridge University Press
- Stich, S. 1990 *The Fragmentation of Reason*, MIT Press, Cambridge Mass.
- Stich S. and Nichols, S. (1992) "Folk Psychology: Simulation or Tacit Theory?", *Mind and Language*, 7, 1-2
- Stjernberg, F. (1991) *The Public Nature of Meaning*, Acta Universitatis Stockholmiensis, Stockholm Studies in Philosophy 10, Almquist and Wiskell Int.
- Thagard, P. (1988) *Computational Philosophy of Science*, MIT Press, Cambridge, Mass.
- Warmbrod, K. "The Need for Charity in Semantics", *Philosophical Review*, C, 3, 515, 431-458