

CHARITY AND ERROR-THEORETIC NOMINALISM

Arvid Båve

Abstract

I here investigate whether there is any version of the principle of charity both strong enough to conflict with an error-theoretic version of nominalism about abstract objects (EN), and supported by the considerations adduced in favour of interpretive charity in the literature. I argue that in order to be strong enough, the principle, which I call “(Charity)”, would have to read, “For all expressions *e*, an acceptable interpretation must make true a sufficiently high ratio of accepted sentences containing *e*”. I next consider arguments based on (i) Davidson’s intuitive cases for interpretive charity, (ii) the reliability of perceptual beliefs, and (iii) the reliability of “non-abstractive inference modes”, and conclude that none support (Charity). I then propose a diagnosis of the view that there must be some universal principle of charity ruling out (EN). Finally, I present a reason to think (Charity) is false, namely, that it seems to exclude the possibility of such disagreements as that between nominalists and realists.

Nominalists about abstract objects have given remarkably disparate answers to the question: how, if there are no abstract objects, should we view statements that seem to involve purported reference to or quantification over such objects?¹ According to the

¹ I define nominalism as the view not merely that there are no abstract objects, but also that *nothing is an abstract object*, contra certain “noneists” like Graham Priest (2005), who accept the

error-theoretic variant of nominalism (EN), these statements should be taken at face value, both semantically and pragmatically, i.e., taken both as literally entailing that there are abstract objects, and also as used and interpreted literally by ordinary speakers. On this view, most of our everyday claims involving abstract singular terms are untrue due to reference failure, and existential quantifications over abstract objects are simply false.

Among the difficult questions that arise for such a theory are:

- (1) What explains why we use the relevant vocabulary, given that it is fraught with such massive error?
- (2) Closely related to the first question, does the discourse in question have a function, in spite of being fraught with massive error, and, if so, what is it?
- (3) What constitutes the meanings of the relevant abstract terms, given that they cannot be constituted by the expressions' having certain extensions, referents, etc.?
- (4) Can the account of the function of the relevant expressions be squared with the account of their meanings?

first claim, but reject the second (see, e.g., his *Towards Non-Being: The Logic And Metaphysics of Intentionality* (Oxford: Oxford University Press, 2005).

- (5) What to say about Quine's and Putnam's plausible assumptions, (i) that we should accept whatever ontological commitments are incurred by our best theories, and (ii) that the best theories will involve quantification over abstract objects?
- (6) What should adherents of (EN) say about the plausible-seeming principle that in assigning meanings to natural language expressions, we should make speakers come out as mostly speaking the truth?

In my 'Deflationist Error-Theory of Properties, forthcoming in *Dialectica*, I propose detailed answers to each of the questions (1)-(5). The present paper is devoted to the sixth question, which I think can be handled without any extended discussion of the others.

Many philosophers hold that the imputation of massive error is reason enough to reject views like (EN). This form of rejection may, in turn, have several sources, roughly corresponding to the six questions above. A major source is the so-called Principle of Charity, pioneered by W. V. O. Quine and Donald Davidson, according to which we should interpret people so as to minimise error (or irrationality) on their behalf.² Crispin

² See esp. W. V. O. Quine, *Word And Object* (Cambridge, MA.: MIT Press, 1960), §13, and Donald Davidson, 'Mental Events', in L. Foster and J. Swanson (eds.), *Experience and Theory* (London: Duckworth, 1970), pp. 79-101, at p. 97), and *Inquiries into Truth And Interpretation* (Oxford: Clarendon Press, 1984), essays 9, 10, 11, 13, 14, Appendix, 'Reply to Burge', *Journal of Philosophy* 85 (1988), pp. 664-665, 'Interpretation: Hard in Theory, Easy in Practice', in M. de Caro (ed.), *Interpretations and Causes: New Perspectives on Donald Davidson's Philosophy* (Dordrecht: Kluwer, 1999).

Wright, for instance, argues that ‘charitable interpretation dictates that we should avoid that charge [of massive error] if we possibly can’ (‘Response to Jackson’, *Philosophical Books* 35 (1994), pp. 169-175, at p. 172).³

Various considerations have been adduced in the literature in favour of “charitable interpretation”. This paper, however, argues that none of these considerations suffice for establishing a principle strong enough to conflict with (EN) (given the facts about what sentences we accept, of course, but I will henceforth leave this implicit). In Section 1, I consider a number of different candidate principles of charity and argue that there is one principle, called simply “(Charity)”, which is *prima facie* reasonable yet strong enough to rule out (EN), and which will be the focus of this paper. I then turn to the arguments for principles of charity from the literature, to wit, Davidson’s intuitions (Section 2), the reliability of perceptual beliefs (Section 3), and the reliability of “non-abstractive inference modes” (Section 4), and argue that none of the relevant arguments succeed in supporting (Charity). In the concluding remarks of Section 5, I first propose a diagnosis of the appeal of a universal principle of charity strong enough to rule out (EN);

³ Other examples of over-the-board arguments against error-theories based on considerations of charitable interpretation include David Wong, *Moral Relativity* (Berkeley: University of California Press, 1984), S. L. Hurley, *Natural Reasons: Personality and Polity* (Oxford: Oxford University Press, 1990), p. 93, Gilbert Harman and Judith Thomson, *Moral Relativism and Moral Objectivity* (Cambridge, MA.: Blackwell, 1996), p. 4, James Dreier, ‘Moral Relativism and Moral Nihilism’, in *Handbook of Ethical Theory*, ed. David Copp (New York: Oxford University Press, 2005), pp. 240-264, at pp. 261f., and Stephen Finlay, ‘The Error in the Error Theory’, *Australasian Journal of Philosophy* 86 (2008), pp. 347-369, at pp. 366f.

then I suggest a weaker and more plausible universal principle, and, finally, I provide a simple argument showing that (Charity) is (most probably) false.

1. A spectrum of interpretive principles

Our question is whether there is a principle of charity that is both strong enough to rule out (EN) (given, as always, which sentences we accept) and supported by argument. The “Principle of Charity” is typically taken to say, very roughly, that an interpretation is good to the extent that it makes the interpretee come out as being right, or as agreeing with the interpreter, or as being rational, or as making sense.⁴

Firstly, what is meant by an “interpretation”? I will take it to be an assignment of truth-conditions (to sentences), satisfaction conditions (to predicates), and reference conditions (to singular terms). Although it is perhaps more customary to take interpretations to involve assignments of *referents* to singular terms, this would clearly be inadequate to our purposes. Taking them to involve merely reference *conditions* seems neutral in the required way, and is arguably plausible for independent reasons (cf. Mark Sainsbury, *Reference without Referents* (Oxford: Clarendon Press, 2005)).

Now, what kind of property of sentences is it that should be maximised, according to our principle of charity, if it is to conflict with (EN)? It seems that it must be

⁴ We should also keep in mind David Lewis’s view that the relevant principle concerns how physical facts metaphysically determine mental and semantic facts, rather than how knowledge of the former epistemically justifies beliefs about the latter (see his ‘Radical Interpretation’, *Synthese* 23 (1974), pp. 331-344). I will formulate the principles in epistemic terms—as Lewis in fact does himself—but this will not affect the discussion.

truth. A principle involving “making sense” seems too loose and ambiguous, but also too weak to rule out (EN) (since, at least on one natural reading of “make sense”, we would make sense even if we were in massive error concerning abstracta). The rationality-maximising principle does not seem to be of the right kind for ruling out (EN) either. The latter is an *error*-theory, after all, not one imputing irrationality. And the principle asking us to maximise agreement in effect coincides with the truth-maximising one in its verdicts as to which interpretation is best, for judging something as true is merely to judge it as true-by-one’s-lights.⁵ (I will speak more or less interchangeably of accepting sentences and believing, and of sentence-meanings and belief-contents.)

Davidson, the father of the principle (the grandfather being Quine), often says that we can allow for *explicable* error, and thus that charity is about minimising inexplicable error. Should we similarly qualify our desired principle so as to demand maximisation of sentences that are true, except those the acceptance of which we can explain? I think not, for (EN) is plausibly committed anyway (by independently plausible principles) to there being an explanation, consistent with (EN), of why we accept sentences entailing the existence of abstracta although they are untrue. If there is such an explanation, then this qualification would make the principle accord with (EN). If not, then (EN) fails anyway, for reasons independent of interpretive charity.

It is clear that truth-maximisation cannot be the only constraint guiding interpretation. If it were, we could just take every sentence to mean that $1=1$. In fact, even

⁵ True, they do not seem to coincide on Lewis’s metaphysical take on the Principle of Charity. On that take, maximisation of agreement with “the interpreter” does not quite make sense. Surprisingly, Lewis fails to notice this oddity—see esp. his ‘Radical Interpretation’, pp. 336f.

if we add both a principle of simplicity and one of learnability (which in turn motivates a principle of compositionality), this would not be enough to rule out the above interpretation (given that it takes all sentences to be atomic and that the speaker(s) follow a general rule to the effect that every sentence is true iff $1=1$). In view of this, it may seem that we also need to make the interpretation consistent with some assumption to the effect that language use is *useful*. Interesting and important as these questions may be, we will here forego them, by simply using a proxy, “constraints *C*”, to cover whatever further constraints should be set on interpretations.

Keeping these points in mind, we might now propose the following principle:

- (C) The correct interpretation of a speaker’s language is that which strikes the best balance between truth-maximisation and constraints *C*.

This is perhaps not aptly called *simply* a principle of charity, since it also sets other constraints, although it does not spell them out. However, this is not of importance to the first point I want to make, which only concerns the charity-part of (C).

(C) is not apt to figure in an argument against (EN), since a nominalist can hold that, given her views about what exists, the best balance between the constraints is one on which most abstract beliefs come out untrue. Thus, regardless of what the other constraints might be, the truth-maximising constraint can be well satisfied by a nominalist theory since it merely asks us to make as many accepted sentences as possible come out true, given the facts. According to the nominalist, there are no abstracta, and given this fact, one can only make so many accepted sentences come out true. That ratio

of true accepted sentences containing abstract terms may be low, but what matters for (C) is merely that it is the highest one possible (given sufficient satisfaction of the other constraints), not that it is high.

For this reason, we need to find a principle that sets a lowest acceptable limit on the ratio of sentences rendered untrue by the interpretation. Thus, on the desired kind of principle, if one cannot state an interpretation, which, given one's ontology, reaches this limit, then one has to change one's ontology.

A very weak variant of the kind of principle I have in mind is:

(W) An acceptable interpretation must make a ratio r of accepted sentences come out true.

Clearly, (W) does not merely ask us to interpret speakers so as to come out as accepting true sentences as far as possible, given our pre-existing views about the world, since it sets a definite ratio of true accepted sentences. We need not, of course, decide the value of this ratio, but simply assume that it is high enough for (EN) to be ruled out by (W), given what sentences we accept. We will soon see that (W) is not adequate to our purposes either, but since it shares certain important features with the principle I will eventually opt for, I would first like to comment on them.

Firstly, the reference to "constraints C " is missing in (W) (and will not crop up in the principles to come either). The following line of thought should show why they need not be mentioned, given the reading of (W): assume that (EN) satisfies the other constraints to degree d . Whatever d may be, if the argument from charity against (EN) is

to work, there must be a lowest ratio (given that value of d) of true accepted sentences consistent with the assumption that the interpretation is acceptable. Just let r be that ratio, and (W) will be of the right form to conflict in the appropriate way with (EN).

Perhaps it is easier to see this point by noting that even if (EN) scores *maximally* with respect to the other constraints, this should not be sufficient for making (EN) an acceptable interpretation, if there indeed is a sound argument against (EN) based on some principle of charity. Thus, we can in what follows assume that (EN) does indeed score maximally on that count, and hence take r to be some ratio high enough for (EN) to violate (W). (Note, by the way, that it is not enough that $r > 0$, since (EN) takes some accepted abstract sentences to be true, e.g., “There is no greatest number”).⁶

Now, despite being of the right form in the respects we have noted, (W) fails to conflict with (EN) for the simple reason that one can, for any ratio less than 1, make that ratio of sentences come out true without making that ratio of sentences *within a certain discourse* come out true (cf. Chris Daly and David Liggins, ‘In Defence of Error Theory’, *Philosophical Studies* 149 (2010), pp. 209-230, at p. 212). Thus, one could propose an interpretation which makes true fewer than r *abstract* sentences that speakers accept, but which makes true more than r accepted sentences *tout court*.

⁶ Another caveat I should mention, if only to put it out of our way, is that a more plausible principle of interpretation would not merely set a ratio of accepted sentences that should come out true, but also take into account the *degree of firmness* to which sentences are accepted, and probably more. Nothing below will turn on such niceties, however, so in what follows, I will stick to simple formulations like (W).

The right kind of principle must take this point into account. However, the talk about “discourses” is too vague. How do we decide which “discourses” there are and which sentences belong to which? But there is a clear formulation with which I think we can make essentially the same point, namely by universally quantifying over expressions. Such, indeed, will be the formulation of the principle I wish to focus on in this paper:

(Charity) For each expression e , an acceptable interpretation must make a sufficiently high ratio of accepted sentences containing e come out true.

Assuming the “sufficiently high ratio” is high enough, (Charity) is guaranteed to conflict with (EN), given what sentences we accept. Clearly, it is not possible to make (Charity) come out consistent with (EN) by “compensation elsewhere”, as was possible with (W). The question that will occupy us in sections 4-6 is whether the considerations that have been taken to support the idea of interpretive charity can be used to motivate (Charity) or some restricted version of it that is still strong enough to conflict with (EN). Of course, there are weaker principles conflicting with (EN), e.g., the variant of (Charity) had by restricting the expressions to abstract expressions, but since the idea is that there is something in general about interpretation that makes (EN) implausible, I will stick to (Charity).

Before embarking on the main project of the paper, I would like to note a general problem with the idea of refuting (EN) by appeal to a principle of charity. If I am right that (Charity) is the right principle for this project, the project itself may seem doubtful in

view of the fact that (Charity) seems too strong to be true. To wit, it seems to have the consequence that speakers cannot use empty singular terms they falsely believe to refer, e.g., “Thor” and “Odin”. And, surely, the ratio of true sentences containing “Odin” that the Vikings accepted will not be high enough. The most obvious way to avoid this consequence would be to restrict the quantification in (Charity) to terms with a referent. But then, the principle would clearly fail to conflict with (EN). So how could one exclude those expressions in a principled and plausible way without also excluding abstract singular terms? More intuitively, if there are empty terms believed to refer, why could not abstract terms be among them? Any principle of charity strong enough to refute the claim that there are such empty terms is just *ipso facto* unreasonable, so why think there is a reasonable principle strong enough to conflict with (EN)?

Although I think this is a serious point, I will not press it further, and none of the arguments below turn on it. (One might try arguing that abstract terms, as opposed to names of gods, are *useful or indispensable to our best scientific theories*, and claim that this is the principled demarcation sought for. This is of course a common line of reasoning, but note that the present question is rather if there is a principle of interpretation which can be used to refute (EN), and “indispensability arguments” do not concern interpretation.)

2. Davidson’s cases of absurd beliefs

Davidson’s arguments for interpretive charity divide into roughly two kinds. Firstly, there are transcendental arguments that (radical) interpretation is possible only if guided by

interpretive charity. We will briefly consider such general considerations toward the end of section 5. Secondly, he has argued in numerous texts by way of thought experiments, intended to show the intuitive absurdity of grossly uncharitable interpretations. Here is a representative example:

Let someone say [...], “There’s a hippopotamus in the refrigerator”; am I necessarily right in reporting him as having said that there is a hippopotamus in the refrigerator? Perhaps; but under questioning he goes on, “It’s roundish, has a wrinkled skin, does not mind being touched. It has a pleasant taste, at least the juice, and it costs a dime. I squeeze two or three for breakfast.” After some finite amount of such talk we slip over the line where it is plausible or even possible to say correctly that he said there was a hippopotamus in the refrigerator, for it becomes clear he means something else by at least some of his words than I do. The simplest hypothesis so far is that my word ‘hippopotamus’ no longer translates his word ‘hippopotamus’; my word ‘orange’ might do better. (*Inquiries*, pp. 100f.)

While these examples may well support various important principles of charity, our question is whether it supports (Charity). Note first that if it does, it clearly cannot be demonstrably (deductively); rather, it must be non-demonstratively, i.e., either abductively or inductively.

Now, assuming the subject in the example is sincerely expressing his beliefs, it certainly seems that the sheer absurdity of these beliefs is sufficient to rule the interpretation out. Thus, one may think, one simply cannot interpret someone’s expressions in such a way that many of their beliefs turn out (obviously) untrue. However, as should be clear from the foregoing sections, we have to be careful in making such sweeping claims about “making beliefs come out untrue”. Plausibly, the

interpretation above is implausible because it violates (C). To wit, it makes the subject come out wrong not because it involves a controversial assumption about the world (such as (EN)), but because, given certain obvious facts about the nature of hippos and oranges, it makes the subject come out badly wrong when more charitable alternatives are available.

While Davidson's examples might be used to infer (C), they seem clearly inapt as grounds for (Charity). The reason is that they all involve errors consisting in the acceptance of sentences that are *false due to mispredication*. But of course, (EN) does not say that we can refer to abstract objects but have massively false beliefs about them. The type of massive error that (EN) imputes consists rather in the acceptance of sentences that are *untrue due to reference failure*. Thus, these intuitions cannot be appealed to in support of (Charity). (I might here also remind the reader of Colin McGinn's persuasive counter-example against principles ruling out massive error due to mispredication: the ancients who believed that stars are apertures in the dome of the sky, letting through light from outside the dome—see his 'Charity, Interpretation, and Belief', *Journal of Philosophy* 74 (1977), pp. 521-535.) An important lesson to learn from this, I believe, is that great care should be taken with the ambiguous phrase "massive error". It may well be that some have come to doubt that massive error of many kinds are impossible on the basis of intuitions about what is in fact a rather specific kind.

3. Arguments from the reliability of perceptual beliefs.

Many philosophers mainly discuss charity in relation to empirical beliefs. For instance, in their detailed discussion of principles of charity, Ernest LePore and Kirk Ludwig think

Davidson is best interpreted as arguing for a principle of charity for ‘beliefs prompted by the speaker’s environment’ (*Donald Davidson: Meaning, Truth, Language, and Reality* (Oxford: Oxford University Press, 2005), at p. 189). Let us assume that interpretation of “observation sentences” indeed requires consistency with the assumption that speakers mostly accept true and only true such sentences (or, for convenience, that perceptual beliefs are mostly true). The question here is whether we can infer (Charity) from this fact.

It should be clear that there is no way of inferring (Charity) inductively from the reliability of our perceptual beliefs. The fact in question is simply not a proper inductive base for the former, which concerns beliefs in general. Let us therefore look at the case for abduction. In order to show that such an inference fails, we need to argue that there is an alternative explanation that is not clearly inferior to (Charity), and which does not entail it. There are probably several such explanations, but it will suffice to present one, since this suffices for refuting the abduction (pending further argument).

The alternative “explanation” I have in mind could be expounded in roughly the following way: Take a perceptual belief, e.g., that a given object before me is white. This belief is, *qua* the belief it is, typically evidenced/caused by perceptions representing something as white. This is an essential aspect of its cognitive role. Let us now ask what it might be for a perception to represent something as white. Surely, it cannot be an intrinsic property of the event that is the perception. It is more likely something which includes at least the property of having some intrinsic property that is reliably (but not universally) instantiated in the presence of suitably placed white things and only in their

presence, i.e., that “tracks” white things. (Note that this is much weaker than the kind of causal theory that faces the “disjunction problem”.)

If this holds for beliefs that something is white, it must hold of all such simple beliefs involving simple recognitional concepts. Similarly simple beliefs involving complex recognitional concepts, like “Trafalgar Square”, “cat”, etc., presumably derive their reliability from that of beliefs involving simpler recognitional concepts. If some story of this kind is correct, then perceptions are by their very nature mostly veridical and then, because of their constitutive connection with perceptions, perceptual beliefs are mostly true by *their* very nature.

This is an explanation of our explanandum very different from one which simply takes (Charity) to be the explanans, and which clearly does not entail (Charity). Further, it might for all we know be the best explanation of the explanandum. Since (Charity) could, if at all, only be abductively inferred from the explanandum, the fact that we cannot decide which explanation is better means we cannot (at present) infer (Charity) from the reliability of perceptual beliefs at all.

4. The reliability of non-abstractive inferences

Could one not, by close analogy to the kind of argument for (Charity) above, think that the reliability of our inferential practice in general indicates the truth of some general principle like (Charity)? Firstly, one could not, of course, assume in such an argument that *abstractive inferences* are reliable (i.e., mostly truth-preserving), where abstractive inferences are ones that run between nominalistic sentences to sentences entailing the existence of abstract objects, e.g.,

(P) x is $F \Leftrightarrow F$ -ness is a property of x .

But perhaps, it may be thought, we need only assume that our *non-abstractive demonstrative and/or non-demonstrative inference modes* are reliable, and then argue that this gives good grounds for some general principle like (Charity).

As before, I do not wish to deny this assumption, but I think it can be explained without commitment to a general principle of charity like (Charity). However, we must first notice an ambiguity in our description of the explanandum: on one reading, it is the fact that the specific inference modes we employ (induction, particular logical inferences, etc.) are reliable. To explain this, one must look at each individual inference rule, and explain why it is reliable. And the explanation must be rather trivial; for instance, induction is reliable basically because the world is such as to make it reliable, and similarly for demonstrative and other non-demonstrative inferences (for note that reliability is here defined in terms of truth and not rationality or justification). This kind of explanation must be accepted by anyone who agrees at all that the inference mode in question indeed is reliable. But such explanations (of course) do not entail (Charity). For instance, one can gladly accept this kind of explanation while holding that abstractive inference modes are unreliable, which would violate (Charity).

On another reading, the explanandum is rather the fact that the inference modes we employ (whatever they are) are reliable. The alternative explanation I would like to point to here consists, very roughly, of two explanantia: (1) non-abstractive inferences modes that are unreliable would not be useful, but, on the contrary, detrimental to our

interests, and (2) we tend to employ only useful modes of inference and avoid detrimental ones.⁷ Though this explanation can certainly be elaborated, I think the above suffices for showing, by the same kind of reasoning as above, that an abductive inference from the reliability of our non-abstractive non-demonstrative inference modes to (Charity) would be premature. However sketchy, it would be premature to say that we can already see that an explanation involving (Charity) is better than all the various more precise explanations that can be extracted from this sketch.

Note that this point immediately answers one of the first arguments for interpretive charity in the literature, namely, Quine's section about translating logical constants, in *Word and Object*, § 13. In brief, his idea is that if a foreign expression is translated as a given constant in the home-language, say, a connective, in such a way that, on this translation, the translatee comes out as accepting many logical falsehoods and invalid inferences, then the translation is *ipso facto* a bad one. If this is right, one might suspect that this must hold more generally, or at least for inferences featuring abstract terms, such as (P) above. For instance, the generalisation might hold for all inferences defining *non-empirical expressions*, or for all inferences that are *a priori* or *analytic*. Such principles, while significantly weaker than (Charity), would still contradict (EN). But in reply to this, we can simply point to (1) and (2) as an alternative explanation of his observation.

⁷ Williamson argues against the idea that our beliefs are true because it is useful to have true beliefs (*The Philosophy of Philosophy* (Oxford: Blackwell, 2007), section 8.2). But his argument essentially concerns non-connective propositions and so does not touch the present explanation.

Admittedly, it may be thought that Quine wanted to say more than merely that most of our logical inferences are valid, perhaps that it is in some sense “constitutive” of (logical) concepts that people infer with them reliably. But note that even if a given constant is such that it is inconceivable and impossible that someone uses it without mostly inferring with it in accordance with certain rules *that happen to be valid*, one cannot infer that it is impossible to infer with a constant in accordance with invalid rules. And there seems to be nothing in Quine’s example preventing nominalists simply to reject the generalisation to abstract terms. (We will see anon, however, that there are several differences they can point to in order to motivate the claim that abstract terms are exceptions.) Further, the sense of unintelligibility might stem merely from the presumption that concept-users cannot use grossly detrimental concepts (like “tonk”). If so, then appealing to (1) and (2) would not be to ignore the force of Quine’s observation.

Even if (1) and (2) allow us to explain the reliability of non-abstractive inference modes, it may seem that I have unduly restricted (1) so as to cover only non-abstractive principles. One might object, what reason is there to think that non-abstractive principles are any different with respect to usefulness vis-à-vis reliability? Well, such reasons have been elaborated at great length by nominalists. For instance, Hartry Field has argued that we can “facilitate inferences” by postulating abstract objects (*Science without Numbers* (Princeton: Princeton University Press, 1980), Chapter 1). In my forthcoming ‘Deflationist Error-Theory of Properties’, I suggest instead that the purpose of abstract discourse is to increase the expressive power of our language, in a certain sense. To wit, with an expression meaning *property*, we can formulate sentences with an inferential behaviour unlike that of any other sentence in the “old” language. Since nothing like this

can be intelligibly said of non-abstractive inference modes, this idea does not undermine (1) by suggesting that non-abstractive inference-modes might be unreliable yet useful, too. However controversial these proposals may be, it should be clear that discussing them would take us too far away from the considerations about interpretive charity, which constitute the subject matter of this paper.

There is another obvious difference between abstractive and non-abstractive inference modes that does not pertain to usefulness. To wit, non-abstractive inferences (both demonstrative and non-demonstrative) have that in common that their reliability does not depend on what objects there are. Not so, of course, for abstractive inference modes. If it is more reasonable to assume that speakers are wrong about what things there are than wrong in their basic inference modes that do not rely on the existence of certain objects, then that, too, tells against arguing against (EN) on the basis of Quine's observation about logical constants (or from the reliability of non-abstractive inference modes in general). It is plausible (though I will not argue this point here) that there is a connection between this difference and the difference pertaining to usefulness.

5. Concluding remarks

In the foregoing sections, I have argued that the most representative arguments in favour of interpretive charity do not support (Charity), and hence this principle, while conflicting with (EN), is currently without support. The foregoing discussion has also provided us with enough material, I think, to give a diagnosis of the belief in a universal charity strong enough to rule out (EN). It seems clear that, for many different—and rather *disparate*—types of expression, it would be absurd to interpret people as being in

massive error in their use thereof. To wit, we have seen that one cannot reasonably interpret someone as being in massive error due to mispredication, nor in one's perceptual beliefs, and nor in one's basic (non-abstractive) inferences. To this, we might add the fact that non-semantic interpretation, i.e., inferring what speakers mean when they speak non-literally, also seems to require the assumption that people typically mean something true (cf. Paul Grice's Maxim of Quality and Dan Sperber and Deirdre Wilson's similar pragmatic principle in their Postface to the second edition of *Relevance: Communication and Cognition* (Oxford: Blackwell, 1995)).

Taken together, the number and disparity of these cases provide convincing *prima facie* reason to think that some important universal principle of charity must hold. And it is not surprising if one should also suspect that this universal principle of charity should be strong enough to rule out an error-theory like (EN). But we have seen that for each of the examples we have treated, there is an explanation that explains the relevant reliability by appeal to something particular to the case at hand (the constitutive relationship with perceptions, the different types of usefulness of different kinds of inference). This indicates that one could find exceptions to the rule, e.g., abstract terms.

This is not to say that there are no true, important, universal principles of charity that these various examples point to. For instance, it is common among nominalists to think that abstract discourse, by the special inferential tools it affords us, makes us more efficient in arriving at new true nominalistic claims, even if the "way there" involves false assumptions. Their *raison d'être*, then, is after all to help us arrive at more truths. Thus, adherents of (EN) can accept the following universal principle of charity:

(Ch) For every expression e , a good interpretation must be consistent with the assumption that using e overall helps us arrive at more truths.

(Ch) is a true, universal principle of charity that involves truth, and yet is consistent with (EN). Although I will here rest content with (Ch), there may well be other such principles, and each of them will plausibly help adherents of (EN) explain the appeal of universal charity consistently with (EN).

So far, I have only argued that (Charity) is not supported by the most representative arguments in the literature. A more head-on strategy for defending (EN) is to argue that it is false. And, indeed, it seems that the very existence of nominalists (by which I mean adherents of (EN)) casts doubt on it. For, surely, adherents of (EN) should be interpreted the same way as realists—both mean the same thing by “property”, etc. Yet they differ radically on abstraction principles like (P). So one of them must be in massive error concerning abstracta. But (Charity) does not allow this kind of massive error. Thus, assuming there really are “adherents of (EN)” (which I think could be denied, but not with much plausibility), (Charity) must be false.

One might wonder how nominalists and realists can indeed mean the same if they disagree about such fundamental principles about (P). But there is a well-known answer to this worry, which is that they mean the same because they share a *defeasible disposition* to accept the relevant principle (whose manifestation, however, is inhibited by

Arvid Båve

certain philosophical considerations in the nominalist's case). This is controversial, of course, but this does not seem to affect the plausibility of her case against (Charity).⁸

Department of Philosophy, Linguistics, and Theory of Science

Gothenburg University

Box 200

405 30 Gothenburg

Sweden

arvidbave@gmail.com

⁸ Thanks to the audience at the workshop “Truth And Nothing But the Truth”, Humboldt University, Berlin 2010, for comments on an earlier draft of this paper.