EXTRAPOLATION AND SCIENTIFIC TRUTH
Louis Caruana

I am taking extrapolation to mean the act of inferring more widely from a limited range of known facts. This notion of extrapolation, when especially applied to past events, has recently been used to formulate a pragmatic definition of truth. The argument in this paper is meant to show that this definition has serious problems. The pragmatic definition of truth has been formulated in the context of the discussion concerning internal realism. In this paper, the basic internal realist starting point, as expounded in the works of H. Putnam and N. Jardine, will be taken to be the claim that the distinction between judgements which are true from judgements which are false is explainable from within our judgement-forming faculty — no recourse to the external world is necessary. There are three steps in the argument I am presenting. I will start first by showing that the pragmatic definition of truth involves a counterfactual statement. The second step will be to describe how the counterfactual definition is apparently justified by an inductive argument. In the final step, I show that this justification is not successful, and I will then draw out some problems with the definition of truth.

Putnam's definition of truth may be expressed in the following way,[1] which I will be referring to as statement S1:

S1:   X is true if and only if X is verifiable in the long run by an ideal community of inquirers.

---

[1] A discussion clearly on this point can be found in one of his recent writings: H. Putnam, 'Replies,' *Philosophical Topics*, vol. 20(1) (1992), pp. 363f.

There are two difficulties to be faced when working with S1. The first one concerns the future: is the long run possible? It seems that, given our limited economic possibilities, a community of inquirers can never envisage an inquiry which takes an unlimited time. Costs of research get bigger and bigger. Moreover, some moral obligations concerning the resources of the planet may also constrain particular types of inquiry to favour others; for example instead of exploring more about Higgs bosons by building a Superconducting Supercollider, one may feel obliged to search for fairer methods of distributing food.[2] Such constraints on the long run makes us realise that the idea behind S1 may probably be better expressed as a hypothetical statement of the following form: if an inquiry in the long run were possible, then truth may be expressed according to S1.

A second problem with S1 will push more in this same direction. The first one concerned the future. This concerns the past. It seems unquestionable that the past has been, in its time, as determinate as the present is now. We are therefore obliged to hold that there are many propositions about the past whose truth for us is inaccessible. For example, we cannot doubt that there was a definite number of tyrannosaurus roaming the earth eighty million years ago, but we are practically certain that there is now no extant information from which this number is recoverable. Such questions will remain unresolved even if human inquiry had no economic or moral restrictions but had all the resources it needs to continue for ever. Our question therefore is: how is S1 to cater for evidence-transcendent truth-values? I am here considering only the evidence we expect to have in the actual world we live in. I am not including all possible evidence. For instance, I am not considering the imaginary situation where the dinosaurs wrote their population figures on tablets which will be found in the year 2050. I am not including all *possible* evidence because, when we talk of evidence-transcendent truth-values of a contingent kind as we have here, we are assuming a

---

[2] US Congress ruled against the continuation of the SSC project in Nov., 1993. See D. Ritson, 'Demise of the Texas supercollider', *Nature* vol. 366, 16th December, 1993.

situation where we claim that the evidence required is inaccessible *because* of the laws of nature as we know them. These laws of nature act as a fixed background with reference to which evidence for given statements is judged accessible or inaccessible. The way one may make S1 account for evidence-transcendent truth values is to formulate it as a counterfactual conditional in the following way: if a human being had been present fifty million years ago, then the number of such creatures could have been counted.

If these qualifications are made to S1 then the formulation will in fact be very similar to that suggested by N. Jardine[3] and which I will call statement S2:

S2:   X is true if and only if X is verifiable by evidentially unrestricted observers.

Notice that this hides also a hypothetical structure, because what S2 is in fact saying is the following. For the dinosaur case, for example, it is saying that 'If I were an evidentially unrestricted observer, then I would be able to count so many tyrannosaurus'. For the general case, this formulation becomes: 'For any X, if I were an evidentially unrestricted observer, X is true if and only if X is verifiable by me'.

I hope this is enough to illustrate the counterfactual nature of the pragmatic theory of truth I am investigating. Some may think there is something wrong with this account of truth, because it is not only hypothetical but also counterfactual. They may say that for human beings to be out and about in prehistoric forests with pencil and paper in hand eighty million years ago is physically impossible, especially if we consider, as I suggest we should, that time-travel is a conceptual impossibility.

---

[3] N. Jardine, *The Fortunes of Inquiry*, Clarendon, 1986, p. 29.

Now I come to the second step of my argument. I will mention how we are led to think that such a definition is in fact justified. The justification is an inductive argument. In everyday life, we very often extrapolate from a situation of known parameters to one of unknown parameters. For example, knowing that cats do not disappear into thin air, and that my cat went into the garage (which has no windows), then I know that if I go into the garage I will find it there. I am here employing some simple laws of nature. Even if I do not own any cats, the above inference may be made without the slightest problem. In other words, in everyday life we are accustomed to consider counterfactual conditionals as true or false. Put simply, for a law-statement to support a counterfactual it must be the case that the law-statement plus the fully stated antecedent of the counterfactual must entail the consequent of the counterfactual.[4]

In a similar way, we may employ our theories to conclude for example that if the fundamental constants in the early universe were different from the values we know, then no life would have appeared on Earth.[5] This type of extrapolation in everyday and in scientific discourse may be carried out in other ways. For example, I may perhaps hold that 'If I were the same person as Aristotle, then I would not like karaoke'. The statement S2 is trading on the kind of extrapolation exemplified by these cases. By induction, we claim that since extrapolation holds in the everyday case, like the cat-example, then it holds for the example concerning the fundamental constants and also for the example concerning me and Aristotle. Hence, by keeping this sort of inductive justification in mind, we can understand better our previous tyrannosaurus case. What time-travelling observers who have unlimited evidence-gathering capabilities perceive into the past of eighty million years ago is allegedly

---

[4] Law-statements in many cases do, but statements of local and Humean uniformity generally do not, support counterfactuals. Laws are therefore not Humean uniformities, so the argument goes. See D.M. Armstrong, *What is a Law of Nature?* Cambridge University Press, 1983, p 46-52.

[5] Examples of fundamental constants are the speed of light in vacuum c, the Plank constant h, the elementary charge e, the rest mass of the electron and proton m_e and m_p. The weak anthropic principle holds that if these were different from the current values, we would not be here to measure them. The strong anthropic principle holds that they *cannot in principle* be different from the current values.

nothing more than an extrapolation of present theories. If I had been there, I would have seen them. This, it is claimed, is what makes S2 a plausible account of truth.

But is it? In the final third section of the paper I will attempt to bring out some serious problems with this inductive argument.

First of all, let us notice the difference between the three examples given above. In the cat example, the antecedent of the counterfactual 'if the cat goes into the garage' is a thought-situation, or possible world, which is not very different from the actual one I usually am in.[6] Hence, the same laws of nature are assumed to apply. In the fundamental-constants example, the antecedent 'if the fundamental constants had a value different from the current values' is a thought-situation which is more distant, as it were, from our actual one. But still we feel confident to say that the laws of nature apply in the new situation as well. This is because the antecedents in both cases are determinate. Their meaning is clear and they are well defined in the context of the relevant laws of nature. In the third case, however, the one involving me and Aristotle being the same person, we are introducing a thought-situation which is not only very far from our actual one. The example is also one whose antecedent is indeterminate. Given the laws of nature as we know them, it seems to me that we face great difficulty when we try to see what being the same person as Aristotle might mean.

So the differences between the three examples can be brought out in terms of two attributes related to the antecedent, namely: (1) the *distance* of the new thought-situation, and (2) the *determinacy* of the meaning of the antecedent. I will simplify matters by not allowing these attributes to have degrees. I will not discuss here thought-situations that are slightly more distant or slightly less distant that others. I will not discuss antecedents that are slightly more determinate or slightly less

---

[6] For my purposes here, I prefer 'thought-situation' to 'possible world'. My argument does not depend on the debate whether possible worlds are real or not.

determinate that others. If we limit ourselves to this working assumption, we have four possible combinations: distant and determinate, non-distant and determinate, distant and non-determinate, non-distant and non-determinate. I will not consider the last case. My suspicion is that one can never have a possible world which is not distant from the actual one and at the same time described by an antecedent which is not determinate. The other three combinations however correspond to the three examples above. The cat-example involves a thought-situation which is non-distant, and its antecedent is fully determinate. The example about fundamental constants introduces a thought-situation which is distant from the actual world we live in, but at least the antecedent is fully determinate. The example concerning me and Aristotle is problematic on both fronts: it is distant and also non-determinate.

Now with this conceptual baggage we are ready to tackle the original question. What kind of counterfactual statement is involved in S2? In other words, what kind of counterfactual is involved in the statement: 'For any X, if I were an evidentially unrestricted observer, X is true if and only if X is verifiable by me'? I will argue that the inclusion of evidentially unrestricted observers makes this counterfactual of the same kind as the third example. It is problematic on both fronts. It is both distant and non-determinate. I will discuss first the distance, and then the non-determinacy.

First, to show that the inclusion of evidentially unrestricted observers makes the counterfactual distant. To do this I will employ some logical machinery formulated specifically to cater for counterfactuals by Nicholas Rescher.[7] The first step is to begin with a group G of accepted statements representing our knowledge — these statements are the facts of the case. Then we add to this group the counterfactual hypothesis H. The combined group $G + H$, call it G', is inconsistent. That is what we

---

[7] N. Rescher, *Hypothetical Reasoning*, North Holland Publishing Company, 1964.

mean by saying that *H* is counterfactual. The third step is to realise that some statements of *G′* can be retained while others must be dropped in the interests of logical consistency. Suppose, after the 'pruning' of *G′*, as Rescher calls it, we arrive at a self-consistent *G**. Then, we can follow Rescher in claiming that the counterfactual conditional 'If *H* were so, then *X* would be so' is a *proper* counterfactual if *X* is a consequence of *G**, the pruned set of beliefs.

The crucial step is this pruning of *G′* to obtain *G**. What governs the pruning is the fact that we are attached to some of the original accepted beliefs more than to others. In fact, to find *G** is not strictly speaking a logical procedure.

The extent of the lack of consistency in *G′* depends on the nature of *H*. Some counterfactual hypotheses will contradict only a small number of statements in the original *G*, while other counterfactual hypotheses may be in contradiction with all, or nearly all, the statements in *G*. If the pruning of *G′* involves the dropping of only a small number of beliefs to obtain *G** — in other words, if *G′* and *G** are nearly the same — then the counterfactual is a proper one. If our definition of truth S2 had involved only such well behaved counterfactuals, then it would have been acceptable. But unfortunately it seems that the counterfactual in S2 is of the sort that necessitate a tremendous difference between *G′* and *G**. The inclusion of evidentially unrestricted observers makes the antecedent contradict an intolerably great number of the beliefs we start with, namely those governed by the laws of nature as we know them. The great number of contradicted prior beliefs shows that the thought-situation brought in by the antecedent is very distant from the actual one we know how to cope with.

Now the second point: to show that including evidentially unrestricted observers makes the antecedent non-determinate. This can be seen from the fact that if I say: 'suppose I am an evidentially unrestricted observer', it seems very difficult, if not impossible, to see what independent variables I will have to keep constant and

what independent variables I have to change in order to qualify as an evidentially unrestricted observer. J. Aronson, in his debate with N. Goodman about the way we are to account for a simple counterfactual like 'if we strike the match, it will light', rightly points out that, even in such simple cases, we do not know all the variables and laws involved.[8] The counterfactual I am analysing here is an extreme case of this. This is the same case as me being the same person as Aristotle. The new thought-situation goes against so many of our normal thought constraints that it seems very difficult to admit that such phrases have any meaning even though they seem to have a *prima facie* meaning. The upshot is that the particular counterfactual we are dealing with in S2 is non-determinate.

So this concludes my diagnosis. My line has been the following. The pragmatic definition of truth is presented in a counterfactual formulation. This counterfactual formulation can apparently be justified by induction, using our everyday use of counterfactuals. But it turns out that not all counterfactuals are proper. Some suffer from being distant, and some from being non-determinate, and some from both. The one involved in the formulation of pragmatic truth turns out to be both distant and non-determinate. Hence it is not satisfactory.[9]

---

[8] J. Aronson, *A Realist Philosophy of Science*, Macmillan 1984, pp 237-259. Notice the interesting distinction made by Aronson, *ibid.* p. 259n8: 'disagreement over the truth of a counterfactual may not reflect contrasting views of similarity among possible worlds, as Lewis believes (*Counterfactuals*, pp. 91-5), but that the disputants are simply working with rival theories about the world.'

[9] An earlier draft of this paper was read and discussed at the Cambridge Philosophy Faculty Graduate Seminar. I am grateful to those who were present at the seminar, and also to N. Jardine and P. Lipton for useful comments.