“Epistemologies and Apologies” first appeared in *Dialectica* Vol. 40, No. 1 (1986), pp. 45-58.

Quite a bit of recent work by epistemologists can be viewed as being part of a struggle between two theories of knowledge, viz., foundationalism and fallibilism.

Foundationalism is the older of the two epistemologies, being most commonly associated with Descartes. Fallibilism is somewhat younger, being most commonly associated with Peirce. I would like to examine the difference between these two views from a certain angle, namely, the angle of the justification of the principles of reasoning. Specifically, I would like to present the thesis that while foundationalism leads to the traditional problem of induction, fallibilism, which does not face that problem, faces a functional analogue. This "functional analogue," which I shall call *the problem of convergence*, seems to be nearly as difficult as the traditional problem of induction. The epistemology you hold will dictate the apology you must make for the principals of reasoning you employ.

The terms "foundationalism" and "fallibilism" are notoriously hard to define. In what follows, the definitions I give will be guided by the vision of Descartes and Locke as paragons of foundationalism, and Peirce as a paragon of fallibilism. But the labeling of figures is at best procrustean, and at worst misleading or even erroneous. So take my definitions as arising from, but not necessarily describing, certain historical positions. In particular, I think it is most profitable to look at the motivation behind each view.

The foundationalist is concerned with skeptical doubt. We intuitively feel that we do indeed have knowledge about many things. By "knowledge" we will mean true justified belief. (Gettier-type problems will not be of relevance in what follows.) But here skeptical doubt enters in. How can I tell if my beliefs are true and justified? How do I know that I am not completely deceived by them all? (Descartes' image of the Evil Demon serves to make this doubt systematic). We all know the skeptical doubt game: I ask you for your justification of claim *p1* and you cite evidence *p2*. I then ask for your evidence for *p2*, and you can only cite some other evidence *p3*. And so the regress goes.

The traditional foundationalist response is to stop the regress at some set of beliefs which are held to be self-evident. Traditional foundationalism, which I shall call *strong foundationalism*, held that:

* 1. There are foundational beliefs {*f1*,*f2*, ... } which can be/do form the basis for all other beliefs which can/are known.[[1]](#footnote-1)
  2. The {*fi*) are all incorrigible/indubitable/self-evident/certain.[[2]](#footnote-2)
  3. A person can truly claim to know *p* if and only if he can prove *p* on the basis of some of the {*fi*}).
  4. A person can truly prove a belief if and only if he can know by inspection that his proof is logically adequate.

The research program which grew out of strong foundationalism accordingly focused upon the nature of the {*fi*}, and sketching out the rational reconstructions of beliefs. Were the foundations general truths known by the light of reason (as the rationalists believed), or instead particular truths known by direct observation (as the empiricists held)? Did the reconstruction involve only deductively valid inferences or probabilistic ones as well? Should we reconstruct beliefs about minds in terms of the overt behavior of the body, or reconstruct beliefs about observable objects in terms of mental images?

Problems arose in carrying out the strong foundationalist research program. Fallibilism arose more or less in reaction to the failures of the old theory, and while characterizing fallibilism is more difficult (if for no other reason than the philosophy is still evolving rapidly), we can characterize it tolerably well by opposing it to strong foundationalism:

(1\*) The starting points for any justification of beliefs are those beliefs {*s1*, *s2*...} held to be self-evident by one's community.

(2\*) No beliefs, not even the {*si*}, are ever incorrigible, or "certain"/"self-evident" in any community-transcendent sense. Rather, the {*si*} are those beliefs that are held by the community to be not worth questioning.

(3\*) A person can legitimately claim to know *p* in a community if and only if he can prove *p* on the basis of some of the {*si*} of that community.

(4\*) A person can legitimately claim to have proven *p* in a community if and only if his proof is in accordance with the logical canons of that community.

The differences between the (strong) foundationalist and the fallibilist are profound. To begin with, reconsider the question of motivation. While the motivation for the foundationalist is theoretical—answering *skeptical* doubt, the motivation for the fallibilist is ultimately economic—answering *real* doubt. To use Peirce's term for it, the "economics of research" dictate that we actually spend our effort in answering real doubts, real questions set for us by our community. Skeptical doubt isn't the spur to scientific research, but real doubt is. Part of the training of a scientist is to learn both the questions and the methodology for answering those questions in his domain of research. (Naturally, the questions and ever the methodologies change over time. This is a point to which we shall return.)

The point about the difference in motivations brings up another difference in perspective. While the foundationalist typically seeks certainty in all knowledge via the certain reconstruction of beliefs upon incorrigible foundations, the fallibilist seeks reasonable security rather then certainty. We can, the fallibilist claims, have true knowledge, but we can't ever be certain about anything we believe—at least, we can't have the absolute certainty sought by the foundationalists. Everything is open to later revision.

This latter point (that everything is open to later revision) is no minor difference. For the fallibilist, even the laws of logic are potentially open to revision; indeed, Peirce spent a good deal of time investigating many-valued logics. To the fallibilist, deciding which logic is "the true logic" is a question that calls for detailed empirical (psychological, anthropological, historical and sociological) research and is not a question of absolute self-evidence. Peirce's analogy with geometry is apt: for many centuries people supposed Euclidean geometry to be indubitable, but the work done by Gauss, Riemann, Bolyai, Lobachevski et al. proved otherwise.

However, from the point of view of this paper, the most important difference between the two theories of knowledge is that between (4) and (4\*). For the strong foundationalist, knowledge is solitary: knowledge is something an individual sees that he has or lacks, depending upon whether he can actually carry out the justification and verify its adequacy. For the fallibilist, knowledge is social: knowledge is something a person claims, his claim's merit being dependent upon whether the community assents to his starting points and his logical methodology, i.e., the validity of his canons of reasoning. (Of course, he can "lead the community" as in a scientific revolution, that is, he can get his community to reject or modify some of the starting points or even some of their canons of reasoning, but only by appeal to the remainder of the beliefs or canons.)

As I said earlier, the fallibilist research program is still relatively young, and there does not seem to be widespread agreement among fallibilists about (for instance) how one individuates epistemic communities (by beliefs, language?), whether the fallibilist needs to reject the KK thesis, and so on. But I want now to return to strong foundationalism, because what I will later say about fallibilism grows out of what I will now say about foundationalism.

The strong foundationalist faces a problem, growing out of his acceptance of (4): you can truly prove a belief if and only if you can know by inspection that your proof is logically adequate. Knowing that your proof is adequate involves two things. First, you must know that the steps connecting *p* (the belief in question) to the foundation statements *fi*, *fj*,... upon which it is based must be in accordance with the canons of logic, and second you must know that your logical system is valid (or acceptable or adequate). Waiving problems about the possible corrigibility of your belief that your steps are all instances of your rules, you still have the problem of knowing that your logic is adequate. Traditionally, strong foundationalists have automatically assumed that demonstrative (deductively valid) inference is self-evidently adequate, whereas the principles of non-demonstrative inference are not self-evident, and need some justification. But, as Hume so well indicated, this leads to a problem.

The problem is this. To be justified in using an inductive logic N (according to the strong foundationalist view), you must be able to prove that N is adequate, i.e., that the arguments to which N assigns a high degree of inductive probability lead from true premises to true conclusions most of the time.[[3]](#footnote-3) But how can this be done?

The reader is no doubt familiar with most of the attempts to justify induction; is not, the second chapter of Skyrms (1975) has an excellent survey.[[4]](#footnote-4)

Obviously, we cannot give a deductively valid proof of the adequacy of N. For we have in the premises of such a proof only information about how well N has worked in the past, and that information cannot logically entail any statement to the effect that N will work in the future. Yet to prove N is adequate, we must show that the arguments which it sanctions lead from true premises to true conclusions most of the time past, present and future.

One the other hand, how would one construct an acceptable inductively strong proof of the adequacy of N? If one proves N is justified by using N, doesn't that beg the question? It seems to.

Of course, those who seriously attempt to meet the challenge do so in ways more subtle, though in the end no more satisfactory than those indicated above. For instance, a more sophisticated attempt at an inductive justification of induction would be this.[[5]](#footnote-5) Begin by distinguishing *levels* of arguments, with *level 1* arguments being arguments about individual things and events; *level 2* arguments being arguments about level 1 arguments (such as: "Some level 1 arguments are valid; all level 1 arguments which are valid are explicable formally; therefore, some level 1 arguments are explicable formally"); *level 3* arguments being arguments about level 2 arguments, etc.

Then we can distinguish between levels of inductive logic, with N1 being that inductive logic which evaluates level 1 arguments, etc. Now, we can justify induction by showing that N1 has been adequate in the past, and arguing that it will therefore be adequate in the future. That is, we justify N1 by giving a level 2 argument which is sanctioned by N2. If the skeptic has doubts about N2 we can justify it by a level 3 argument sanctioned by N3 and so on.

Alas, this layer-cake approach isn't satisfactory. After all, one could play the same shell- game with a counter-inductive logic Nc, which sanctioned arguments of the form "such and such happened in the past, therefore, it will not happen in the future". You could justify such level 1 arguments by appeal to Nc1. But, of course, given the way our world works, level 1 arguments sanctioned by Nc1 will lead to inaccurate predictions. Fine, because we can then formulate a level 2 argument to the effect that that arguments sanctioned by Nc1 have not led to true predictions in the past, so they will in the future; and thus level 2 arguments are sanctioned by Nc2. This raises (quite proper) skeptical doubts about Nc2, doubts about Nc2 by Nc3, and so on.

Again, Reichenbach and Feigl floated a more sophisticated deductive defense of induction. Very crudely put, the defense consists of showing that, if any inductive logic will work, N will. (Here the precise formulation of the logical system N becomes crucial.) But we must use some system of inductive logic, even if it is only spinning a bottle. So we are justified in using N.

Unfortunately, this sophistication buys us little. The essence of the defense amounts to recognizing that even if some crazy system of making predictions (say, spinning a bottle of Jack Daniels whiskey) works (makes accurate predictions) on the level of individual objects and events, then scientific induction N will have at least one accurate prediction, namely, the prediction that that crazy system will continue to work. Put another way, the essence of this deductive justification of induction is the recognition that to even say that a method "works" and to advocate its continued usage involves generalization in accordance with N.

Yet this insight is too feeble to do the job required. As Skyrms puts it:

Now this is quite different from showing that if any method works on any level then scientific induction will also work on that level, or even from showing that if any method works on level 1 then scientific induction will work on level 1. Instead what has been shown is that if any other method is generally successful on level 1 then scientific induction will have at least one notable success on level 2; it will eventually predict the continued success of that method on level 1.[[6]](#footnote-6)

And as Skyrms notes, this is quite compatible with the possibility that scientific induction (system N) doesn't work at the level of ordinary world of things (as opposed to arguments).

The problem of induction is nothing new. But what is somewhat more recent is the recognition that the same sort of problem faces any logic, deductive logic not excepted. To be justified in using any system of deductive logic L (on strong foundationalist grounds), one must prove that L is adequate, i.e., that the arguments L sanctions lead from premises to true conclusions all of the time. But as Haack observed, it is no less difficult to justify deduction (in the sense required) than it is to justify induction:

... we cannot justify deduction inductively, because to do so would be, at best, to show that usually, when the premises of a deductive argument are true, the conclusion is true too—which would be too weak; and we cannot justify deduction deductively either, because such a justification would be circular.[[7]](#footnote-7)

And, as before, increased sophistication gets us no closer to meeting the challenge. For instance, we might distinguish between levels of deductive arguments, and justify (say) modus ponens on one level by using it at a higher level. But Haack devises a counter rule, Modus

Morons ("From A B and B to infer A"), which can be justified on one level by using it on a higher level.[[8]](#footnote-8)

Of course, the strong foundationalist might throw up his hands in frustration and insist that the steps in any proof are directly knowable by inspection. But any particular step is valid only if any step of the same form would be valid (otherwise, the difficulty would be compounded: he would have to say why *this* step is allowable for him and not for others). But this gets back to the notion of a logical rule. Perhaps the strong foundationalist might allege that the rules of logic (of whatever variety) are simply self-evident. But this seems very dicey, indeed. For millennia Euclidean geometry seemed self-evident, but it isn't. In the realm of deductive logic (to elaborate upon a point mentioned earlier), the classically valid *Tertium non dater”* (i.e., “*p v ~p*") is not valid in threshold logic, nor is the law "*p* (*q* *p*)" valid in

relevant logics.[[9]](#footnote-9)

If the laws of demonstrative inference are thus not knowledge by inspection, then how much more dicey such a claim (that the steps of reasoning are knowable by direct inspection) is when applied to non-demonstrative inference! For example, consider the rule dubbed by Haack "Modus Morons", more usually called the Fallacy of Affirming the Consequent:

p q

q

\_\_\_

p

This argument is deductively invalid. But great controversy surrounds this argument form regarding its status from the point of view of inductive logic. Put in terms of confirmation the rule

T O

O

\_\_\_\_\_\_\_\_\_\_

O

(the arrow representing "predicts" in any way you care to take it—as material implication, strict entailment, or whatever) is held to be inductively strong by corroborationists: successful predictions serve to establish a given hypothesis. Again, put the argument form in terms of explanation:

F

H F

\_\_\_\_\_\_\_\_\_\_\_\_\_

H

(Where now the arrow now symbolizes "is the best explanation of"), and you have an argument form held to be good (either inductively strong or else good in some other sense). Peirce and Hanson called this form of reasoning variously "retroduction", "abduction", and "hypothesis". Harman has called it the "inference to the best explanation".

Now, some writers (such as Popper and Lauden) think the inference to the best explanation is simply bad reasoning. Some (such as Fumerton[[10]](#footnote-10)) view it as either confused or enthymematic for some other sort of reasoning. Some view it merely weak. And some (Hanson, Peirce, Harman and others) view it as acceptable. Similar controversy surrounds reasoning by analogy as well as enumerative induction (the straight rule).

In sum: thesis (4) (and (2)) of strong foundationalism lead to skepticism. Strong foundationalism has as its chief motive the defeat of skepticism, but has as its chief result the creation of an even more powerful skepticism.

Lately, some philosophers have tried to devise a foundationalism which does not face the problems of justifying either foundations or the principles of reasoning from those foundations. For example, Alston has floated a version of foundationalism much weaker than the traditional sort I have discussed above.[[11]](#footnote-11) Essentially, this weak foundationalism only holds:

* 1. There are foundational beliefs {*f1* ,*f2*..}which form the basis of all other justified beliefs;
  2. A person can truly know *p* if and only if he can justify *p* on the basis of the {*fi*}.

This form of foundationalism is indeed clearly weaker than the traditional form. Unlike the strong foundationalist, the weak foundationalist is not committed to the incorrigibility of the foundations in any sense of the term "incorrigible". And under the weak foundationalist view, a person may be justified in believing something, yet not be able to "justify" his justification in the strong sense of proving adequate the logical system used.

One the other hand, weak foundationalism seems utterly useless. The major motivation for foundationalism was to refute skepticism, and all that remains of the original anti-skepticism is a nostalgic attachment to the concept of "foundations". One has no assurance that either the foundations or the derivations are sound.

Perhaps the fallibilist is no better off—we shall see. But weak foundationalism still involves an empirical thesis—that knowledge is based upon a fixed type or set of foundations.[[12]](#footnote-12) In the face of the abandonment of the anti-skeptic motivation, why bother holding on to such a dubious claim? The claim is certainly not empirically attractive: can you name these community- transcendent foundations? Sense-data, perhaps? But theory-neutrality is no longer in vogue, to say the least.

Why not accept instead the more empirically adequate fallibilist claim that in any research setting there are some agreed upon starting points, which may well be highly theoretical or subject to revision? Such a view seems much more plausible: one's belief in rain tomorrow is really based more often on the word of the weatherman than some set of sense-data or such.

To be fair, however, we haven't spent much time criticizing the fallibilist perspective. Lest I be accused of special pleading, I would like to turn now to the problem of convergence, the problem which I view as the functional equivalent (in fallibilism) of the problem of induction.

To begin with, it is clear that in any epistemology one needs some principle governing the use of justification techniques, if not the strong on discussed above. Otherwise, any proof technique, any logic would be acceptable. And so anybody could prove or justify any belief he wanted to. It is not a good bargain to swap skepticism, the view that nobody knows anything, for pansophia, i.e., the view that all people know everything. And indeed, the fallibilist does have a restriction built in regarding the logical rules employed in justifying a belief: those rules must be accepted by the community of which one is a member.

Yet this restriction seems hardly enough. Remember that the fallibilist is not a special type of skeptic—a skeptic, say, with a particular sociology of knowledge. The fallibilist, while perhaps not motivated by the goal of answering skeptical doubts, in fact does assert that knowledge is actually achievable. In reviewing the fallibilist theses (1\*)—(4\*), we see only the preconditions for people being justified in thinking that they know. My belief in p may be based upon my community's starting points, my proof may be acknowledged by one and all to accord with the canons of reasoning my community (including myself) accepts, and there may be no real reason to doubt those canons or starting points. Yet one hundred years later, my claim— however legitimate—may be seen to be false, along with the whole basis upon which it rested. May not that be true of every possible community of researchers?

So far as theses (1\*)—(4\*) are concerned. I believe the answer is—yes. The fallibilist position so far defined only establishes that people are often justified in claiming to know— something quite compatible with skepticism. It seems clear that what is required to demarcate fallibilism from skepticism is something like:

(5\*) There has been, is now or will be some community whose starting points are true, and whose canons of reasoning are valid. (Call this *the ideal community*)[[13]](#footnote-13)

With such a principle, it becomes rather easy to demarcate fallibilism from skepticism. I have, truly have, knowledge if my proof is such that it is acceptable in the ideal community. (By "ideal" I by no means imply that that community is somehow unachievable; rather, this postulated community is "ideal" in the sense of being perfect for the job of defining genuine knowledge.)

The concept of the ideal community, whatever its drawbacks, does enlighten us in just the way fallibilists want to enlighten us. To take just one instance, we can (from this expanded fallibilist perspective) better understand the history of science, which is full of people who were justly honored in their times as having discovered (or known or established) some alleged fact or theory later found to be wrong, and full of people treated with scorn for their beliefs or methodologies, later found to be correct. (Such historical phenomena are notoriously hard for foundationalists to account for: if there are absolute foundations and obviously correct logical principles, why did intelligent people of times gone by not recognize them?)

But I just did something sly. In describing the historical cases, I assumed that our community is the ideal one. But surely that is epistemologically debatable, not to say arrogant. Which community or communities are ideal? Might not (say) Aristotle and his followers have been the ideal community? Or Hitler and his minions?

If we don't want to arrogantly assume that our community is the ideal one, and we don't know which if any of part communities is the ideal one, we have to resort to some sort of open- ended characterization such as:

(6\*) The ideal community is the community whose beliefs and logical practices are those which would be the result of an infinite long run of research.

More succinctly, genuine knowledge is belief converged upon in the infinite long run. Fallibilist writings contain expressions of this view, sometimes expressed as a theory of truth, other times—more recent times—expressed as a version of realism. Some sample quotes:

Truth is a character which attaches to an abstract proposition, such as a person might utter. It essentially depends upon that proposition's not professing to be exactly true. But we hope that in the progress of science its error will indefinitely diminish, just as the error of 3.14159, the value given for will indefinitely diminish as the calculation is carried to more and more places of decimals. What we call is an ideal limit to which no numerical expression can be perfectly true. If our hope is vain; if in respect to some question—say that of the freedom of the will—no matter how long the discussion goes on, no matter how scientific our methods may become, there never will be a time when we can fully satisfy ourselves either that the question has no meaning, or that one answer or the other explains the facts, then in regard to that question there certainly is no truth.

But whether or not there would be perhaps any reality is a question for the metaphysician, not the logician. Even if the metaphysician decides that where there is no truth there is no reality, still the distinction between the character of truth and the character of reality is plain and definable. Truth is that concordance of an abstract statement with the ideal limit towards which endless investigation would tend to bring scientific belief, which concordance the abstract statement may possess by virtue of the confession of its inaccuracy and one-sidedness, and this confession is an essential ingredient of truth. A further explanation of what this concordance consists in will be given below. Reality is that mode of being by virtue of which the real thing is as it is, irrespectively of what any mind or any definite collection of minds may represent it to be. The truth of the proposition that Caesar crossed the Rubicon consists in the fact that the further we push our archaeological and other studies, the more strongly will that conclusion force itself on our minds forever—or would do so, if study were to go on forever. An idealist metaphysician may hold that therein also lies the whole reality behind the proposition; for though men may for a time persuade themselves that Caesar did not cross the Rubicon, and may contrive to render this belief universal for any number of generations, yet ultimately research—if it be persisted in—must bring back the contrary belief. But in holding that doctrine, the idealist necessarily draws the distinction between truth and reality.[[14]](#footnote-14)

... all the followers of science are animated by a cheerful hope that the processes of investigation, if only pushed far enough, will give one certain solution to each question to which they apply it. One man may investigate the velocity of light by studying the transits of Venus and the aberration of the stars; another by the oppositions of Mars and the eclipses of Jupiter's satellites; a third by the method of Fizeau; a fourth by that of Foucault; a fifth by the motions of the curves of Lissajoux; a sixth, a seventh, an eighth, and a ninth, may follow the different methods of comparing the measures of statical and dynamical electricity. They may at first obtain different results, but as each perfects his method his processes, the results are found to move steadily together toward a destined center. So [it is] with all scientific research. Different minds may set out with the most antagonistic views, but the progress of investigation carries then by a force outside of themselves to one and the same conclusion. This activity of thought by which we are carried, not where we wish, but to a foreordained goal, is like the operation of destiny. No modification of the point of view taken, no selection of other facts for study, no natural bent of mind even, can enable a man to escape the predestinate opinion. This great hope is embodied in the conception of truth and reality. The opinion which is fated to be ultimately agreed to by all who investigate is what we mean by truth, and the object represented in this opinion is the real. That is the way I would explain reality.

But it may be said that this view is directly opposed to the abstract definition which we have given of reality, inasmuch as it makes the characters of the real depend on what is ultimately thought about them. But the answer to this is that, on the other hand, reality is independent, not necessarily of thought in general, but only of what you or I or any finite number of men may think about it; and that, on the other hand, though the object of the final opinion depends on what that opinion is, yet what that opinion is does not depend on what you or I or any man thinks. Our perversity and that of others may indefinitely postpone the settlement of opinion; it might even conceivably cause an arbitrary proposition to be universally accepted as long as the human race should last. Yet even that would not change the nature of the belief, which alone could be the result of investigation carried sufficiently far; and if, after the extinction of our race, another should arise with faculties and disposition for investigation that true opinion must be the one which they would ultimately come to. "Truth crushed to earth shall rise again," and the opinion which would finally result from investigation does not depend on how anybody may actually think. But the reality of that which is real does depend on the real fact that investigation does not depend on how anybody may actually think. But the reality of that which is real does depend on the real fact that investigation is destined to lead, if continued long enough, to a belief in it.[[15]](#footnote-15)

But all this requires a conceived identification of one's interests with those of an unlimited community. Now, there exist no reasons, and a later discussion will show that there can be no reasons, for thinking that the human race, or any intellectual race, will exist forever. On the other hand, there can be no reason against it; and, fortunately, as the whole requirement is that we should have certain sentiments, there is nothing in the facts to forbid our having a hope, or calm and cheerful wish, that the community may last beyond any assignable date.

It may seem strange that I should put forward three sentiments, namely, interest in an indefinite community, recognition of the possibility of this interest being made supreme, and hope in the unlimited continuance of intellectual activity, as indispensable requirements of logic. Yet when we consider that logic depends on a mere struggle to escape doubt, which, as it terminates in action, must begin in emotion, and that, furthermore, the only cause of our planting ourselves on reason or that other methods of escaping doubt fail on account of the social impulse, why should we wonder to find social sentiment presupposed in reasoning?[[16]](#footnote-16)

As to reality, one finds it defined in various ways; but it that principal of terminological ethics that was proposed be accepted, the equivocal language will soon disappear. For “realis” and “realitas” are not ancient words. They were invented to be terms of philosophy in the thirteenth century, and the meaning they were intended to express is perfectly clear. That is real which has such and such characters, whether anybody thinks it to have those characters or not. At any rate, that is the sense in which the pragmatist uses the word. Now, just as conduct controlled by ethical reason tends toward fixing certain habits of conduct, the nature of which (as to illustrate the meaning, peaceable habits and not quarrelsome habits) does not depend upon any accidental circumstances, and in that sense may be said to be destined; so, thought, controlled by a rational experimental logic, tends to the fixation of certain opinions, equally destined, the nature of which will be the same in the end, however the perversity of thought of whole generations may cause the postponement of the ultimate fixation. If this be so, as every man of us virtually assumes that it is, in regard to each matter the truth of which he seriously discusses, then, according to the adopted definition of "real," the state of things which will be believed in that ultimate opinion is real.[[17]](#footnote-17)

Matters stand thus. A fully articulated foundationalism, one that faces squarely the problem that prompted it, faces the problem of induction, or more broadly, the problem of justifying the principles of reasoning (and for that matter, the incorrigibility of the foundations). A fully articulated fallibilism escapes that problem, since the truth of my claim that I know something grows out of whether the ideal community would accept my principles and starting points, and I may not be able to show that it would.[[18]](#footnote-18)

But as I said at the outset, I think the fallibilist faces a problem analogous to the (generalized) problem of induction. The problem of convergence can be stated as three nested questions:

1. How can convergence be defined and measured?
2. (assuming (a) has been answered) How do we know there will be convergence to some limit?
3. (assuming (b) has been answered) Why would there be convergence?

We can get some insight into the first question by thinking about theoretical terms of temporally successive theories. In what sense are theories converging? Laudan has argued that the history of science is replete with cases of theories that were successful during their time, but are now (i.e., our community's perspective) seen to be non-referential with respect to the central theoretical terms. He mentions such examples as 19th century aether theories, the crystalline spheres of ancient and medieval astronomy, the phlogiston theory, theories of spontaneous generation and so on.[[19]](#footnote-19)

But if historically successive theories may not even refer to the same things, how can we speak of convergence of belief? Convergence of belief in what? How can we tell of two communities are converging in belief, as opposed to talking about different things entirely?

Thus the question "how can convergence be defined and measured?" leads one into a hornets' nest of difficulties (about the nature of reference, of "approximate truth", of the incommensurability of theories, and so on). Perhaps one way to answer the question yet avoid the hornets' nest is to focus, not upon the convergence of theories, but instead the convergence of belief in theories. That is, one might shift focus from explaining how one theory can be said to be more approximately true than another (a syntactic/semantic issue) to explaining how one can identify how a community comes to agree to reject one theory and adopt another theory in any given domain (which is more of a pragmatic issue).

Even assuming that there can be some way to detect convergence, we still face the problem of uniqueness. Isn't it possible that even faced with all possible data, there might still be alternative theories that fit? Quine aptly speaks of the "under-determination of theories by data", and says of Peirce's theory:

Peirce was tempted to define truth outright in terms of scientific method, as the ideal theory which is approached as a limit when the (supposed) canons of scientific method are used unceasingly on continuing experience. But there is a lot wrong with Peirce's notion, besides its assumption of a final organon of scientific method and its appeal to an infinite process. There is a faulty use of numerical analogy in speaking of a limit of theories, since the notion of limit depends on that of "nearer than" which is defined for numbers and not for theories. And even if we by-pass such troubles..., still there is trouble in the imputation of uniqueness ("the ideal result"). For ... we have no reason to suppose that man's surface irritations even unto eternity admit of any one systematization that is scientifically better or simpler that all possible others. It seems likelier, if only on account of asymmetries or dualities, that countless alternative theories would be tied for first place. Scientific method is the way to truth, but it affords even in principle no unique definition of truth. Any so-called pragmatic definition of truth is doomed to failure equally.[[20]](#footnote-20)

Again, I don't mean to suggest that the difficulty here is an impossibility. We might argue that convergence would be reached on matters of beauty or elegance, if not on the basis of data alone. That is, even given that explanatory power and simplicity might not ensure uniqueness of final theory, perhaps there is some set of virtues that would, a set including beauty, consilience, and so on.

More cogently, Quine's objection ignores the fact that the limit is reached by an historical route. Even supposing that we looked at all data on the day of final judgment and decided that we could devise other theories, the community might have converged on one single theory just as a matter of fact. This agreed-upon theory would (we are imagining) account for all the data equally simply, elegantly, and so on as any other theory anybody could devise, so there would be no temptation to switch. Remember the difference in motivation between the foundationalist and the fallibilist: the fallibilist is not interested in skeptical doubts about his theory—so long as it works well, so long as there is no real reason to doubt it, he doesn't change it. And in such a scenario as we are imagining, there would be no such doubt.

The third difficulty seems like the least bothersome, but I suspect that it is in reality the most profound. This objection, you will recall, is that it is unclear why these would be convergence of opinion at all. Assume we can measure convergence, assume that such convergence would be convergence to but one point of view, still, *why would there be convergence at all?* You almost automatically say: because there is one real world out there that forces (or would force) eventual agreement. But on what basis do we know there is "one world" "out there"? An empiricist foundationalist would say "we just see it!", but (waiving the problems inherent in such a view) that is not really a reply the fallibilist can make. There is no certainly in perception, even if my community agrees with my perceptual claims. The only reply open to the fallibilist seems to be: the statement "there is one world out there" is true because a community of researchers would converge to agreement upon that claim.

When Descartes had finished *The Meditations*, he sent the work out to a number of philosophers for comment. Arnauld pointed out what has since come to be known as "the Cartesian Circle". Descartes argues early on that whatever we clearly and distinctly perceive is true, then uses this criterion to prove that God exists, and then proves that the criterion is correct because God's existence (somehow) guarantees it. This, as Arnauld recognized, is circular.

The fallibilist similarly faces a circle: we know that in the ideal community, there will be eventual convergence because there is a real world out there, and we know there is a real world out there because the ideal community will converge in belief that there is. It is not clear to me how the fallibilist can reply.

One reply to the problem of convergence might be to weaken fallibilism. If we call theses (1\*) — (4\*) weak fallibilism and theses (1\*) — (6\*) strong fallibilism, the argument might run as follows. Since the problem of convergence arises only when we move from weak fallibilism to strong, why not just stick to weak fallibilism? Granted, weak fallibilism isn't incompatible with skepticism, but we fallibilists don't have exaggerated worries about skepticism anyway. We leave the hyperbolic doubt to others.

In a curious way this argument is reminiscent of the argument for weak foundationalism, and is open to the same sort of reply: The weak theory of knowledge can't do the job for which the strong version was intended. In particular, the fallibilist theory was devised to shed light on the economics of research: why some questions are investigated rather than others, why some research programs and not others are undertaken, and so on. But the weak fallibilist has no basis for doing any normative work, he can only describe. This physical theory is preferred by physicists today, but I (the weak fallibilist) can't really say that that theory is any better or closer to the truth than its predecessors. This community checks for lung cancer by chest X-rays, that community by reading chicken guts. There is no right or wrong about the matter.

A close reading of one of the earlier quotes by Peirce suggests another reply to the dilemma of circularity or weakness. The start of the quote which reads:

... all the followers of science are animated by a cheerful hope that the processes of investigation if only pushed far enough, will give one certain solution to each question to which they apply it...

The phrase "are animated by a cheerful hope" originally read "are fully persuaded", indicating that Peirce weakened the claim that we are sure there is an ideal community to the claim that we hope there is one. We might accordingly say that we fallibilists don't know for certain that there is a single world out there that will force a unique convergence, but we have no real reason to doubt what seems so obvious to us, and until we do, we ought to keep that conviction. In short, perhaps the way to answer the problem of convergence is to be a more consistent fallibilist.

In sum, the weakened versions of foundationalism and fallibilism seem altogether ad hoc and uninteresting. Thus faced with a choice between strong foundationalism and strong fallibilism, I am inclined to choose fallibilism.[[21]](#footnote-21)

There are reasons for this choice. An empirical approach to logic and related matters (such as theory choice) allows the epistemologist to learn more from his colleagues in the cognitive sciences (cognitive psychology, history of science, artificial intelligence, and so on). Also, the fallibilist perspective is more conducive to a diachronic approach as Glymour's "bootstrapping" view of theory confirmation, and more generally to epistemological theories which stress learning and discovery, as opposed to the retrospective justification of the already learned.

Finally, there might just be a solution to the problem of convergence. There might not be such a solution, of course. But I take it as being an inescapable feature of the act of choosing a research program that there be is no guarantee of success.

1. Quite often foundationalists offer their epistemology more as a program for reconstructing our beliefs to achieve true knowledge than as an analysis of how our belief systems are in fact structured. [↑](#footnote-ref-1)
2. There are of course important differences between various concepts of certainty, incorrigibility, indubitability, and so on. Those differences will not be important in what follows. [↑](#footnote-ref-2)
3. Note that the skeptic here is *not* (as is sometimes charged) trying “to turn induction into deduction.” The skeptic does not demand that we show that the arguments which N sanctions. [↑](#footnote-ref-3)
4. Skyrms, Brian *Choice and Chance (2nd ed.)*, Belmont, CA: Dickinson Publishing Company, 1975. [↑](#footnote-ref-4)
5. *Ibid* pp. 41-47. [↑](#footnote-ref-5)
6. *Ibid* p. 46. [↑](#footnote-ref-6)
7. Haack, Susan “The Justification of Deduction,” *Mind* 85 (1976), pp. 54-62. [↑](#footnote-ref-7)
8. Haack, op. cit., p. 57. [↑](#footnote-ref-8)
9. This is a bit quick. The “” connective is not the same as the material implication symbol “”. “*p* ( *q* *p*)” is still valid in most, though not all, relevant logics. In any case, the point still remains that the ordinary language statement “if *p*, then if *q* then *p*” may well not be a law of logic. [↑](#footnote-ref-9)
10. Fumerton, R.A., “Induction and Reasoning to the Best Explanation” *Philosophy of Science* 47 (1980), pp. 589- 600. [↑](#footnote-ref-10)
11. Alston, William “Two Types of Foundationalism” *Journal of Philosophy* Vol. 78 No. 7 (1976), pp. 165-185. Alston does not commit himself to the view he calls “simple foundationalism” which is equivalent to the view I call “weak foundationalism”. See also James Van Cleve “Foundationalism, Epistemic Principles, and the Cartesian Circle” in *The Philosophical Review* 88, No.1 January 1979. [↑](#footnote-ref-11)
12. Or, at least, weak foundationalism involves the normative claim that knowledge *should* be based on the {*f*i}. [↑](#footnote-ref-12)
13. Strictly speaking, we might want to speak of “an ideal community.” But if there were two ideal communities, we could be led to contradictory ascriptions f knowledge unless both communities in fact were one (meaning, shared belief systems). In my argument, I assume that the law of non-contradiction is held to be true by the ideal community. [↑](#footnote-ref-13)
14. *Collected Works of Charles Sanders Peirce* Vol. V, Section 556. [↑](#footnote-ref-14)
15. *Ibid.,* Vol. V, Section 407-408. [↑](#footnote-ref-15)
16. *Ibid.,* Vol. II, Section 264-265. [↑](#footnote-ref-16)
17. *Ibid*., Vol. V Section 430. [↑](#footnote-ref-17)
18. But then again, maybe I could. Richard Feldman, in "Fallibilism and Knowing That One Knows" (*Philosophical Review* XC, No. 2 April 1981) argues that the KK thesis may be compatible with fallibilism. [↑](#footnote-ref-18)
19. Laudan, Larry "A Confutation of Convergent Realism" *Philosophy of Science* 48 (1981) p. 33. Note that Landon's concern is with convergent realism, rather than the convergence theory of truth, and note also that his objections are somewhat different from the one I am discussing here (viz., that convergence is hard to define). For a reply to Laudan, see Clyde Hardin and Alexander Rosenberg "In Defense of Convergent Realism" *Philosophy of Science* 49 (1982) pp. 604- 615. Laudan's counter response to Hardin and Rosenberg is to be found in *Philosophy of Science* 51 (1984) pp. 156-162. [↑](#footnote-ref-19)
20. W.V.O. Quine *Word and Object.* [↑](#footnote-ref-20)
21. Coherentists may resent me for not having mentioned their theory of knowledge as an alternative. But regarding the thesis of this paper, I think the coherentist will face the same sort of problems as the foundationalist and fallibilist/pragmaticist. The coherentist believes that beliefs are justified by being part of a coherent web. That is, beliefs don't usually require justification, but require it only when coherence demands. But, how can we define and measure coherence? How do we know that there would be one coherent web (as opposed to an infinity of webs)? And why would there be coherence anyway? [↑](#footnote-ref-21)