1. Conjectural Knowledge: My Solution of the Problem of Induction

The growth of unreason throughout the nineteenth century and what has passed of the twentieth is a natural sequel to Hume’s destruction of empiricism.

BERTRAND RUSSELL

Of course, I may be mistaken; but I think that I have solved a major philosophical problem: the problem of induction. (I must have reached the solution in 1927 or thereabouts.) This solution has been extremely fruitful, and it has enabled me to solve a good number of other philosophical problems.

However, few philosophers would support the thesis that I have solved the problem of induction. Few philosophers have taken the trouble to study—or even to criticize—my views on this problem, or have taken notice of the fact that I have done some work on it. Many books have been published quite recently on the subject which do not refer to any of my work, although most of them show signs of having been influenced by some very indirect echoes of my ideas; and those works which take notice of my ideas usually ascribe views to me which I have never held, or criticize me on the basis of straightforward misunderstandings or misreadings, or with invalid arguments. This chapter is an

---

1 I had earlier (in the winter of 1919–20) formulated and solved the problem of demarcation between science and non-science and I did not think it worth publishing. But after I had solved the problem of induction I discovered an interesting connection between the two problems. This made me think that the problem of demarcation was important. I started to work on the problem of induction in 1923, and I found the solution about 1927. See also the autobiographical remarks in Conjectures and Refutations (C. & R. for short), chapters 1 and 11.

This chapter was first published in Revue internationale de Philosophie, 25th année, no. 95–6, 1971, fasc. 1–2.
attempt to explain my views afresh, and in a way which contains
a full answer to my critics.

My first two publications on the problem of induction were
my note in Erkenntnis of 1933, in which I briefly presented my
formulation of the problem and my solution, and my Logik der
Forschung (L.d. F.) of 1934. The note and also the book were very
compressed. I expected, a little optimistically, that my readers
would find out, with the help of my few historical hints, why my
peculiar reformulation of the problem was decisive. It was, I think,
the fact that I reformulated the traditional philosophical prob-
lem which made its solution possible.

By the traditional philosophical problem of induction I mean some
formulation like the following (which I will call ‘Tr’):

Tr What is the justification for the belief that the future will
be (largely) like the past? Or, perhaps, What is the justification
for inductive inferences?

Formulations like these are wrongly put, for several reasons.
For example, the first assumes that the future will be like the past
—an assumption which I, for one, regard as mistaken, unless the
word ‘like’ is taken in a sense so flexible as to make the assump-
tion empty and innocuous. The second formulation assumes that
there are inductive inferences, and rules for drawing inductive
inferences, and this, again, is an assumption which should not
be made uncritically, and one which I also regard as mistaken.
Therefore I think that both formulations are simply uncritical,
and similar remarks would hold for many other formulations.
My main task will be, therefore, to formulate once more the
problem which I think lies behind what I have called the traditional
philosophical problem of induction.

The formulations which by now have become traditional are
historically of fairly recent date: they arise out of Hume’s criti-
cism of induction and its impact upon the commonsense theory
of knowledge.

I shall return to a more detailed discussion of the traditional
formulations after presenting, first, the commonsense view, next

---

1. Conjectural Knowledge

Hume’s view, and then my own reformulations and solutions of
the problem.

1. The Commonsense Problem of Induction

The commonsense theory of knowledge (which I have also
dubbed ‘the bucket theory of the mind’) is the theory most
famous in the form of the assertion that ‘there is nothing in our
intellect which has not entered it through the senses’. (I have
tried to show that this view was first formulated by Parmenides—in
a satirical vein: Most mortals have nothing in their erring
intellect unless it got there through their erring senses.)

However, we do have expectations, and we strongly believe in
certain regularities (laws of nature, theories). This leads to the
commonsense problem of induction (which I will call ‘Cs’):

Cs How can these expectations and beliefs have arisen?

The commonsense answer is: Through repeated observations
made in the past: we believe that the sun will rise tomorrow
because it has done so in the past.

In the commonsense view it is simply taken for granted
(without any problems being raised) that our belief in regulari-
ties is justified by those repeated observations which are respon-
sible for its genesis. (Genesis sum justification—both due to
repetition—is what philosophers since Aristotle and Cicero have
called ‘epagoge’ or ‘induction’.

2. Hume’s Two Problems of Induction

Hume was interested in the status of human knowledge or, as
he might have said, in the question of whether any of our beliefs
—and which of them—can be justified by sufficient reasons.

He raised two problems: a logical problem ($H_L$) and a psycho-
logical problem ($H_P$). One of the important points is that his
two answers to these two problems in some way clash with each
other.

---

* See my Conjectures and Refutations (C. & R.), Addendum 8 to the third edn.,
* Cicero, Topica, X. 45; cp. De inventione, Book I: xxxi. 51 to xxxiii. 61.
* See David Hume, Enquiry Concerning Human Understanding, ed. L. A. Selby-
Hume's logical problem is:

Hₐ: Are we justified in reasoning from [repeated] instances of which we have experience to other instances [conclusions] of which we have no experience?

Hume's answer to Hₐ is: No, however great the number of repetitions.

Hume also showed that the logical situation remained exactly the same if in Hₐ, the word 'probable' is inserted before 'conclusions', or if the words 'to other instances' are replaced by 'to the probability of other instances'.

Hume's psychological problem is: Hₚₐ, Why, nevertheless, do all reasonable people expect, and believe, that instances of which they have no experience will correspond to those of which they have experience? That is, why do we have expectations in which we have great confidence?

Hume's answer to Hₚₐ is: Because of 'custom or habit'; that is, because we are conditioned, by repetitions and by the mechanism of the association of ideas; a mechanism without which, Hume says, we could hardly survive.

3. Important Consequences of Hume's Results

By these results Hume himself—one of the most rational minds ever—was turned into a sceptic and, at the same time, into a believer: a believer in an irrationalist epistemology. His result that repetition has no power whatever as an argument, although it dominates our cognitive life or our 'understanding',

Hume, Treatise on Human Nature, ed. Selby-Bigge, Oxford, 1888, 1960, Book I, Part III, section vi, p. 91; Book I, Part III, section xii, p. 139. See also Kant, Prolegomena, pp. 14 ff., where he calls the problem of the existence of a priori valid statements 'Hume's problem'. To my knowledge I was the first to call the problem of induction 'Hume's problem'; though of course there may have been others. I did so in 'Ein Kriterium der empirischen Charakters theoretischer Systeme', Erkenntnis, 3, 1933, pp. 445 ff., and in L.d.F., section 4, p. 7, where I wrote: 'If, following Kant, we call the problem of induction "Hume's problem", we might call the problem of demarcation "Kant's problem". This very brief remark of mine (supported by a few remarks such as on p. 29 of L. Sc. D., that Kant took the problem of induction as a priori valid) contained hints of an important historical interpretation of the relationship between Kant, Hume, and the problem of induction. See also this volume, Chapter 5, pp. 85 ff. and p. 93, where these points are discussed more fully.

8 See Treatise, pp. 91, 139.

led him to the conclusion that argument or reason plays only a minor role in our understanding. Our 'knowledge' is unmasked as being not only of the nature of belief, but of rationally indefensible belief—of an irrational faith.

I hope that it will become clear from my next section, 4, and from sections 10 and 11, that no such irrationalist conclusion can be derived from my solution of the problem of induction.

Hume's conclusion was even more forcefully and desperately stated by Russell, in the chapter on Hume of his A History of Western Philosophy, published in 1946 (thirty-four years after his Problems of Philosophy, which contained a beautifully clear statement of the problem of induction without reference to Hume). Russell says about Hume's treatment of induction: 'Hume's philosophy . . . represents the bankruptcy of eighteenth-century reasonableness' and, 'It is therefore important to discover whether there is any answer to Hume within a philosophy that is wholly or mainly empirical. If not, there is no intellectual difference between sanity and insanity. The lunatic who believes that he is a poached egg is to be condemned solely on the ground that he is in a minority . . .'

Russell goes on to assert that if induction (or the principle of induction) is rejected, 'every attempt to arrive at general scientific laws from particular observations is fallacious, and Hume's scepticism is inescapable for an empiricist'.

Thus Russell stresses the clash between Hume's answer to Hₐ and (a) rationality, (b) empiricism, and (c) scientific procedures.

I hope that my discussions in sections 4 and 10 to 12 will show that all these clashes disappear if my solution of the problem of induction is accepted: there is no clash between my theory of non-induction and either rationality, or empiricism, or the procedures of science.

9 Since Hume, many disappointed inductivists have become irrationalists (just as have many disappointed Marxists).

10 Hume's name does not occur in chapter VI ('On Induction') of Russell's The Problems of Philosophy (1912 and many later reprints), and the nearest to a reference is in chapter VIII ('How A Priori Knowledge is Possible'), where Russell says of Hume that he inferred the far more doubtful proposition that nothing could be known a priori about the connection of cause and effect. No doubt, causal expectations have an inborn basis: they are psychologically a priori in the sense that they are prior to experience. But this does not mean that they are a priori valid. See C. & R., pp. 47 ff.

11 The quotations are from Bertrand Russell, A History of Western Philosophy, London, 1946, pp. 69 ff. (The italics are mine.)
4. My Way of Approaching the Problem of Induction

(1) I regard the distinction, implicit in Hume's treatment, between a logical and a psychological problem as of the utmost importance. But I do not think that Hume's view of what I am inclined to call 'logic' is satisfactory. He describes, clearly enough, processes of valid inference; but he looks upon these as 'rational' mental processes.

By contrast, one of my principal methods of approach, whenever logical problems are at stake, is to translate all the subjective or psychological terms, especially 'belief', etc., into objective terms. Thus, instead of speaking of a 'belief', I say, of a 'statement' or of an 'explanatory theory'; and instead of an 'impression', I speak of an 'observation statement' or of a 'test statement'; and instead of the 'justification of a belief', I speak of 'justification of the claim that a theory is true', etc.

This procedure of putting things into the objective or logical or 'formal' mode of speaking will be applied to $H_L$, but not to $H_P$; however:

(2) Once the logical problem, $H_L$, is solved, the solution is transferred to the psychological problem, $H_P$, on the basis of the following principle of transference: what is true in logic is true in psychology. (An analogous principle holds by and large for what is usually called 'scientific method' and also for the history of science: what is true in logic is true in scientific method and in the history of science.) This is admittedly a somewhat daring conjecture in the psychology of cognition or of thought processes.

(3) It will be clear that my principle of transference guarantees the elimination of Hume's irrationalism: if I can answer his main problem of induction, including $H_P$, without violating the principle of transference, then there can be no clash between logic and psychology, and therefore no conclusion that our understanding is irrational.

(4) Such a programme, together with Hume's solution of $H_L$, implies that more can be said about the logical relations between scientific theories and observations than is said in $H_L$.

(5) One of my main results is that, since Hume is right that there is no such thing as induction by repetition in logic, by the principle of transference there cannot be any such thing in psychology (or in scientific method, or in the history of science): the idea of induction by repetition must be due to an error—a kind of optical illusion. In brief: there is no such thing as induction by repetition.

5. The Logical Problem of Induction: Restatement and Solution

In accordance with what has just been said (point (1) of the preceding section 4), I have to restate Hume's $H_L$ in an objective or logical mode of speech.

To this end I replace Hume's 'instances of which we have experience' by 'test statements'—that is, singular statements describing observable events ('observation statements', or 'basic statements'); and 'instances of which we have no experience' by 'explanatory universal theories'.

I formulate Hume's logical problem of induction as follows:

$L_1$ Can the claim that an explanatory universal theory is true be justified by 'empirical reasons'; that is, by assuming the truth of certain test statements or observation statements (which, it may be said, are 'based on experience')?

My answer to the problem is the same as Hume's: No, it cannot; no number of true test statements would justify the claim that an explanatory universal theory is true.12

But there is a second logical problem, $L_2$, which is a generalization of $L_1$. It is obtained from $L_1$ merely by replacing the words 'is true' by the words 'is true or that it is false':

$L_2$ Can the claim that an explanatory universal theory is true or that it is false be justified by 'empirical reasons'; that is, can the assumption of the truth of test statements justify either the claim that a universal theory is true or the claim that it is false?

To this problem, my answer is positive: Yes, the assumption of the truth of test statements sometimes allows us to justify the claim that an explanatory universal theory is false.

This reply becomes very important if we reflect on the problem situation in which the problem of induction arises. I have in mind a situation in which we are faced with several explanatory theories which compete qua solutions of some problem of explanation—for example a scientific problem; and also with the fact that we have to, or at least wish to, choose between them. As we have seen, Russell says that without solving the problem of

12 An explanatory theory goes essentially beyond even an infinity of singular test statements; even a law of low universality does so.
induction, we could not decide between a (good) scientific theory and a (bad) obsession of a madman. Hume too had competing theories in mind. Suppose [he writes] a person . . . advances propositions, to which I do not assent, . . . that silver is more fusible than lead, or mercury heavier than gold. . . ."

This problem situation—that of choosing between several theories—suggests a third reformulation of the problem of induction:

$L_3$ Can a preference, with respect to truth or falsity, for some competing universal theories over others ever be justified by such 'empirical reasons'?  

In the light of my answer to $L_2$ the answer to $L_3$ becomes obvious: Yes; sometimes it can, if we are lucky. For it may happen that our test statements may refute some—but not all—of the competing theories; and since we are searching for a true theory, we shall prefer those whose falsity has not been established.

6. Comments on My Solution of the Logical Problem

(1) According to my reformulations, the central issue of the logical problem of induction is the validity (truth or falsity) of universal laws relative to some 'given' test statements. I do not raise the question, 'How do we decide the truth or falsity of test statements?', that is, of singular descriptions of observable events. The latter question should not, I suggest, be regarded as part of the problem of induction, since Hume's question was whether we are justified in reasoning from experienced to unexperienced 'instances'.

Neither Hume nor any other writer on the subject before me has to my knowledge moved on from here to the further questions: Can we take the 'experienced instances' for granted? And are they really prior to the theories? Although these further questions are some of those problems to which I was led by my solution of the problem of induction, they go beyond the original problem. (This is clear if we consider the kind of thing for which philosophers have been looking when trying to solve the problem of induction: if a 'principle of induction', permitting us to derive universal laws from singular statements, could be found, and its claim to truth defended, then the problem of induction would be regarded as solved.)

(2) $L_1$ is an attempt to translate Hume's problem into an objective mode of speech. The only difference is that Hume speaks of future (singular) instances of which we have no experience—that is, of expectations—while $L_1$ speaks of universal laws or theories. I have at least three reasons for this change. First, from a logical point of view, 'instances' are relative to some universal law (or at least to a statement function which could be universalized). Secondly, our usual method of reasoning from 'instances' to other 'instances' is with the help of universal theories. Thus we are led from Hume's problem to the problem of the validity of universal theories (their truth or falsehood). Thirdly, I wish, like Russell, to connect the problem of induction with the universal laws or theories of science.

(3) My negative answer to $L_1$ should be interpreted as meaning that we must regard all laws or theories as hypothetical or conjectural; that is, as guesses.

This view is by now fairly popular, but it took quite a time to reach this stage. It is, for example, explicitly combated in an otherwise excellent article of 1937 by Professor Gilbert Ryle. Ryle argues (p. 36) that it is wrong to say 'that all the general propositions of science . . . are mere hypotheses'; and he uses the term 'hypothesis' in exactly the same sense in which I have always used it and in which I am using it now: as a 'proposition . . . which is only conjectured to be true' (loc. cit.). He asserts against a thesis like mine: 'We are often sure, and warranted in being sure, of a law proposition' (p. 38). And he says that some general propositions are 'established': 'These are called 'laws'', and not 'hypotheses'.'

This view of Ryle's was indeed almost the 'established' standard at the time I wrote $L.d.F.$, and it is by no means dead. I first turned against it because of Einstein's theory of gravity: there never was a theory as well 'established' as Newton's, and it is unlikely that there ever will be one; but whatever one may think about the status of Einstein's theory, it certainly taught us to look at Newton's as a 'mere' hypothesis or conjecture.


15 See Mr. Stove's commencing remark in Australas. Journ. of Philos. 38, 1960, p. 179.
A second such case was the discovery by Urey in 1931 of deuterium and heavy water. At that time, water, hydrogen, and oxygen, were the substances best known to chemistry, and the atomic weights of hydrogen and oxygen formed the very standards of all chemical measurement. Here was a theory upon the truth of which every chemist would have staked his life, at least before Soddy's isotope conjecture in 1910, and in fact long afterwards. But it was here that a refutation was found by Urey (and thus a theory of Bohr's corroborated).

This led me to look more closely into other 'established laws' and especially into the three standard examples of the inductivists: 17

(a) that the sun will rise and set once in 24 hours (or approximately 90,000 pulse beats),
(b) that all men are mortal,
(c) that bread nourishes.

In all three cases I found that these established laws were actually refuted in the sense in which they were originally meant.

(a) The first was refuted when Pytheas of Marseilles discovered 'the frozen sea and the midnight sun'. The fact that (a) was intended to mean 'wherever you go, the sun will rise and set once in 24 hours' is shown by the utter disbelief with which his report was met, and by the fact that his report became the paradigm of all travellers' tales.

(b) The second—or rather, the Aristotelian theory on which it is based—was also refuted. The predicate 'mortal' is a bad translation from the Greek: thanatos means 'bound to die' or 'liable to die', rather than merely 'mortal', and (b) is part of Aristotle's theory that every generated creature is essentially bound to die after a period which, though its length is part of the creature's essence, will vary a little according to accidental circumstances. This theory was refuted by the discovery that bacteria are not bound to die, since multiplication by fission is not death, and later by the realization that living matter is not in general bound to decay and to die, although it seems that all forms can be killed by sufficiently drastic means. (Cancer cells, for example, can go on living.)

(c) The third—a favourite of Hume's—was refuted when people eating their daily bread died of ergotism, as happened in a catastrophic case in a French village not very long ago. Of course (c) originally meant that bread properly baked from flour properly prepared from wheat or corn, sown and harvested according to old-established practice, would nourish people rather than poison them. But they were poisoned.

Thus Hume's negative reply to \( H_F \) and my negative reply to \( L_4 \) are not merely far-fetched philosophical attitudes, as implied by Ryle, and by the commonsense theory of knowledge, but are based on very practical realities. In a vein similarly optimistic to that of Professor Ryle, Professor Strawson writes: 'If... there is a problem of induction, and... Hume posed it, it must be added that he solved it'—that is, by Hume's positive answer to \( H_F \), which Strawson seems to accept, describing it as follows: 'our acceptance of the “basic canons” [of induction]... is forced upon us by Nature... Reason is, and ought to be, the slave of the passions.' 18 (Hume had said: 'ought only to be'.)

I have not seen anything before which illustrates so well the quotation from Bertrand Russell's *A History of Western Philosophy*, p. 609, which I have chosen as a motto for the present discussion.

Yet it is clear that 'induction'—in the sense of a positive reply to \( H_F \) or \( L_4 \)—is inductively invalid, and even paradoxical. For a positive reply to \( L_4 \) implies that our scientific account of the world is roughly true. (With this I agree, in spite of my negative reply to \( L_4 \).) But from this it follows that we are very clever animals, precariously placed in a surrounding that differs greatly from almost every other place in the universe: animals that strive courageously to discover, by some method or other, the true regularities which rule the universe and thereby our surroundings. It is clear that whatever method we might use, our chances of finding true regularities are slim, and our theories will contain many mistakes which no mysterious 'canon of induction', whether basic or otherwise, will prevent us from committing. But this is just what my negative reply to \( L_4 \) says. Thus, since the positive reply entails its own negation, it must be false.

If anybody should wish to moralize about this story, he could

---

17 These examples, which I have often used in my lectures, have also been used in Chapter 2 (pp. 97 ff., and footnote 48). I apologize for the overlap, but these two chapters were written independently and I feel that they should be kept self-contained.

say: critical reason is better than passion, especially in matters touching on logic. But I am quite ready to admit that nothing will ever be achieved without a modicum of passion.

(4) $L_2$ is merely a generalization of $L_1$, and $L_3$ is merely an alternative formulation of $L_0$.

(5) My answer to $L_2$ and $L_3$ provides a clear answer to Russell's questions. For I can say: yes, at least some of the ravings of the lunatic can be regarded as refuted by experience; that is, by test statements. (Others may be non-testable and thereby distinguished from the theories of science; this raises the problem of demarcation.\textsuperscript{19})

(6) Most important, as I stressed in my first paper on the problem of induction, my answer to $L_3$ is in agreement with the following somewhat weak form of the principle of empiricism: Only 'experience' can help us to make up our minds about the truth or falsity of factual statements. For it turns out that, in view of $L_4$ and the answer to $L_4$, we can determine at most the falsity of theories; and this indeed can be done, in view of the answer to $L_4$.

(7) Similarly, there is no clash between my solution and the methods of science; on the contrary, we are led by it to the rudiments of a critical methodology.

(8) Not only does my solution throw much light upon the psychological problem of induction (see section 11, below), but it also elucidates the traditional formulations of the problem of induction and the reason for the weakness of these formulations. (See sections 12 and 13, below.)

(9) My formulations and my solutions of $L_1$, $L_2$, and $L_3$, fall entirely within the scope of deductive logic. What I show is that, generalizing Hume's problem, we can add to it $L_2$ and $L_3$, which allows us to formulate a somewhat more positive answer than the one to $L_4$. This is so because from the point of view of deductive logic there is an asymmetry between verification and falsification by experience. This leads to the purely logical distinction between hypotheses which have been refuted, and others which have not; and to the preference for the latter—if only from a theoretical point of view which makes them theoretically most interesting objects for further tests.

7. Preference for Theories and the Search for Truth

We have seen that our negative reply to $L_1$ means that all our theories remain guesses, conjectures, hypotheses. Once we have fully accepted this purely logical result, the question arises whether there can be purely rational arguments, including empirical arguments, for preferring some conjectures or hypotheses to others.

There may be various ways of looking at this question. I shall distinguish the point of view of the theoretician—the seeker for truth, and especially for true explanatory theories—from that of the practical man of action; that is, I will distinguish between theoretical preference and pragmatic preference. In this section and the next I shall be concerned only with theoretical preference and the quest for truth. Pragmatic preference and the problem of 'reliability' will be discussed in the next section but one.

The theoretician, I will assume, is essentially interested in truth, and especially in finding true theories. But when he has fully digested the fact that we can never justify empirically—that is, by test statements—the claim that a scientific theory is true, and that we are therefore at best always faced with the question of preferring, tentatively, some guesses to others, then he may consider, from the point of view of a seeker for true theories, the questions: What principles of preference should we adopt? Are some theories 'better' than others?

These questions give rise to the following considerations.

(1) It is clear that the question of preference will arise mainly, and perhaps even solely, with respect to a set of competing theories; that is, theories which are offered as solutions to the same problems. (See also point (8) below.)

(2) The theoretician who is interested in truth must also be interested in falsity, because finding that a statement is false is the same as finding that its negation is true. Thus the refutation of a theory will always be of theoretical interest. But the negation of an explanatory theory is not, in its turn, an explanatory
theory (nor has it as a rule the ‘empirical character’ of the test statement from which it is derived). Interesting as it is, it does not satisfy the theoretician’s interest in finding true explanatory theories.

(3) If the theoretician pursues this interest, then finding where a theory breaks down, apart from giving theoretically interesting information, poses an important new problem for any new explanatory theory. Any new theory will not only have to succeed where its refuted predecessor succeeded, but it will also have to succeed where its predecessor failed; that is, where it was refuted. If the new theory succeeds in both, it will at any rate be more successful and therefore ‘better’ than the old one.

(4) Moreover, assuming that this new theory is not refuted at the time \( t \) by a new test, it will, at any rate at the time \( t \), be ‘better’ in yet another sense than the refuted theory. For it will not only explain all that the refuted theory explained, and more, but it will also have to be regarded as possibly true, since at the time \( t \) it has not been shown to be false.

(5) Yet the theoretician will value such a new theory not only because of its success, and its being perhaps a true theory, but also because it may perhaps be false: it is interesting as an object of further tests; that is, of new attempted refutations which, if successful, establish not only a new negation of a theory, but with it a new theoretical problem for the next theory.

We can sum up points (1) to (5) as follows.

The theoretician will for several reasons be interested in non-refuted theories, especially because some of them may be true. He will prefer a non-refuted theory to a refuted one, provided it explains the successes and failures of the refuted theory.

(6) But the new theory may, like all non-refuted theories, be false. The theoretician will therefore try his best to detect any false theory among the set of non-refuted competitors; he will try to ‘catch’ it. That is, he will, with respect to any given non-refuted theory, try to think of cases or situations in which it is likely to fail, if it is false. Thus he will try to construct severe tests, and crucial test situations. This will amount to the construction of a falsifying law; that is, a law which may perhaps be of such a low level of universality that it may not be able to explain the successes of the theory to be tested, but which will, nevertheless, suggest a crucial experiment: an experiment which may refute,

depending on its outcome, either the theory to be tested or the falsifying theory.

(7) By this method of elimination, we may hit upon a true theory. But in no case can the method establish its truth, even if it is true; for the number of possibly true theories remains infinite, at any time and after any number of crucial tests. (This is another way of stating Hume’s negative result.) The actually proposed theories will, of course, be finite in number; and it may well happen that we refute all of them, and cannot think of a new one.

On the other hand, among the theories actually proposed there may be more than one which is not refuted at a time \( t \), so that we may not know which of these we ought to prefer. But if at a time \( t \) a plurality of theories continues to compete in this way, the theoretician will try to discover how crucial experiments can be designed between them; that is, experiments which could falsify and thus eliminate some of the competing theories.

(8) The procedure described may lead to a set of theories which are ‘competing’ in the sense that they offer solutions to at least some common problems, although each offers in addition solutions to some problems which it does not share with the others. For although we demand of a new theory that it solves those problems which its predecessor solved and those which it failed to solve, it may of course always happen that two or more new competing theories are proposed such that each of them satisfies these demands and in addition solves some problems which the others do not solve.

(9) At any time \( t \), the theoretician will be especially interested in finding the best testable of the competing theories in order to submit it to new tests. I have shown that this will at the same time be the one with the greatest information content and the greatest explanatory power. It will be the theory most worthy of being submitted to new tests, in brief ‘the best’ of the theories competing at time \( t \). If it survives its tests, it will also be the best tested of all the theories so far considered, including all its predecessors.

(10) In what has just been said about ‘the best’ theory it is assumed that a good theory is not \textit{ad hoc}. The ideas of \textit{adhocness} and its opposite, which may perhaps be termed ‘boldness’, are very important. \textit{Ad hoc} explanations are explanations which are
approximation towards this new theory. Nothing, of course, can make sure that for every theory which has been falsified we shall find a 'better' successor, or a better approximation—one that satisfies these demands. There is no assurance that we shall be able to make progress towards better theories.

(13) Two further points may be added here. One is that what has been said so far belongs, as it were, to purely deductive logic—the logic within which $L_1$, $L_2$, and $L_3$ were posed. Yet in trying to apply this to practical situations arising in science, we come up against problems of a different kind. For example, the relationship between test statements and theories may not be as clear-cut as is here assumed; or the test statements themselves may be criticized. This is the kind of problem which always arises if we wish to apply pure logic to any lifelike situation. In connection with science it leads to what I have called methodological rules, the rules of critical discussion.

The other point is that these rules may be regarded as subject to the general aim of rational discussion, which is to get nearer to the truth.

8. Corroboration: The Merits of Improbability

(1) My theory of preference has nothing to do with a preference for the 'more probable' hypothesis. On the contrary, I have shown that the testability of a theory increases and decreases with its informative content and therefore with its improbability (in the sense of the calculus of probability). Thus the 'better' or 'preferable' hypothesis will, more often than not, be the more improbable one. (But it is a mistake to say, as does John C. Harsanyi, that I have ever proposed an 'improbability criterion for the choice of scientific hypotheses'; see not only do I have no general 'criterion', but it happens quite often that I cannot prefer the logically 'better' and more improbable hypothesis, since someone has succeeded in refuting it experimentally.) This result has of course been regarded as perverse by many, but my main arguments are very simple (content = improbability), and they have recently been accepted even by some proponents

---

20 See, especially, 'Naturgesetze und theoretische Systeme', in Gesammelte Werke, ed. S. Menger, Innsbruck, 1940, pp. 43 ff., and 'The Aim of Science', Ratio, 1, 1957, now respectively the Appendix and ch. 5 below.

of inductivism and of a probabilistic theory of induction, such as Carnap.\(^{23}\)

(2) I originally introduced the idea of *corroboration*, or ‘degree of corroboration’, with the aim of showing clearly that every probabilistic theory of preference (and therefore every probabilistic theory of induction) is absurd.

By the degree of corroboration of a theory I mean a concise report evaluating the state (at a certain time \(t\)) of the critical discussion of a theory, with respect to the way it solves its problems; its degree of testability; the severity of tests it has undergone; and the way it has stood up to these tests. Corroboration (or degree of corroboration) is thus an evaluating *report of past performance*. Like preference, it is essentially comparative: in general, one can only say that the theory \(A\) has a higher (or lower) degree of corroboration than a competing theory \(B\), in the light of the critical discussion, which includes testing, *up to some time \(t\)*. Being a report of past performance only, it has to do with a situation which may lead us to prefer some theories to others. *But it says nothing whatever about future performance, or about the ‘reliability’ of a theory.* (Of course this would in no way be affected should anybody succeed in showing that, in certain very special cases, my or someone else’s formulae for the degree of corroboration can be given a numerical interpretation.\(^{24}\)

The main purpose of the formulae which I proposed as definitions for the degree of corroboration was to show that, in many cases, the more improvable (improbable in the sense of the calculus of probability) hypothesis is preferable, and to show clearly in which cases this holds and in which it does not hold. In this way, I could show that *preference cannot be a probability in the sense of the calculus of probability*. Of course, one may call the preferable theory the more ‘probable’ one: *words do not matter*, as long as one is not misled by them.

To sum up: We can sometimes say of two competing theories,


\(^{24}\) It seems to me that Professor Lakatos suspects that the actual contribution of numbers to my degree of corroboration, if possible, would render my theory inductivist in the sense of a probabilistic theory of induction. I see no reason whatever why this should be so. Cp. pp. 410-12 of *The Problem of Inductive Logic*, I. Lakatos and A. Musgrave (eds.), North Holland, Amsterdam, 1968. (Added in proofs: I am glad to learn that I have misunderstood the passage.)

\(A\) and \(B\), that in the light of the state of the critical discussion at the time \(t\), and the empirical evidence (test statements) available at the discussion, the theory \(A\) is preferable to, or better corroborated than, the theory \(B\).

Obviously, the degree of corroboration at the time \(t\) (which is a statement about preferability at the time \(t\)) says nothing about the future—for example, about the degree of corroboration at a time later than \(t\). It is just a report about the state of discussion at the time \(t\), concerning the logical and empirical preferability of the competing theories.

(3) I must emphasize this, because the following passage of my *Logic of Scientific Discovery* has been interpreted—or rather misinterpreted—as showing that I was using corroboration as an index of the *future* performance of a theory: ‘Instead of discussing the “probability” of a hypothesis we should try to assess what tests, what trials, it has withstood; that is, we should try to assess how far it has been able to prove its fitness to survive by standing up to tests. In brief, we should try to assess how far it has been “corroborated”’.\(^{25}\)

Some people thought\(^{26}\) that the phrase ‘proves its fitness to survive’ shows that I had here intended to speak of a fitness to survive in the future, to stand up to future tests. I am sorry if I have misled anybody, but I can only say that it was not I who mixed the Darwinian metaphor. Nobody expects that a species which has survived in the past will therefore survive in the future: all the species which ever failed to survive some period of time \(t\) have survived up to that time \(t\). It would be absurd to suggest that Darwinian survival involves, somehow, an expectation that every species that has so far survived will continue to survive. (Who would say that the expectation for our own species to survive is very high?)

(4) It may perhaps be useful to add here a point about the degree of corroboration of a statement \(s\) which belongs to a theory \(T\), or follows logically from it, but is logically much weaker than the theory \(T\).

Such a statement \(s\) will have less informative content than the theory \(T\). This means that \(s\), and the deductive system \(S\) of all those statements which follow from \(s\), will be less testable and less corroborable than \(T\). But if \(T\) has been well tested, then we


\(^{26}\) See *Mind*, New Series, 69, 1960, p. 100.
can say that its high degree of corroboration applies to all the statements which are entailed by it, and therefore to \( s \) and \( S \), even though \( s \) could, on its own, never attain as high a degree of corroboration as it can as part of \( T \) or relative to \( T \).

This rule may be supported by the simple consideration that the degree of corroboration is a means of stating preference with respect to truth. But if we prefer \( T \) with respect to its claim to truth, then we have to prefer with it all its consequences, since if \( T \) is true, so must be all its consequences, even though they can be less well tested separately.

Thus I assert that with the corroboration of Newton's theory, and the description of the earth as a rotating planet, the degree of corroboration of the statement \( s \) 'The sun rises in Rome once in every twenty-four hours' has greatly increased. For, on its own, \( s \) is not very well testable; but Newton's theory, and the theory of the rotation of the earth are well testable. And if these are true, \( s \) will be true also.

A statement \( s \) which is derivable from a well-tested theory \( T \) will, so far as it is regarded as part of \( T \), have the degree of corroboration of \( T \); and if \( s \) is derivable not from \( T \) but from the conjunction of two theories, say \( T_1 \) and \( T_2 \), it will qua part of two theories have the same degree of corroboration as the less well tested of these two theories. Yet \( s \) taken by itself may have a very low degree of corroboration.

(5) The fundamental difference between my approach and the approach for which I long ago introduced the label 'inductivist' is that I lay stress on negative arguments, such as negative instances or counter-examples, refutations, and attempted refutations—in short, criticism—while the inductivist lays stress on 'positive instances', from which he draws 'non-demonstrative inferences', and which he hopes will guarantee the 'reliability' of the conclusions of these inferences. In my view, all that can possibly be 'positive' in our scientific knowledge is positive only in so far as certain theories are, at a certain moment of time, preferred to others in the light of our critical discussion which consists of attempted refutations, including empirical tests. Thus even what may be called 'positive' is so only with respect to negative methods.


This negative approach clarifies many points, for example the difficulties encountered in explaining satisfactorily what is a 'positive instance' or a 'supporting instance' of a law.

9. Pragmatic Preference

So far I have discussed why the theoretician's preference—if he has any—will be for the 'better', that is, more testable, theory, and for the better tested one. Of course, the theoretician may not have any preference: he may be discouraged by Hume's, and my, 'sceptical' solution to the problems \( H_2 \) and \( L_2 \); he may say that, if he cannot make sure of finding the true theory among the competing theories, he is not interested in any method like the one described—not even if the method makes it reasonably certain that, if a true theory should be among the theories proposed, it will be among the surviving, the preferred, the corroborated ones. Yet a more sanguine or more curious 'pure' theoretician may well be encouraged, by our analysis, to propose again and again new competing theories in the hope that one of them may be true—even if we shall never be able to make sure of any one that it is true.

Thus the pure theoretician has more than one way of action open to him; and he will choose a method such as the method of trial and the elimination of error only if his curiosity exceeds his disappointment at the unavoidable uncertainty and incompleteness of all our endeavours.

It is different with him qua man of practical action. For a man of practical action has always to choose between some more or less definite alternatives, since even inaction is a kind of action.

But every action presupposes a set of expectations; that is, of theories about the world. Which theory shall the man of action choose? Is there such a thing as a rational choice?

This leads us to the pragmatic problems of induction:

\[ Pr_1 \] Upon which theory should we rely for practical action, from a rational point of view?

\[ Pr_2 \] Which theory should we prefer for practical action, from a rational point of view?

My answer to \( Pr_1 \) is: From a rational point of view, we should not 'rely' on any theory, for no theory has been shown to be true, or can be shown to be true.
My answer to $P_{r_2}$ is: But we should prefer as basis for action the best-tested theory.

In other words, there is no ‘absolute reliance’; but since we have to choose, it will be ‘rational’ to choose the best-tested theory. This will be ‘rational’ in the most obvious sense of the word known to me: the best-tested theory is the one which, in the light of our critical discussion, appears to be the best so far, and I do not know of anything more ‘rational’ than a well-conducted critical discussion.

Of course, in choosing the best-tested theory as a basis for action, we ‘rely’ on it, in some sense of the word. It may therefore even be described as the most ‘reliable’ theory available, in some sense of this term. Yet this does not say that it is ‘reliable’.

It is not ‘reliable’ at least in the sense that we shall always do well, even in practical action, to foresee the possibility that something may go wrong with our expectations.

But it is not merely this trivial caution which we must derive from our negative reply to $L_1$ and $P_{r_1}$. Rather, it is of the utmost importance for the understanding of the whole problem, and especially of what I have called the traditional problem, that in spite of the ‘rationality’ of choosing the best-tested theory as a basis of action, this choice is not ‘rational’ in the sense that it is based upon good reasons for expecting that it will in practice be a successful choice; there can be no good reasons in this sense, and this is precisely Hume’s result. (In this our answers to $H_L$, $L_1$, and $P_{r_1}$ all agree.) On the contrary, even if our physical theories should be true, it is perfectly possible that the world as we know it, with all its pragmatically relevant regularities, may completely disintegrate in the next second. This should be obvious to anybody today; but I said so before Hiroshima: there are infinitely many possibilities of local, partial, or total disaster.

From a pragmatic point of view, however, most of these possibilities are obviously not worth bothering about because we cannot do anything about them: they are beyond the realm of action. (I do not, of course, include atomic war among those disasters which are beyond the realm of human action, although most of us think in this way, just because most of us cannot do more about it than about an act of God.)

All this would hold even if we could be certain that our physical and biological theories were true. But we do not know it. On the contrary, we have reason to suspect even the best of them; and this adds, of course, further infinities to the infinite possibilities of disaster.

It is this kind of consideration which makes Hume’s and my own negative reply so important. For we can now see very clearly why we must beware lest our theory of knowledge proves too much. More precisely, no theory of knowledge should attempt to explain why we are successful in our attempts to explain things.

Even if we assume that we have been successful—that our physical theories are true—we can learn from our cosmology how infinitely improbable this success is: our theories tell us that the world is almost completely empty, and that empty space is filled with chaotic radiation. And almost all places which are not empty are occupied either by chaotic dust, or by gases, or by very hot stars—all these in conditions which seem to make the application of any method of acquiring physical knowledge locally impossible.

To sum up, there are many worlds, possible and actual worlds, in which a search for knowledge and for regularities would fail. And even in the world as we actually know it from the sciences, the occurrence of conditions under which life, and a search for knowledge, could arise—and succeed—seems to be almost infinitely improbable. Moreover, it seems that if ever such conditions should appear, they would be bound to disappear again, after a time which, cosmologically speaking, is very short.

10. Background to My Restatement of Hume’s Psychological Problem of Induction

Historically, I found my new solution to Hume’s psychological problem of induction before my solution to the logical problem: it was here that I first noticed that induction—the formation of a belief by repetition—is a myth. It was first in animals and children, but later also in adults, that I observed the immensely powerful need for regularity—the need which makes them seek for regularities; which makes them sometimes experience regularities even where there are none; which makes them cling to their expectations dogmatically; and which makes them unhappy and may drive them to despair and to the verge
of madness if certain assumed regularities break down. When Kant said that our intellect imposes its laws upon nature, he was right—except that he did not notice how often our intellect fails in the attempt: the regularities we try to impose are psychologically a priori, but there is not the slightest reason to assume that they are a priori valid, as Kant thought. The need to try to impose such regularities upon our environment is, clearly, in-born, and based on drives, or instincts. There is the general need for a world that conforms to our expectations; and there are many more specific needs, for example the need for regular social response, or the need for learning a language with rules for descriptive (and other) statements. This led me first to the conclusion that expectations may arise without, or before, any repetition; and later to a logical analysis which showed that they could not arise otherwise because repetition presupposes similarity, and similarity presupposes a point of view—a theory, or an expectation.

Thus I decided that Hume’s inductive theory of the formation of beliefs could not possibly be true, for logical reasons. This led me to see that logical considerations may be transferred to psychological considerations; and it led me further to the heuristic conjecture that, quite generally, what holds in logic also holds—provided it is properly transferred—in psychology. (This heuristic principle is what I now call the ‘principle of transference.’) I suppose it was largely this result which made me give up psychology and turn to the logic of discovery.

Quite apart from this, I felt that psychology should be regarded as a biological discipline, and especially that any psychological theory of the acquisition of knowledge should be so regarded.

Now if we transfer to human and animal psychology that method of preference which is the result of our solution of $L_3$, we arrive, clearly, at the well-known method of trial and error-elimination: the various trials correspond to the formation of competing hypotheses; and the elimination of error corresponds to the elimination or refutation of theories by way of tests.

This led me to the formulation: the main difference between Einstein and an amoeba (as described by Jennings⁹) is that

Einstein consciously seeks for error elimination. He tries to kill his theories: he is consciously critical of his theories which, for this reason, he tries to formulate sharply rather than vaguely. But the amoeba cannot be critical vis-à-vis its expectations or hypotheses; it cannot be critical because it cannot face its hypotheses: they are part of it. (Only objective knowledge is criticizable: subjective knowledge becomes criticizable only when it becomes objective. And it becomes objective when we say what we think; and even more so when we write it down, or print it.)

It is clear that the method of trial and error-elimination is largely based upon inborn instincts. And it is clear that some of these instincts are linked with that vague phenomenon called by some philosophers ‘belief’.

I used to take pride in the fact that I am not a belief philosopher: I am primarily interested in ideas, in theories, and I find it comparatively unimportant whether or not anybody ‘believes’ in them. And I suspect that the interest of philosophers in belief results from that mistaken philosophy which I call ‘inducivism’. They are theorists of knowledge, and starting from subjective experiences they fail to distinguish between objective and subjective knowledge. This leads them to believe in belief as the genus of which knowledge is a species (‘justification’ or perhaps a ‘criterion of truth’ such as clarity and distinctness, or vivacity,³⁹ or ‘sufficient reason’, providing the specific difference).

This is why, like E. M. Forster, I do not believe in belief.

But there are other reasons, and more important ones, for being wary concerning belief. I am quite ready to admit that there exist some psychological states which may be called ‘expectations’, and that there are shades of expectations, from the lively expectation of a dog which is about to be taken for a walk, to the almost non-existent expectation of a schoolboy who knows, but does not really believe, that if only he lives long enough, he will one day be an old man. But it is questionable whether the word ‘belief’ is used by philosophers to describe psychological states in this sense. It seems that they more often use it to denote not momentary states but what may be called ‘settled’ beliefs, including those countless unconscious expectations which make up our horizon of expectations. It is a far cry

³⁹ See Hume, Treatise, p. 265.
from these to formulated hypotheses, and therefore also to statements of the form 'I believe that ...'.

Now almost all such formulated statements can be considered critically; and the psychological states which result from a critical consideration seem to me very different indeed from an unconscious expectation. Thus even a 'settled' belief changes when it is formulated, and again after it has been formulated. If the result of its critical consideration is 'acceptance', it can range from that fanatical acceptance which attempts to suppress one's doubts and scruples to that tentative acceptance which is ready for reconsideration and revision at a moment's notice, and which may even be linked with an active search for refutations.

I do not think that such distinctions between different 'beliefs' are of any interest for my own objectivist theory of knowledge; but they ought to be interesting for anybody who takes the psychological problem of induction seriously—which I do not.

11. Restatement of the Psychological Problem of Induction

For the reasons just explained, I do not regard the psychological problem of induction as part of my own (objectivist) theory of knowledge. But I think that the principle of transference suggests the following problems and answers.

$P_1$ If we look at a theory critically, from the point of view of sufficient evidence rather than from any pragmatic point of view, do we always have the feeling of complete assurance or certainty of its truth, even with respect to the best-tested theories, such as that the sun rises every day?

I think the answer here is: No. I suggest that the feeling of certainty—the strong belief—which Hume tried to explain was a pragmatic belief; something closely connected with action and the choice between alternatives, or else with our instinctive need for, and expectation of, regularities. But if we assume that we are in a position to reflect on the evidence, and what it permits us to assert, then we shall have to admit that the sun may not rise tomorrow over London after all—for example because the sun may explode within the next half-hour, so that there will be no tomorrow. Of course we shall not consider this possibility 'seriously'—that is, pragmatically—because it does not suggest any possible action: we just can't do anything about it.

Thus we are led to consider our pragmatic beliefs. And these can be very strong indeed. We may ask:

$P_2$ Are those 'strong pragmatic beliefs' which we all hold, such as the belief that there will be a tomorrow, the irrational results of repetition?

My reply is: No. The repetition theory is untenable anyway. These beliefs are partly inborn, partly modifications of inborn beliefs resulting from the method of trial and error-elimination. But this method is perfectly 'rational' since it corresponds precisely to that method of preference whose rationality has been discussed. More especially, a pragmatic belief in the results of science is not irrational, because there is nothing more 'rational' than the method of critical discussion, which is the method of science. And although it would be irrational to accept any of its results as certain, there is nothing 'better' when it comes to practical action: there is no alternative method which might be said to be more rational.

12. The Traditional Problem of Induction and the Invalidity of all Principles or Rules of Induction

I now return to what I call the traditional philosophical problem of induction.

What I call by this name is, I suggest, the result of seeing the commonsense view of induction by repetition challenged by Hume, without taking the challenge quite as seriously as it should be taken. Even Hume, after all, remained an inductivist; thus not every inductivist challenged by Hume can be expected to see that Hume's challenge is one to inductivism.

The fundamental schema of the traditional problem may be stated in various ways, for example:

$T_1$ How can induction be justified (in spite of Hume)?

$T_2$ How can a principle of induction (that is, a non-logical principle justifying induction) be justified?

$T_3$ How can one justify a particular principle of induction, such as 'the future will be like the past', or perhaps the so-called 'principle of the uniformity of nature'?

As I briefly indicated in my Logik der Forschung, I think that Kant's problem 'How can synthetic statements be valid a priori?'
was an attempt to generalize $T_1$ or $T_2$. This is why I regard Russell as a Kantian, at least in some of his phases, for he tried to find a solution for $T_2$ by some a priori justification. In the Problems of Philosophy, for example, Russell’s formulation of $T_2$ was: ‘... what sort of general belief would suffice, if true, to justify the judgement that the sun will rise tomorrow...?’

From my point of view, all these problems are badly formulated. (And so also are the probabilistic versions such as the one implicit in Thomas Reid’s principle of induction, ‘What is to be will probably be like what has been in similar circumstances.’) Their authors do not take Hume’s logical criticism sufficiently seriously; and they never seriously consider the possibility that we can, and must, do without induction by repetition, and that we actually manage without it.

It seems to me that all the objections to my theory which I know of approach it with the question of whether my theory has solved the traditional problem of induction—that is, whether I have justified inductive inference.

Of course I have not. From this my critics deduce that I have failed to solve Hume’s problem of induction.

It is, among other reasons, especially for the reason stated in section 9 that the traditional formulations of the principle of induction have to be rejected. For they all assume not only that our quest for knowledge has been successful, but also that we should be able to explain why it is successful.

However, even on the assumption (which I share) that our quest for knowledge has been very successful so far, and that we know something of our universe, this success becomes miraculously improbable, and therefore inexplicable; for an appeal to an endless series of improbable accidents is not an explanation. (The best we can do, I suppose, is to investigate the almost incredible evolutionary history of these accidents, from the making of the elements to the making of the organisms.)

Once this has been seen, not only Hume’s thesis that an appeal to probability cannot change the reply to $H_1$ (and therefore to $L_1$ and $P_{\text{true}}$) becomes perfectly obvious, but also the invalidity of any ‘principle of induction’.

The idea of a principle of induction is that of a statement—to be regarded as a metaphysical principle, or as valid a priori, or as probable, or perhaps as a mere conjecture—which, if true, would give good reasons for our reliance upon regularities. If by ‘reliance’ is meant merely pragmatic reliance, in the sense of $P_{\text{true}}$, upon the rationality of our theoretical preferences, then clearly no principle of induction is needed: we do not need to rely on regularities—that is, on the truth of theories—to justify this preference. If, on the other hand, ‘reliance’ in the sense of $P_{\text{true}}$ is intended, then any such principle of induction would simply be false. Indeed in the following sense it would even be paradoxical. It would entitle us to rely on science; whereas today’s science tells us that only under very special and improbable conditions can situations arise in which regularities, or instances of regularities, can be observed. In fact, science tells us, such conditions occur hardly anywhere in the universe, and if they occur somewhere (on earth, say) they are likely to occur for periods which will be short from a cosmological point of view.

Clearly this criticism applies not only to any principle which would justify inductive inference based on repetition, but also to any principle which would justify ‘reliance’, in the sense of $P_{\text{true}}$, on the method of trial and error-elimination, or on any other conceivable method.

13. Beyond the Problems of Induction and Demarcation

My solution of the problem of induction occurred to me a considerable time after I had solved, at least to my own satisfaction, the problem of demarcation (the demarcation between empirical science and pseudoscience, especially metaphysics).

Only after the solution of the problem of induction did I regard the problem of demarcation as objectively important, for I had suspected it of giving merely a definition of science. This seemed to me of doubtful significance (owing perhaps to my negative attitude towards definitions), even though I had found it very helpful for clarifying my attitude towards science and pseudoscience.

I saw that what has to be given up is the quest for justification, in the sense of the justification of the claim that a theory is true. All theories are hypotheses; all may be overthrown.

On the other hand, I was very far from suggesting that we give up the search for truth: our critical discussions of theories are dominated by the idea of finding a true (and powerful) explanatory theory; and we do justify our preferences by an appeal
to the idea of truth: truth plays the role of a regulative idea. *We test for truth*, by eliminating falsehood. That we cannot give a justification—or sufficient reasons—for our guesses does not mean that we may not have guessed the truth; some of our hypotheses may well be true.²¹

The realization that all knowledge is hypothetical leads to the rejection of the ‘principle of sufficient reason’ in the form ‘that a reason can be given for every truth’ (Leibniz) or in the stronger form which we find in Berkeley and Hume who both suggest that it is a sufficient reason for unbelief if we ‘see no [sufficient] reason for believing’.²²

Once I had solved the problem of induction, and realized its close connection with the problem of demarcation, interesting new problems and new solutions arose in rapid succession.

First of all I soon realized that the problem of demarcation and my solution, as stated above, were a bit formal and unrealistic: *empirical refutations could always be avoided*. It was always possible to ‘immunize’ any theory against criticism. (This excellent expression which, I think, should replace my terms ‘conventionalist stratagem’ and ‘conventionalist twist’ is due to Hans Albert.)

Thus I was led to the idea of *methodological rules* and of the fundamental importance of a *critical approach*; that is, of an approach which avoided the policy of immunizing our theories against refutation.

But at the same time, I also realized the opposite: the value of a *dogmatic* attitude: somebody had to defend a theory against criticism, or it would succumb too easily, and before it had been able to make its contributions to the growth of science.

The next step was the application of the critical approach to the test statements, the ‘empirical basis’: I stressed the conjectural and theoretical character of all observations, and all observation statements.

This led me to the view that all languages are theory-impregnated; which meant, of course, a radical revision of empiricism. It also made me look upon the critical attitude as characteristic of the rational attitude; and it led me to see the significance of the argumentative (or critical) function of language; to the idea of deductive logic as the organon of criticism, and to stressing the retransmission of falsity from the conclusion to the premises (a corollary of the transmission of truth from the premises to the conclusion). And it further led me to realize that only a *formulated* theory (in contradistinction to a believed theory) can be objective, and to the idea that it is this formulation or objectivity that makes criticism possible; and so to my theory of a ‘third world’ (or, as Sir John Eccles prefers to call it, ‘world 3’).²³

These are just a few of the many problems to which the new approach gave rise. There are other problems which are of a more technical character, such as the many problems connected with probability theory, including its role in quantum theory, and the connection between my theory of preference and Darwin’s theory of natural selection.


²² Berkeley, *Three Dialogues Between Hylas and Philonous*, second Dialogue: “It is to me a sufficient reason not to believe... if I see no reason for believing.” For Hume, see C. III B., p. 21 (where the *Enquiry Concerning Human Understanding*, Section V, Part I is quoted).