18

The Nature of Science and the Validity of Psychotherapy

DAVID H. MALAN, M.D,

Since there will be no lack of evidence in the later pages of this essay, I shall start by asking a number of questions and answering them— most unscientifically—without evidence.

The first question is, What is the truth about the effectiveness of dynamic psychotherapy?

More than twenty years of systematic follow-up work have convinced me of the following four-part answer:

1. On the whole, dynamic psychotherapy as usually practiced is ineffective.
2. On the other hand, in certain patients dynamic psychotherapy is extremely effective.
3. This effect occurs through the operation of specific factors, which can be identified and have been understood for decades.
4. But such patients represent a very small proportion of the general psychotherapeutic population.

The second question is, What is the truth about "spontaneous remission"?

My answer to this is that true improvements do occur without therapy and that the mechanisms mostly involve normal maturation interacting with life experience.

These statements together account at once for the fact that so many controlled studies of psychotherapy have given results that are positive but *not* statistically significant: the genuine good results of psychotherapy are too few to outweigh significantly the cases of spontaneous remission.

Now a third question: If these statements represent the truth, why have they not been demonstrated unequivocally by thirty years of research?

The answer is complex, but I believe a large part of it to be as follows:

The majority of psychotherapy researchers have been trained in a tradition that overvalues scientific method and basically does not understand the true nature of science. The result has been the blind and compulsive application of methods that are inappropriate because they are too exact, and the neglect of indirect and less exact methods that are the only ones likely to give the true answers. This is one of the tragedies of our field and is the theme of the present essay.

The rest of the essay falls into two main parts. First I examine aspects of the history of four sciences, contrasting the situation in bacteriology, on the one hand, with that in biology, chemistry, and cosmology on the other. As I hope to show, at certain vital stages in the development of the latter three sciences, direct methods were impossible. The indirect methods that had to be used lay very far from the crucial experiments so often demanded in the field of psychotherapy and involved forms of argument that lay very far from the rigorous logic that so many people consider essential in any subject deserving the name of science.

I then turn to a detailed examination of two studies of brief psychotherapy (Malan 1963,1976), frequeptly drawing parallels with the logic of these other sciences. I show how the cumulative circumstantial evidence leads inescapably to the conclusion that in these carefully selected cases major therapeutic effects follow from the operation of specific factors.

THE TRUE NATURE OF SCIENCE

Bacteriology

Here we may take as an example the problem of the cause of anthrax. Much condensed, the sequence of events was as follows; (1) the discovery by a French physician named Davaine of the *invariable presence* of certain rod-shaped bodies in the blood of animals that had died of anthrax; (2) the demonstration by Koch that these rods could be grown in culture outside the animal's body; and (3) Koch's further demonstration that these cultures, when injected into healthy animals, reproduced the disease.

These three observations in combination (which have come to be generalized as "Koch's postulates") virtually proved that the rod-shaped bodies were the causal agents of anthrax.

The point of this summary of events is that it represents not an., idealized situation in scientific inquiry, but an *ideal* situation, because it really happened. But the trouble is that many people are led to believe that this is the only type of situation to occur, which they illustrate by carefully selected examples of crucial experiments which confirm or disprove a given hypothesis, such as the observation in 1919 of the bending of light by a gravitational field, which seemed to confirm the general theory of relativity. But this ideal situation does not always occur—crucial experiments are not always possible—and then it is necessary to fall back on methods that are., very different. In the excitement and fascination of describing the ideal logic of science, these other situations—which may indeed be the rule rather than the exception—tend to be forgotten. To emphasize this again and again and provide additional examples, we will look at aspects of the development of three other sciences.

Biology: The Origin of Species

The traditional view of the origin of species, which Darwin himself originally shared, was that each species was immutable and independently created by God, perfectly adapted to its environment. However, the more observations he made, the more evidence he began to see which suggested an alternative hypothesis, namely, that species had changed and evolved over the course of time, in response to natural selection or the "survival of the fittest."

Here then were two alternative hypotheses. How was it possible to choose between them? The natural answer that springs to mind is that he must perform crucial experiments. But any attempt to use this type of scientific method met difficulties that were totally insuperable. The difference between the problem facing Koch and that facing Dárwin is purely in the time scale: anthrax takes a matter of days to infect its victims, while the time scale of major steps in evolution is often measured in millions of years. Thus the sequence: hypothesis, prediction, crucial experiment, though not impossible *in i principle,* is impossible *in fact.* The result was that Darwin was f forced to act on the principle that "something might perhaps be made out on this question by patiently accumulating and reflecting on all sorts of facts which could possibly have a bearing on it"—in other words, on the principle of *cumulative, indirect, and circum-stantial* evidence. The result was a vast accumulation of independent lines of evidence which favored one theory over another—so much so that today the theory of evolution is regarded as a scientific fact.

A few examples will give a feeling of the arguments used in this hard-hitting yet extraordinarily nonpolemical report of scientific work.

The arguments cover a series of observations, each of which can readily be explained on the theory of natural selection, but seem arbitrary or irrational on the theory of special creation:

1. The presence of special adaptations that are *irrelevant* to the environment in which a given species is living, e.g., webbed feet in upland geese which rarely go near the water. Presumably these geese, originally living near water, have succeeded in colonizing a new environment, to which they are well adapted in other ways.

2. The fact that the species found in very similar oceanic islands, e.g., the Galapagos Islands off South America and the Cape Verde Islands off Africa, resemble not each other but those living on the nearest mainland. This can easily be explained by the hypothesis that colonists came to these islands from the mainland and subsequently underwent modification.

3. The fact that corresponding or "homologous" parts in different animals, e.g., the bones of the foreleg in mammals, seem to have been produced by extensive modification of a basic ground plan rather than by a special design suited to each function.

After describing each of these observations, he writes a passage rather like the following:

On the ordinary view of the independent creation of each being, we can only say that so it is;—that it has pleased the Creator to construct all the animals and plants in each class on a uniform plan; but this is not a scientific explanation.

The explanation is to a large extent simple on the theory of successive slight modifications. . . . [Darwin 1859, p. 435]

The argument really involves two principles, one being the universally recognized principle of *economy of hypothesis;* while the other consists essentially of what we may call *intuitive probability,* based on speculation about the mind of the Creator.

But one thing that this argument does not do, of course, is to disprove the hand of God in the creation and distribution of species. God may be postulated at any point in the process; if any of the observations that Darwin describes seem arbitrary, well "God moves in a mysterious way," and perhaps he had his reasons. It is, after all, a pretty unproductive exercise to speculate about the mind of the Creator—an argument that can be used to dismiss every one of the points that Darwin puts forward and is quite unanswerable.

The final result is that the Origin *of Species* (185.9).reads as much like a speech in a court of law as an exposition of scientific argument, with Darwin presenting his case like an advocate—not quite like a judge giving a summing up, because the jury is being very strongly directed. But it differs from an advocate's speech in that there is a total absence of special pleading or polemics or any attempt to gloss over the difficulties. He is an advocate who has become convinced that the difficulties and improbabilities will one day be overcome. It is up to us, the jury, to make up our own minds, but in the end the verdict will be based, like any verdict in law, on what it seems reasonable to believe and on nothing else. This is the major difference from the crucial experiments of Davaine and Koch described in the previous section.

Chemistry: Atomic Weights and the Search for Meaningful Regularities in Nature

The story of atomic weights is touched on in every elementary

textbook of chemistry. It is a subject which many students find infinitely boring, and yet it has a profound bearing on the nature of science as it really is and not as philosophers of science would like it to be. The theme once more is the problem of reaching a conclusion when crucial experiments are impossible.

Toward the end of the eighteenth century the atomic hypothesis was put forward, according to which each chemical element consists of indivisible particles which combine with one another in simple numerical proportions. An intrinsic property of each element was then the weight of one of its atoms relative to that of the lightest of all elements, hydrogen, a quantity called the atomic weight.

Now the problem of determining the atomic weight of a given element was one of the most intractable in the history of science. In the words of Aaron Ihde in *The Development of Modern Chemistry* (1964), this problem "was so formidable that it taxed the ingenuity of the best chemists of Europe for half a century." The reason is as follows: it was easy enough to determine the combining weight, i.e., the weight of the element that combines with unit quantity, say one gram, of hydrogen. For oxygen this is about 8 grams, and the result is water. If we assume that one atom of hydrogen combines with one atom of oxygen, i.e., that the chemical formula of water is HO, then the combining weight and the atomic weight are equal; but of course this is not necessarily so. If the chemical formula of water is H_2O, then the atomic weight of oxygen is twice times 8, or 16; if it is HO_2 the atomic weight is 8 divided by 2, or 4; and so on. All that can be obtained, therefore, is a series of possible values; there is no way of knowing which of these to choose, and no further progress can be made. In Ihde's words: "The dilemma is evident—formulas are , needed to calculate atomic weights, and atomic weights are needed to determine formulas." Once more, crucial experiments are *I* impossible.

The only hope is to make a start somewhere, which has to be done either by a guess, or else by using what evidence there is and making an *assumption* about how this is to be interpreted. The leader in the field, the Swedish chemist Berzelius, made use of an observation of Gay-Lussac's, namely that gaseous elements seem to combine in simple proportions by volume, for instance *two* volumes of hydrogen combine with *one* volume of oxygen to form water. If we *assume* that equal volumes of gaseous elements contain equal numbers of atoms (which of course on purely common-sense grounds seems improbable since one could be forgiven for supposing that heavier atoms would occupy larger volumes) then this suggests that the formula of water is H_2O, giving an atomic weight for oxygen of 16.

By arguments such as this, Berzelius succeeded in drawing up a list of atomic weights which possessed self-consistency if nothing else.

Each time a new element was discovered, however, chemists were faced with the same problem. The combining weight could be measured, but since the chemical formulae of its compounds were unknown, the atomic weight was undiscoverable.

The only way of getting round this problem was to *argue by similarity*. When a new element was discovered, its properties were compared with those of known elements, and if sufficient similarity was found with any particular element, then the inference was made that the chemical formulae were similar. Since the formulae of the compounds of known elements were themselves based on insecure assumptions, this was an example of inference piled on inference, assumption on assumption. Moreover, similarity is a highly subjective concept. Which are the *fundamental* properties and which the incidental ones? There was nothing to indicate the answer except intuitive insight based on a wide experience of chemistry. This was thus an example of subjective methods playing a major part in the determination of absolute and immutable quantities in physical science. Yet with the aid of this method and with the good fortune that the assumed formula of water was correct, Berzelius's list of atomic weights bears an astonishing resemblance to that used today. Another method which helped to choose the correct values for atomic weights resulted from the discovery of an *empirical law,* or in other words an unexplained regularity. It was discovered that if one took a given solid element and multiplied its supposed atomic weight by its specific heat, giving a quantity that was named the "atomic heat," then in a high proportion of cases one obtained approximately the same number. With modern atomic weights this number usually lies between 6.0 and 6.5. This is the Law of Dulong and Petit.

At this point the following kinds of inference can be made: First, since one has found an unexpected regularity, one is "on to something"; the regularity reflects-something important in nature which one day will be explained.- The argument here is based,on a fundamental *assumption:* that in *nature regularities do not happen by*

chance. Yet of course exactly the same argument could be used against this as was used by the opponents of Darwin's theory of evolution, namely that "it has pleased the Creator" to make this regularity and that the explanation is thus essentially unfathomable. Well, if it pleased the Creator to make this regularity, there were times when it didn't please Him, since the Law of Dulong and Petit possesses notable exceptions. The atomic heat of carbon in the form of diamond, for instance, is about 1.2. What about this?

Here one is immediately faced with a dilemma. There are three possibilities: (1) The Law of Dulong and Petit is wrong; (2) the value used for the atomic weight of carbon is wrong and should be altered to bring the atomic heat in line with that of other elements; or (3) carbon is simply an exception to the law, and one day this too will be explained. It was in fact the third of these possibilities that was accepted.

Since these accounts of problems in other sciences are all given for their bearing on the validity of psychotherapy, I find it useful to , pause and consider what a hostile counsel would make of the evidence. One can hear him asking: "What is the use of an empirical law to which you arbitrarily admit exceptions? Either it is a law or isnt."

To this the defending counsel can only answer: "But the evidence in favor of the law for a large number of elements is overwhelming. I cannot even explain the law, let alone the exceptions, but I believe that one day both the law and the exceptions will be explained.

This reply may seem pretty lame and will be dismissed by the opposing counsel with contempt, but in fact, of course, it is correct. In the twentieth century both explanations were found.

However, precisely because crucial experiments were impossible, there was room for complete disagreement, and the result was that by the middle of the nineteenth century the whole question of atomic weights was in a state of confusion.

Finally in 1858, Cannizzaro published a method of determining atomic weights with Tar greater certainty—though still with the aid of assumptions and therefore not a crucial experiment—and at last the interlocking measurements contained such a degree of internal consistency that it was almost impossible to doubt their validity. Now the way was paved for the discovery by Mendeleyev of the most fundamental concept in the whole of chemistry, the Periodic Table of the elements.

This happened through a method of investigation that is common enough in science but very far from a crucial experiment; namely the *retrospective arrangement of data into patterns,* with the additional methods of changing trie pattern wherever necessary to suit the data, and of *cheating* whenever the data seemed to contain exceptions and anomalies. The result was a pattern which was extremely striking on the one hand and extremely complex on the other. Its existence seemed to indicate both that the list of atomic weights was essentially correct and that atomic weights reflected some quite fundamental property of the chemical elements.

It is obvious that this kind of arbitrary, retrospective juggling with data is to say the least a hazardous occupation. The point is, however, that—as with the Law of Dulong and Petit—*the regularities were so striking as to outweigh the discrepancies.* This was hardly apparent to many of Mendeleyev's colleagues, who remained sceptical and acted as the hostile counsels postulated in the present essay.

However, even in his lifetime Mendeleyev was completely vindicated. This arose from his courage in making *predictions* from his pattern about elements which remained to be discovered. When these predictions were fulfilled, the hostile counsels were silenced forever. Yet, if these elements had already been discovered, predictions would have been impossible, and the retrospective arrangement of data would have been the only method he could have used.

In all this story there is a fundamental point. As I have implied many times, "truth" in science often arises not from any crucial xperiment but from an *overall assessment of the evidence.* This has nothing to do with rigorous logic, but much more to do with a concept that might be supposed to be more characteristic of law than (of science, namely *a conclusion based on what it seems reasonable to believe.* Because this has nothing to do with logic, circular arguments lare perfectly permissible.

The following is one of the circular arguments implied in the story of atomic weights:

1. We have determined atomic weights bymeans of various uncer tain hypotheses and empirical rules, which themselves seem to represent some underlying regularities in nature.

2. The atomic weights so determined fit into a complex but extraordinarily beautiful pattern.

3. This pattern in turn is only a hypothesis, but it also seems to represent some underlying regularity in nature.

4. We do not believe that this pattern can have arisen by chance.

5. Therefore *both the originalhypotheses* on the one hand, *and the pattern on* the other hand, are essentially valid and reflect the truth.

Nowadays, of course, crucial—or nearly crucial— experiments on all these questions have been performed.

But during the nineteenth century none of this wasfpossible, and all that could be done was to make observations, sea/ch for patterns, and pile one hypothesis on another, On the one hand the story illustrates the utter confusion resulting from this situation; but on the other hand it illustrates that *the correct answer can arise from the confusion without the use of crucial experiments.*. This is the relevance to the problem of the validity of psychotherapy.

Cosmology: The Dimensions of the Universe

The distances of the nearest celestial objects can be measured by triangulation—measuring the angles subtended by the object at the two ends of a baseline of known length, ultimately the diameter of the earth's orbit round the sun. But for the vast majority of celestial objects the differences in angle are too small to measure. Once more, the crucial experiment is impossible.

This deadlock was broken by a complex set of circumstances. At the beginning of the twentieth century Henrietta Leavitt of the Harvard College Observatory began studying the Clouds of Magellan, two dense clusters of stars then thought to be part of our own galaxy. She found that the clouds contained a type of star whose brightness varied over the course of time in a regular manner, usually with a period somewhat more than a day. She then made the crucial discovery of an empirical law, namely that there seemed to be a constant relation between the length of the period on the one hand, and the brightness of the star on the other.

Now since each of the Clouds of Magellan appears to be a cluster of stars, the individual stars in each cloud may be assumed to be all at roughly the same distance from us. If they are all at the same distance, then the empirical relation between the period and the apparent brightness becomes a relation between the period and the true brightness, and if we can find the true brightness, we can find the distance by the inverse square law.

The problem, then, is to calibrate the equation, which—if the empirical law is reasonably exact and always holds—means finding the true brightness of at least one example of this kind of variable star. This, of course, can be done only if we know its distance; and the trouble is that though similar stars occur elsewhere in our galaxy, even the nearest of them is too far away. So we come full circle, facing exactly the same kind of problem as was faced over the question of atomic weights.

When this kind of situation arises in any branch of science, it has to be circumvented by any means available, however precarious and uncertain these may be. Then, once more, the universally used circular argument is brought into play, if the results seem to make intuitive sense or give rise to new regularities, then it is tentatively concluded that both the results and the methods used to reach them are essentially correct.

The method used depended on the following principles:

1. Variable stars that seem to have the same properties as those inthe Clouds of Magellan are found scattered throughout our galaxy.

2. Stars that are hot too far away—though still too far for their distance to be measured by triangulation—change their positions in the sky over the course of years because of their motion at right angles to the line of sight. In other words, their angular velocity can be determined.

3.Might there be a way of determining their true velocity? This, together with their angular velocity, would enable their distance to be calculated.

4. A step toward this can be made by the fact that the component of their velocity in the line of sight can be measured by means of the Dóppler shift in the lines of their spectra (red shift if they are moving away from us, blue shift if they are moving toward us).

5. It is then necessary to make the *assumption* that the directions of their motion are completely random, so that on the average the component of velocity in the line of sight and the component at right angles to it are equal. If this can be carried out for a sufficient number of stars, then it provides a measure of their average velocity at right angles to the line of sight.

6. With this assumption the average actual distance moved can be calculated, and henee, by simple trigonometry, the average distance from us.

7. From the average distance and the average apparent brightness, the average true brightness can be calculated, and the equation is calibrated.

8. The true brightness of the variable stars in the Clouds of Magellan can now be calculated from their periods, and hence finally their distance.

The Clouds of Magellan were then found to be far enough away to be lying outside our own galaxy.

The essential point concerns the number of plausible yet questionable assumptions on which the whole chain of reasoning is based:

1. The assumption that the Clouds of Magellan are not appreciably extended in the line of sight. This was by no means certain—after all, the Milky Way *appears* to be a flat band of stars stretching across the sky, whereas we now know that it is a disc seen edge on with a diameter of about a hundred thousand light years.

2. The use of an *empirical relationship without known explanation* in order to measure unknown quantities. The trouble with this is that there is no way of telling how far the empirical relation is applicable or when there may be exceptions to it—compare the law of Dulong and Petit discussed in the previous section.

3. The assumption that the motions of stars are random, when we know very well that these motions contain all sorts of regularites.

4. The assumption that an empirical relation which holds for stars in the Clouds of Magellan holds for other stars in the galaxy.

Once more, anyone who wished to approach the evidence as a hostile counsel could destroy it to his own satisfaction in a moment.

This method could be used to measure the distance of the nearer galaxies, and other methods, also involving reasonable yet questionable assumptions, were developed for those futher away. Then came another pair of observations for which by now wonder has been dissipated by familiarity. First, when astronomers began to examine the spectra of these distant celestial objects, it became clear that the lines in the spectra of galaxies are all shifted toward the red. The natural explanation for this is that the galaxies are all receding from us at enormous speeds. And second, the speed of recession was found to be roughly proportional to their estimated distance from us—once more this was the discovery of an empirical law, an unexpected and unexplained regularity in nature.

Now it became possible to turn the equation around the other way, for the most distant objects to use the empirical relation between velocity and distance to measure distance. The essentially circular nature of this argument should be clear.

But we can go much further than that. The implied circular argument is similar to that used in connection with atomic weights and the Periodic Table, and runs as follows:

1. By a number of "reasonable" assumptions, we have developed a means of measuring the distance of galaxies.

2. This method leads to the discovery of a regularity.

3. Therefore the assumptions, and the method of measuring distance, were valid.

This is a type of argument that is universally used in science. Again we can go further. There is now a new circular argument:

1. The regularity admits of an extraordinarily simple explanation.

2. Therefore the regularity itself, the explanation for it, and the methods used to discover it, are all valid.

The regularity, as already described, is the relation between velocity of recession.and distance. This can easily be explained if the current state of the universe is the result of a single colossal explosion which occurred several thousands of millions of years ago. Provided there is no large force to change the velocity of the resulting fragments to different degrees, it follows inescapably that after a reasonable lapse of time, the fastest-moving fragments will become the furthest away and that the distance will be proportional to the velocity. Moreover, by simple arithmetic—since *time* equals *distance* divided by *speed*—it now becomes possible to calculate the number of years ago that the explosion must have occurred, which at that time came out to be a few thousand million years.

But did this figure fit in with other evidence? The answer was no.

Much cumulative evidence suggested that the age of the solar system—which, after all, is part of the universe—is about 4.5 thousand million years; since this exceeded the above value for the age of the universe as a whole, there must be something wrong somewhere.

The answer to this paradox illustrates the uncertainties of relying on empirical relations that no one understands. It eventually became clear that the fault lay in the fourth of the assumptions listed above. There were in fact two kinds of variable stars, and the equation had been calibrated with one kind while the distances of galaxies had been measured with the other kind. The discrepancy was such that all these distances, and hence the age of the universe, had to be doubled. The evidence now became self-consistent.

Again we can imagine what a hostile counsel would make of a method of measuring distances that is so uncertain that it apparently gives half the correct value. On the other hand, it could also be argued that we have just illustrated one of the strengths of scientific method, namely that in spite of the circularity of the arguments, there are independent checks that in the end enable us to discover errors and correct them.

THE VALIDITY OF PSYCHOTHERAPY

Of course it is true that the only way of proving or disproving the f. validity of psychotherapy beyond reasonable doubt is by means of a true controlled study. In such a study the sample of patients is *i* divided into two, the first of which is given treatment while the second remains untreated, and the changes in the two are then compared.

The difficulty with this design is that many changes—representing both improvement and deterioration—are known to occur during the period after termination. This means that an essential part of any valid study is a follow-up of many years; therefore the controls must be left untreated for a period equal to the length of therapy plus follow-up. This is virtually impossible for both ethical and practical reasons. The ethical problem is obvious; the practical problem concerns the difficulty of preventing patients from finding treatment elsewhere during so long a period. Additional problems arise from the natural wastage of patients, the funding of such a prolonged research project, and the need to keep the research team together.

The relevance now appears of the examples from other brunches of science given in previous sections. The length of follow-up required can be said to be analogous to the evolutionary time scale or to cosmic distances—in both cases sheer magnitude makes direct observation or measurement impossible. Is it true, therefore, that evidence about the validity of psychotherapy is impossible to obtain?

A reading of the material in the foregoing sections will immediately indicate my answer to this question: No, it is not impossible. Indirect methods not only can, but must be used. This means, as Darwin wrote, that "something might perhaps be made out on this question by patiently accumulating and reflecting on all sorts of facts which would possibly have a bearing on it."

In other words it means the use of all the types of evidence and argument used in other sciences, whether the less exact biological sciences or the extremely exact science of chemistry—methods such, as the examination of cumulative circumstantial evidence, measurement of questionable validity, empirical relationships with exceptions for which one seeks explanations, circular arguments, and finally the examination of the overall picture, with a verdict that can never be anything but subjective and based on "reasonable probability." The whole subject has of covirse not been deficient in hostile counsels, and these will never be convinced; but for those who have , an interest in finding out the truth, the evidence may possibly be I worth considering. It is this basic, universally recognized scientific principle—universally recognized, that is, except in the least exact of the sciences, the psychological, to which it is most applicable—that has occupied my research life for the past many years.

EVIDENCE FROM TWO STUDIES OF BRIEF PSYCHOTHERAPY

The Material

This consists of two series of patients treated by a team of experienced therapists under the leadership of Michael Balint. The original aim of the project was purely clinical, namely to investigate psycho-analytically oriented brief psychotherapy from first principles. The design was elementary in the extreme: to take on patients thought to

be suitable, treat them, and see what happened. I then carried out a single-handed study of the first eighteen patients (the first series), which was published in A *Study of Brief Psychotherapy* (1963,1975); this was followed by a far more rigorous study of the second series (thirty patients), published in *Toward the Validation of Dynamic Psychotherapy* (1976). Follow-up was extremely long—between four and nine years after termination, with very few exceptions.

In the present essay I do not intend to get involved in the "fine print" of the evidence, for which the reader is referred to the above two publications. What I shall present is the broad sweep of the two studies, with special reference to the nature of scientific evidence as illustrated by the examples from other sciences given above. I shall hope to show that the cumulative evidence is very strong, not merely that therapy was responsible for the improvements in these patients, but that the therapeutic factors were specific rather than nonspecific, that there are clear indications of what these specific factors are, and that these in turn are the same as those that psychoanalysts have always believed to be operating in psychoanalysis itself.

Evidence from the Duration of Disturbances

Let us suppose that a disturbance that has lasted twelve years is reported in the fourth therapeutic session as having shown a sudden improvement, ten weeks after the patient was first seen, and that when the patient is followed up seven years after termination the improvement has been maintained (in quoting these figures I have an actual case in mind).

In this particular case the ratio of twelve years (the duration of the disturbance) to ten weeks (the duration of therapy before the disturbance improved), which we may call the *duration ratio,* is about 55. It is surely straining credibility to suggest that the improvement "just happened" to occur at this point; if we could show that our series contained a number of cases of this kind, credibility would be strained even further.

In fact, careful examination of the thirty cases in our second series reveals that there were eleven in which the duration ratio, very conservatively estimated, was greater than ten. This suggests very strongly that some factor or factors concerned with therapy were associated with—indeed, probably *caused*—at least some of the improvements. For simplicity I shall refer to this as a single factor in subsequent discussion.

Evidence from the Relation between Outcome and Number of Sessions.

The distribution of these two variables in the two series taken together is shown in Fig. 1.

FIGURE 1.

DISTRIBUTION OF 46 BRIEF PSYCHOTHERAPY PATIENTS ACCORDING TO LENGTH OF THERAPY AND OUTCOME

NUMBERS OF SESSIONS	negatived	> 0 < 1.0	> 1.0 < 2.0	> 2.0 < 2.5	2.5	> 2.5 < 3.0	> 3.0 < 3.5	3> 3.5
>10	0	9	7	3	2	5	5	3
<10	1	3	4	3	1			

OUTCOME

As will be seen, there is an "overhang" on the right-hand side of the diagram, there being fifteen out of thirty-four longer-stay patients, and only one out of twelve shorter-stay patients, who scored more than 2.4 for outcome; and thirteen longer-stay patients, and *no* shorter-stay patients, who scored more than 2.5.

Now because of the large number of longer-stay patients who gave poor results (the top left-hand corner of the diagram), the I hypothesis suggested by this distribution is not that, statistically i speaking, the more sessions the patient has the better the outcome, but that *a certain minimum number of sessions is a necessary condition to a satisfactory outcome.*

This evidence marks the beginning of a long progression which steadily narrows down the factors that we are seeking. The evidence from "duration ratio" suggests only that there is some factor con-

cerned with therapy that is associated with favorable outcome. The evidence just presented suggests that really favorable outcome oc-1 curs only in association with a certain minimum quantity of this factor, whatever it may be. In other words, the factor is probably something that is given in therapy, whether specific or nonspecific, and not simply the fact of being taken on for treatment.

The Study of Motivation

In the first series I studied eight different selection criteria and found that the only one that gave a significant correlation with outcome was motivation.

This result needed to be examined very carefully. First, we have to ask, motivation for what? The answer to this was that motivation must not be merely for help but for the kind of help that is being offered, which is essentially concerned with ijgsighjtand nothing else. Moreover, most patients do not know that this is the kind of help that they will be offered until some point after therapy has got under way, which will therefore be the point at which motivation may best be judged. In accordance with this, it was found that when fluctuations in motivation during the first few sessions were taken into account, the relation to outcome was strongest.

In the second study I had the help of an independent judge, Dr. Peter Dreyfus, who could score motivation entirely uncontaminated by knowledge of outcome. Shorn of excessive detail the results were as follows:

1. Of ten criteria studied at initial assessment, motivation gave the highest correlation, but this could clearly have arisen by chance ($p < 0.1$ But > 0.05)7

2. When fluctuations in motivation were considered, none of the correlations with outcome was significant during the first Four contacTs with the Clinic, but there was a strong tendency for them to become significant during contacts five to eight.

The Study of Focality

One of BalintY original hypotheses about this kind of therapy was that it should be "focal," i.e., the therapist should plan in advance the main area of pathology which he wanted to work through, and by selective attention and selective neglect should concentrate his interpretations upon this planned theme.

In order to test this hypothesis, Dreyfus worked out a method of judging the degree to which the therapist had succeeded in keeping his interpretations on a single theme. This variable, called *focality,* was scored by both of us independently on each of the first eight contacts in the second series.

It is important to note that whereas motivation is defined as motivation for insight, in the patient focality is complementary to this, being concerned with the giving of insight by the therapist.

The following observation now emerged: Exactly as with motivation, the scores for focality were at first unrelated to outcome, but tended to become significantly correlated with outcome during contacts five to eight.

Meaning of the Studies of Motivation and Focality

As has already been discussed, the relation between outcome and number of sessions suggested that a minimum quantity of some factor was necessary for a really satisfactory outcome, but this observation offered no evidence about what this factor is, or whether it is specific to psychoanalytic therapy. Clearly it could be some nonspecific factor common to many different kinds of therapy, such as support or warmth. The present evidence, however, suggests that, the factor is indeed specific and consists of insight.

The Transference/Parent Link

The evidence given in the previous section leads naturally and inevitably to yet another question: Insight about what?

A careful study of the first series resulted in two observations:

1. The more successful therapies tended to be those in which, according to clinical judgment, the transference (i.e., the relation to the therapist) was thoroughly interpreted.

2. This relation was made much more striking if a subdivision of transference interpretations was studied, namely making the link between the transference relationship and the past, usually the relation to parents. I refer to this as the T/P link.

The question now arose whether this last observation could be checked by some more objective method. Here we can imagine the following ideal experiment:

1. Every session of these therapies had been videotaped.

2. Judges watched these tapes and made judgments on the therapist's nterventions.

3. These would include judgments of whether or not each intervention was a transference interpretation, whether or not it made the T/P link, and what the impact on the patient had been at the time.

4. Some measure integrating both the quantity and impact of these interpretations could be devised and correlated with outcome.

This ideal experiment may be compared with the ideal measurement of the distance of galaxies discussed above; and, as with that experiment, it was impossible. The therapeutic work had been already carried out, without any such judgments in mind; it was conducted in the 1950s, without financial support, before even audiotapes—let alone videotapes—became generally available. All the material that we possessed consisted of accounts of each session dictated by the therapist from memory, with varying degrees of fullness and entirely unknown accuracy. One may well ask what possible scientific use could be made of such uncertain data?

The answer to this question is similar to that given by astronomers to the question of the distance of galaxies: direct measurement is impossible, the measurements that are possible are most unsatisfactory and imperfect, and in consequence all one can do is to make a number of assumptions, make the measurements, and examine the results.

The first assumption was simply that the accounts of therapy, though condensed to a greater or lesser degree, represented a reasonable approximation to what actually happened; and that where the accounts were highly condensed, then in general the crucial moments were included. There was no way of checking this assumption, but it was made more plausible by the fact that all the therapists were well trained and highly competent.

The second assumption was that there was a relation between quantity on the one hand and correctness and impact on the other; This could reasonably be assumed from the fact not only "that they were well trained, but also that they all employed a technique of carefully monitoring feedback and quickly abandoning any line of interpretation to which the patient was not responding.

In order to eliminate as far as possible the effects produced by the different lengths of therapy and differences in the style and completeness of recording, I used not an absolute number but a proposition, for example, as a measure of the "importance" of transference interpretations I used the ratio of the number of transference interpretations to the total number of interpretations recorded for each therapy. Each ratio was expressed as a percentage and was referred to as the "transference ratio," "transference/parent ratio," etc.

Results of the Study of Interpretations

This method was applied to both series. The study of the second series was far more elaborate than that of the first and involved judgments made by two independent raters of the proportions of fifteen different categories of interpretation, and the correlation of these with the scores for outcome. (My collaborator was E.H. Rayner.)

The essential result was that the transference/parent ratio was the only one of these variables that gave a positive and significant correlation with outcome in both series.

Meaning of the Result on Transference/Parent Interpretations

At the end of the sections on motivation and focality I concluded that the evidence suggested an association between favorable outcome and the acquisition of insight. I then asked the question, insight about what? The evidence just presented suggests a very clear answer: insight about the way in which the relation with the therapist repeats patterns derived from relations with parents in the past.

Clinical Meaning of the Quantitative Study

I now wish to introduce a circular argument that is very similar to that used over the question of cosmológica! distances and the age of the universe.

1. We started with two assumptions: (a) that case notes dictated from memory bear a reasonable relation to what actually happened; and (b) that a purely quantitative measure of interpretations bears a relation to their correctness and impact.

2. We have found a significant correlation between a purely quantitative measure of transference/parent interpretations and out come.

3. Therefore not only are both assumptions (a) and (b) correct, but there is an association between the *impact* of transference/parent interpretations and outcome.

Moreover, this result admits of a simple explanation in terms of known phenomena. Suddenly we realize that the relation between quantity of interpretations and outcome probably represents the well-known fact that a given piece of insight does not merely need to be experienced but needs to be worked through—which means experienced many times in different contexts. This phenomenon, recognized for many years in long-term therapy, apparently applies to short-term therapy as well. In other words, assumptions, evidence, j; and explanations make sense as a whole, and it is a reasonable» inference to say provisionally that they are all correct,

Cause and Effect

So far, the only statements that have been made concern an association between certain factors and favorable outcome, and hardly anything has been said about cause and effect.

It has to be emphasized that this was a naturalistic study, and whether or not the patient received T/P interpretations was the result of self-selection. The inferences to be drawn from such a study need careful thought.

There are three main types of possibility, representing the three main ways in which correlations may be interpreted: (1) The first is the obvious one, that T/P interpretations cause favorable outcome, while there are also two alternative hypotheses, namely (2) that favorable outcome in some way causes the therapist to make T/P interpretations, and (3) that favorable outcome and T/P interpretations are both associated with some third factor, as yet unknown.

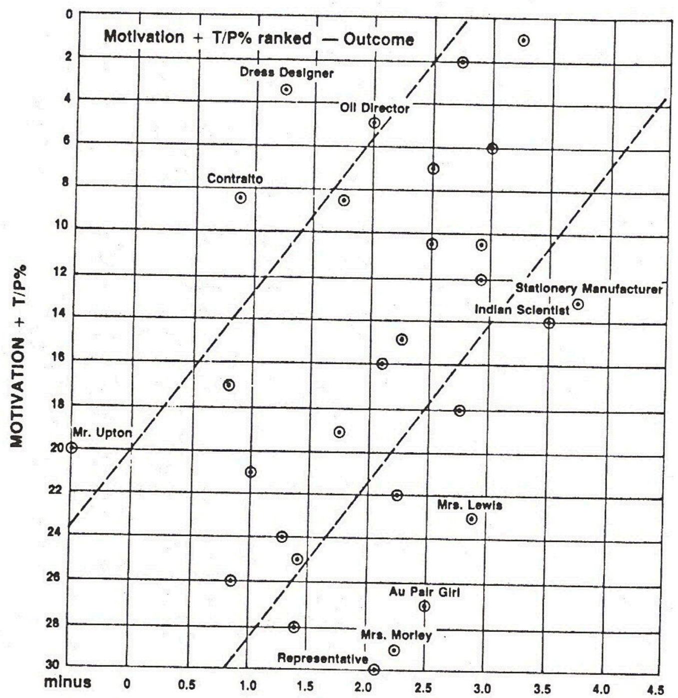
Those who maintain the alternative hypotheses (2) and (3) have usually not thought the situation through. They are under obligation to offer not merely alternative explanations, but more plausible ones. First, the suggestion that favorable outcome in some way causes the therapist to make T/P interpretations simply does not fit in with the facts. Every therapist knows that transference develops rapidly in many therapies and is interpreted as it arises, usually long before any therapeutic effects are manifested—many of which in any case only appear after termination.

The other explanation offersa more viable alternative, One could put up the hypothesis that these patients tend to get better because they interact with people in their life outside in a particular way; that they also show the same kind of interaction in therapy; and that this in turn causes the therapist to perceive the relation with the past and to make interpretations about it. This may even be partly true, but the argument tends to turn back on itself, since if this kind of interaction is therapeutic in life outside, why should it not be therapeutic in the relation with the therapist?

Moreover, the evidence provided by motivation and focality suggests strongly that the factor associated with favorable outcome is not the interaction itself, but insight about it, Most of these patients had been showing this kind of interaction for years, but insight about it was acquired only when they came to interact with a therapist. Therefore this line of thought, after a brief detour, leads straight back to the inference that therapy itself is causal.

Motivation and Transference/Parent Interpretations in Combination

There is a final piece of evidence which further reinforces all that has been presented hitherto. If the thirty patients in the second series are ranked for overall motivation during the first eight contacts and the rank numbers so obtained are.added to. the .rank numbers for T/P percent, a further series of rank numbers is obtained for "motivation +T/P percent." This represents a combined rank for the two most; importantTáctors that correlate with improvement. If this variable is now plotted against outcome, the scatter diagram shown in Figure 2 is obtained. In this diagram there is an obvious trend from "below left" to "above right," but there are also certain patients who lie away from the general trend. These patients are those who lie outside the two arbitrary lines drawn on the diagram.



There are six patients lying to the right of the lower line, in whom outcome was more favorable than the combination of motivation and T/P percent would predict. Of these, the Indian Scientist was in no way exceptional on any other grounds. However, all five of the others showed unusual characteristics on purely clinical grounds, which set them apart from the main body of relatively young, purely neurotic patients in the sample as a whole. Thus the Stationery Manufacturer was a near-psychotic man of forty-six; Mrs. Morley was a woman of over sixty thought to be near a severe depressive breakdown; Mrs. Lewis and the Au Pair Girl improved long after termination in response to major life experiences (i.e., showed "spontaneous remission"); and the Representative showed major improvements after a clear "flight into health." In each of these last three patients the evidence suggests a mechanism of improvement independent of therapy.

There are four patients to the left of the upper line. These are patients in whom outcome was less favorable than the combination of motivation and T/P percent would predict. Again there is one patient, the Oil Director—and as before the one nearest to the line— who was not exceptional in any way. But the other three were all patients who turned out to be far more deeply disturbed than was seen atSnitial assessment. The Contralto was eventually seen to be suffering from a "false self," leading to an inability to form real relationships with anyone; the Dress Designer, originally thought to be suffering from a simple phobic illness, showed an extremely deep-seated illness involving primitive fears of loss of identity; and Mr. Upton, originally thought to be suffering from very ordinary conflicts about growing up, at follow-up showed bisexual, near-psychotic, and manic-depressive features.

Although the analogy with Mendeleyev's work should not be pushed too far—he, after all, was dealing with absolute, universal, and immutable quantities—there are interesting parallels. The features in common are the arrangements of observations in such a way as to form a pattern, followed by the inference that the pattern is so clear that the exceptions must be explainable. However, whereas it took half a century to explain the exceptions in the Periodic Table, in the present case the explanations spring to mind immediately.

Davanloo's Evidence

During the past twenty, years Davanloo has been perfecting.a.technique of brief dynamic psychotherapy that is clearly more powerful than any discovered hitherto. He is now in process of examining his own evidence systematically. He states (personal communication) that the ability to give a meaningful T-C-P interpretation in the initial interview is an extremely important favorable prognostic sign and that the more frequently such interpretations are given in the early stages, the shorter the therapy will tend to be. Thus the conclusions presented above have now received confirmation, entirely independently, from a third series of brief therapy patients.

Evidence from Traditional Clinical Judgment

We may ask why it was that this particular hypothesis about T/P interpretations was examined at all. The answer is very clear, namely that this type of interpretation has been regarded as the most important therapeutic factor in psychoanalytic therapy for many years. This point is made by at least four separate writers on technique, Glover (1955), Alexander (1957), Menninger (1958), and Strachey (1969).

Thus we can add to the evidence already described cumulative evidence from other sources. We can now state that a particular type of interpretation, which has been shown to correlate with improvement in five separate studies of dynamic psychotherapy—two of ours and three of Davanloo's—has been considered for decades as one of the essential therapeutic factors by generations of psychoanalysts.

Can this evidence still be generally ignored? Yes, of course it can— rigid theoretical positions seem to be the rule rather than the exception in science and nowhere more than in the field of psychotherapy.

The Final Analogy with Other Sciences

Although I have mentioned parallels with the logic of cosmology and chemistry, the final analogy that I wish to make is with the logic of Darwin's work. Darwin made an exhaustive study of natural living and extinct organisms and came to the conclusion that a dispassionate appraisal of cumulative circumstantial evidence overwhelmingly supported the theory of evolution by natural selection rather than that of special creation. In our own work, cumulative circumstantial evidence can be shown to support—obviously less overwhelmingly but still powerfully—the validity of psychotherapy. However, the analogy with Darwin leads to one word of warning: these indirect methods can be used only with total objectivity.

Otherwise, of course, they can be made to lead in whatever direction anybody chooses. The marriage between polemics and truth tends to be shown, sooner or later, to be a case of incompatibility, leading inevitably to divorce.

A PSYCHOTHERAPY RESEARCHER'S DREAM

In the early part of the twentieth century a dream came true: with the aid of instruments such as the mass spectrograph it finally became possible to determine atomic weights directly with virtual certainty. Could anything like this ever occur for the validity of psychotherapy?

Let us have a dream. A new method of dynamic psychotherapy is discovered, of extraordinary power. The origins of the neurotic patterns in the past emerge with total clarity through the derepres-sion of long buried anxiety-laden feelings. These are experienced with such intensity and completeness that therapeutic effects begin to follow soon after the beginning of therapy, and the whole neurosis dissolves after about fifteen session, leaving the patient not merely improved but recovered. The events in therapy that lead to therapeutic effects can be observed directly, and lengthy follow-up shows that these effects are permanent.

This leads to a situation in which criteria analogous to Koch's postulates can be applied, as follows:

1. Antecedent events or situations, and particularly precipitating factors, enable a strong inference to be made about the nature of the conflict underlying a neurotic symptom.

2. An examination of the symptom itself makes possible a clear-cut hypothesis about the way in which it expresses or symbolizes the underlying conflict.

3. During therapy, the conflict becomes fully conscious and is accompanied by intense and painful feeling.

4. The symptom immediately disappears.

These four criteria, if fulfilled, would virtually prove that in this particular patient the neurotic symptom had its origin in a specific conflict. Moreover, as experience increased, it would be possible to use these criteria predictively, thus performing a true scientific experiment. This would lead to the virtual proof of psychodynamic theory, something that has eluded us ever since its origins in the 1890s.

Since this is a fantasy, there is nothing to stop us from going further. This method of therapy is applicable to severe and chronic conditions: as an example, fifteen weeks of treatment result in

complete recovery from symptoms that have lasted for twenty years—a ratio of about 70. Moreover, some of these conditions have failed to respond to years of psychoanalytic or behavior therapy, or both, This is the method of "duration ratio" carried to the limit. In this type of experiment, of course, the patients act as their own controls, and the result is virtual proof that it was therapy that caused the improvement. Thus psychotherapy is validated without a formal control series.

Yet mention of this possibility suddenly makes us realize that our belief that it is impossible to have a control series of untreated patients is an example of perseveration. As already mentioned, recovery is complete by the time therapy is terminated, and follow-up merely confirms that there is no subsequent relapse. At once the waiting-list control design becomes both valid and ethically and practically feasible. Control patients now only need to wait for as long as the therapy of the experimental patients lasts—not therapy plus follow-up. Changes in the experimentáis at termination are then compared with changes in controls at the end of the waiting period, the passage of time in the two series being equal. As mentioned above, the method of therapy is so powerful that there is not the slightest difficulty in demonstrating the advantage of experimentáis over controls. This would be the crucial experiment that we have been waiting for in vain all these years.

You may say that this is an absurd dream and may question why it should be mentioned in a serious article about the validity of psychotherapy. The dream is so utterly unbelievable that even those who have seen it come true with their own eyes on Davanloo's videotapes, as shown in many national and international symposia on Short-Term Dynamic Psychotherapy, mostly cannot take in the evident fact that it has become a reality. Clinical examples of this work are described in other chapters of the present book and have also already appeared in the Proceedings of the first two International Symposia (see Davanloo 1978); a further book, describing his work in greater detail and entitled *The Relentless Healer,* is in preparation.

An application for a research grant based on the waiting-list control design is also in preparation; if all goes according to plan, hostile counsels will be hard put to it to find fallacies in the evidence. The work described in the present essay will then become, like Berzelius's atomic weights, basically correct but of historical interest

only —an outcome that would be of benefit to mental health throughout the world.

References

Alexander, F. (1957). *Psychoanalysis and Psychotherapy.* London: George Allen and Unwin.

Darwin, C. (1859). *The Origin of Species.* New York: Collier, 1962.

Davanloo, H. (1978). *Basic Principles and Techniques in Short-Term Dynamic Psychotherapy*. New York: Spectrum.

_____(in preparation). *The Relentless Healer.*

Dible, J.H., and Davie, T.B. (1945). *Pathology*. London: Churchill.

Fernie, J.D. (1969), The period-luminosity relation: a historical review. *Publications of the Astronomical Society of the Pacific* 81:707.

Glover, E. (1955). *The Technique of Psycho-Analysis.* London: Bailliére, Tindall and Cox.

Ihde, A.J. (1964). *The Development of Modern Chemistry*. New York: Harper and Row,

Malan, D.H. (1963) A *Study of Brief Psychotherapy*. London: Tavistock; Philadelphia: Lippincott. Republishedby Plenum Press, New York, 1975.

____(1976). *Toward the Validation of Dynamic Psychotherapy*. New

York: Plenum Press,

____(1979). *Individual Psychotherapy and the Science of Psycho-dynamics.* London: Butterworths.

Menninger, K. (1958). *Theory of Psychoanalytic Technique.* New York; Basic Books.

New Frontiers in Astronomy. Readings from *Scientific American.* San Francisco: Freeman.

Strachey, J. (1969). The nature of the therapeutic action of psycho-analysis. *International Journal of Psycho-Analysis* 50:275.