The Principles of Humane Experimental Technique

W.M.S. Russell and R.L. Burch

CHAPTER 5

REPLACEMENT

With respect to the "analogical" ... resemblances between organic beings ...

The Principles of Replacement

The Lack of a General Theory

Replacing techniques are, as we have seen, specially desirable on humane grounds. Apart from great savings of cost and time, their use has been attended by scientific rewards--such as the discovery of new vitamins and viruses--so great that one is in danger of adopting a superstitious attitude. In this field, it seems, humanity is its own reward. But although there may be quite fundamental reasons for the correlation, this belief is no adequate basis for the systematic and rational extension of replacement. As we shall see, replacement is widely used in some fields, while in others it is very far from being exploited to the full, if at all. Moreover, such developments have been largely empirical, and largely independent of each other. They have often occurred because other methods present insuperable obstacles, as frequently in the history of virology. The use of microorganisms for nutritional assays has been one of the most spectacular instances of successful replacement. Yet even this application was suggested some twenty years before it was finally realized (Sykes, 1957). Such isolated and haphazard advance always occurs in the absence of a general theory. Since the advent of replacement has always meant great advances and advantages, a general theory would be really welcome here, and should facilitate progress on a very broad front in the methodology of both biological research and its applications. An attempt has been made to adumbrate the outlines of such a theory, or at least of the field it must cover (Russell, 1957a, on which this section is largely based). Here we shall continue the sketch, fully aware that its realization as a full-scale picture must be the work of others with the requisite logico-mathematical equipment.

We must distinguish two important cases, which arise when we consider the *object* of experiments. Take, for instance, the study of endoparasites of higher animals. A

rational chemotherapy must take account of the fact that the parasites are to be killed inside the host, to whose metabolic process the lethal drug will be exposed. However, long before this stage, it is desired to explore the biochemistry of the parasites, as a prerequisite for the rational development of drugs to destroy them, and to test these drugs purely from the point of view of their efficiency as selective killers. For these purposes, we want the parasite by itself in vitro. To be forced to study it in the living hosts is a restriction unfortunate for both host animal and experimenter. Great attempts will therefore be made to get at the parasite directly, and, by culturing it *in vitro*, to dispense with the host, which is simply an obstacle. Replacement in such fields is hindered only by technical difficulties. These include the unfortunate circularity that the culture is often difficult before study of the parasite's biochemistry, and, sometimes, with protozoan parasites, the problem of an organism that takes quite different forms in vitro and in the living hosts. (This problem has recently been solved for a trypanosome, which has been converted in culture to the *in vivo* form by the addition of vertebrate serum--Steinert and Boné, 1956.) Wherever these conditions apply, the incentives will be maximal, and sustained efforts will be made to solve the technical problems, as is most spectacularly shown in the field of virology. No theoretical argument arises at all, and the problem is merely that of achieving more direct study of the object of investigation. These conditions apply whenever organisms other than vertebrates are to be studied directly--metazoan parasites, infective microorganisms, etc. At the level of routine experiment, this is true for the whole practice of medical and veterinary diagnosis, except only for the recognition and estimation of *virulence*, which may be a property of both pathogen and host.

In the remaining, much larger, class of investigations, we are primarily concerned with the study of the vertebrate organism itself, and more specifically of a small number of species--man and his domesticated animals. Reference to Chapter 3 and the tables will support the view that the largest proportion of all experiments in biology, routine or research, is intended to provide information about the functioning of the *human* body in health and disease, and the effect upon it of a great variety of substances. The next largest proportion is similarly concerned with the bodies of the more important (socially or economically) of the domesticated animals. The much smaller residual proportion is concerned with the study of other vertebrate species for their own sake, though practically all of the knowledge so acquired bears sooner or later upon one or both of the two major purposes. We shall concentrate on the medical objective, which embraces, besides much pure and applied research (and teaching), a substantial proportion of routine pharmacology and chemotherapeutics. That which we shall put forward can easily be reapplied to the veterinary field.

If we are ultimately studying the human body in health and disease, and the effects upon it of substances and pathogenic organisms, the *only* direct method of approach is

to experiment upon the human subject--a procedure always to be viewed with the greatest caution (cf. Editorial, *B.M.J.*, 1955). The human body is the *system* to be studied, and only thus can it be studied directly. Alike in research and routine testing, we must distinguish between clinical and *all other methods*.

Any of these other methods consists, essentially, in setting up a *model* of the system to be studied (i.e. the human organism), and studying the model. (For the importance of such methods in science, cf. Craik, 1943; Young, 1951; Miller, 1955; Ashby, 1956a; Gerard *et al*, 1956; Russell, in press, b.) Instead of direct study of a human in certain conditions, we use a dog or a rat or a mold as a model, from which we hope to infer the behavior of the human body (or *parts* of it) in similar or analogous conditions. We are using the dog or rat or mold as an analogue computer, just like those used by engineers when for reasons of cost or accessibility they cannot directly study the system that interests them.

A perfect model of the human organism (such as that made by Pygmalion, but not those made by Frankenstein or the Rabbi of Prague) would obviously be indistinguishable by any test from its original. Any other model, whether monkey, dog, rat, fish, mold, or bacterium, must depart in some degree from the properties of the original.

There are, however, two factors governing the way in which the model differs from the original. These factors we may call *fidelity* and *discrimination* (Russell, 1957a). Fidelity means overall proportionate difference, and high fidelity (as in sound reproduction) simply means that *all properties* are *equally badly reproduced*. Discrimination, on the other hand, means the extent to which the model reproduces *one particular property* of the original, in which we happen to be interested. Of two models of the same system, one may be of poorer fidelity than the other while at the same time of higher discrimination for one particular property.

This may be rather vividly illustrated in behavioral example (Russell and Russell, in press), as in the presentation to a herring gull chick of two different (literal) models of the head of its parent--especially shown in *Figure 4*. In this instance, the second model, of poor fidelity but high discrimination for certain key properties of the original elicited more begging responses than the first, despite the latter's extremely "hi-fi" quality. For the activation of behavioral releasing mechanisms in lower vertebrates is often a function, not of the overall pattern, but of certain key stimulus features of the natural stimulus object (Tinbergen, 1948, 1951, etc.). To activate them, discrimination is more important than fidelity.

Figure 4. Fidelity and Discrimination

(From Tinbergen and Perdeck, 1950, Figure 33)

The figure shows some results of Tinbergen and Perdeck's experiments on the stimulireleasing, food-begging reactions in the herring gull chick. They presented the chick (in succession) with various models. Every time a model was presented, the experimenters gave an imitation of the call normally given by a parent-bird when about to feed the chicks. The model was then held in front of the chick for thirty seconds, and the number of reactions counted (that is, the number of times the chick pecked at the model). Such tests were repeated a large number of times, and the experimenters were able to add up and compare the number of reactions released by different kinds of models.

On the left side of the figure, two models are shown. The upper one is a threedimensional, accurately shaped and coloured model of the parent-gull's head and beak. It is a model of very *high fidelity*. The lower model is a thin red rod, with three sharply edged white bands at its tip. It is extremely unlike a gull's head. It does, however, present three stimuli which were shown by other experiments to be of special importance for releasing the reaction. Such stimuli are called key stimuli. The three in question are redness, colour contrast, and elongation. The lower model is highly *discriminative* in respect of these properties. (In fact it is superior in these respects to the actual head of a real parent-gull, and may be called *supernormal*).

The bars on the right side of the figure indicate the relative number of reactions released by the two models. The numbers (which are printed at the end of each bar) were expressed as percentages of the number released by the upper model, which is therefore scored as 100. The figure at the bottom shows the absolute number of reactions observed.

The result shows that a highly discriminative model of very poor fidelity releases more reactions than a high fidelity model. The figure is used here as a graphic illustration of the concepts of fidelity and discrimination in models.

In Chapter 1, much stress was laid on the elaborate inter-dependence of all components of the vertebrate (including human) organism. If this were total, fidelity

would be the only valuable requirement of a model used for medical purposes. Fortunately, this is not entirely true. It is possible to analyse and isolate component functions (cf. Russell *et al*, 1954). In more precise terms, the human organism is a *reducible* system (Ashby, 1956a). If this were not so, experimental biology could never have come into existence.

In fact, in many fields, discrimination is recognized in practice to be the more desirable quality. That is, models are employed which give specifically good response over one particular sector of the human physiological spectrum. Species vary considerably in their discriminativeness for special properties of man. If we are interested in studying the human cerebral cortex, primate species may be more suitable than, say rats. In this instance, evolutionary relationship and homology happens to be important. But this is not always so. The luteotrophic hormone of the adenohypophysis was first discovered--and is still assayed--in connection with the growth and shedding of cells of the pigeon crop gland (Riddle et al, 1933). This organ has nothing whatever to do, in terms of homology and phylogeny, with the mammary glands of man and other mammals upon which luteotrophin acts. More dramatically, in some nutritional contexts, particular strains of microorganisms may be more useful models than mammals. Differences are sometimes more useful than similarities. For discriminative assays of the D Vitamins, both rat and chickens are used, precisely because of their differences. (Indeed the assay of Vitamin D^3 is one of the main uses of the latter species--cf. Tables 10, 13.)

Thus, again and again in particular fields, models of high discrimination and often of very poor fidelity have been accepted through sheer necessity as a matter of course. But this process has never been canalized by means of a set of general principles governing the use of models. One general characteristic of all replacing techniques, when contrasted with living intact mammals, is their relatively (often extremely) poor fidelity as models of the human organism. It is our belief that progress in replacement has been restricted by certain plausible but untenable assumptions, which have yielded only gradually and piecemeal to the logic of empirical practice. These assumptions may be summed up as the *high-fidelity*, or *"hi-fi"*, *fallacy* (Russell, 1957a).

The High-Fidelity Fallacy

There have been some medical men who have denied the slightest value to any nonclinical results. One of them is supposed to have declared that what was clinically proven needed no other proof, and that what was not clinically proven was not proven at all. These individuals have usually been antivivisectionists, and need not concern us here. Such utter disbelief in the use of models, without which science could not exist, must by now be on the way out in the medical profession. The more commonly encountered high-fidelity fallacy takes the form *(implicitly)* of an argument running roughly as follows. Man is an eutherian (placental) mammal. A member of a mammalian species, considered as a model of man, is a model of relatively high fidelity, compared with a bird or, still more markedly, a microbe. In other words, in their general physiological and pharmacological properties, mammals are more *consistently* like us than are other organisms. No zoologist, of course, will argue with this minor premise (cf. Woodger, 1945). The major premise states that high fidelity, indeed the highest possible, is always desired in medical research and the testing of biological substances. This premise acquires its great emotional weight from the fact that caution here, whatever irrational forms it takes, seems to be dictated by the demands of public health and safety. The conclusion is that mammals are always the best models. This conclusion is maintained with special stubbornness in some special fields (such as that of toxicity testing). But a similar general assumption, usually entirely implicit, stands like an unshakable monolith in the path of any rational approach to the replacement of mammals by lower organisms.

It would be folly to deny that fidelity is ever necessary or desirable. There is some truth in the notion that the fidelity required of a model is in part a function of our ignorance. If we know practically nothing of the sub-system we are studying (say effects, especially toxic, of a completely new and untried substance), we may feel that the safest bet is to try it on the dog, or on something else as generally like the human organism as possible. At the other extreme, when we know all the properties of a known chemical substance, we may be prepared to assay it with physical and chemical apparatus of very high discrimination indeed, which has virtually nothing in common--not even life--with the human body.

But this brief formulation is misleading as a general principle, and the high-fidelity fallacy is accompanied by three important and still implicit assumptions, which brand it as an obsession rather than a principle. First, the extent of our ignorance may be exaggerated. Second, the fidelity of mammals as models of man may be greatly overestimated. Once a model, through poor fidelity, begins to depart seriously far from the original in respect of some property crucial for the current study, it loses any advantage it may ever have possessed over a model of much poorer fidelity which may happen to be highly discriminative for the property in question. A lower organism may, paradoxically, have something important in common with man that is absent in nonhuman mammal species. Evolutionary conservation or convergence may unite man (a highly unspecialized mammal in many ways) with some very lowly organism, while specialization separates from him most or all of his fellow-mammals. This is no surprise to the zoologist, who knows (for instance) that, although frogs are classified in one group with the earliest amphibia, they differ from these even in bone structure much more profoundly than do modern lizards (Evans, 1944). (After all, in

the matter of tails, we ourselves are more like frogs than monkeys!) Third, and most important of all, the high-fidelity myth tends to ignore all the advantages of *correlation*. We may show that responses of two utterly different systems may be correlated with perfect regularity, so that if a given effect is produced upon one by a given treatment, this will certainly produce a corresponding (but utterly different) effect upon the other. Two such systems may be perfectly mapped, one upon the other. This mapping will not appease the real "hi-fi" enthusiast, for in such connections the fallacy becomes almost a mystique.¹

We may consider a few practical points against this background. First, there are certainly some fields where mammals (and sometimes higher animals in general) are far from reliable guides. "A disturbing feature in the work of testing compounds for anti-tumor activity is that many compounds are effective in laboratory animals but are without effect in a majority of human neoplasm" (Galton, 1957). The antibiotic cycloserine,

"although fairly active *in vitro*, was found to be inactive in mouse and guinea pig tuberculosis and other infections in animals. It would normally have been rejected, but owing to its virtual lack of toxicity in animals it was tried clinically and found to be highly effective in man. This raises the question as to whether the *in vivo* results in animals are any more reliable than the somewhat discredited *in vitro* technique for the assessment of the value of a new antibiotic" (Birkenshaw, 1957).

And here is a nice point raised by a good deal of modern practice: which is the model of higher fidelity to the whole human organism--an intact nonhuman mammal or a culture of human tissue *in vitro?* If a substance produces certain effects at the tissue periphery, these may be masked by metabolic or detoxification mechanisms in nonhuman mammals which are not present in man. As for correlation, the point has been well put by Grove and Randall (1955) in a discussion of chemical and microbiological assays of antibiotics:

"When one demonstrates the ability of an antibiotic to kill or inhibit the growth of a living microorganism, as is done in the microbiological assay, a direct measure of the activity or potency of the antibiotic is obtained. In order for a chemical assay method to be of value, therefore, it must be able to give results that will correlate well with those obtained by microbiological assays. The chemical or physical methods of assay presented [in their book] ... have been shown to give good results in good agreement with those obtained by bioassay."

From the present point of view, we are not interested in the substitution of one of these absolute humane methods for the other. But the general argument is equally valid for the comparison between animal experiments and replacing techniques. All that is required is accurate and reliable parallelism, and we do not need to know anything whatsoever about the reasons--our ignorance here is simply irrelevant. Instances of such correlation could be multiplied; two may suffice here: the very close agreement between tissue culture and *in vivo* tests of the effect of eight different substances on two kinds of mouse tumor (Eichorn *et al*, 1954), and gross correlation between the relative toxicity of eighteen different substances for cultured explants of human skin and embryonic chick spleen on one hand, and on the other their irritant effects on the skin of living human patients and rabbits (Livingood and Hu, 1954). Correlation studies of this kind are often the first steps in the discovery of excellent discriminative models.

When correlation is imperfect, further investigation becomes necessary. An excellent comparison of methods for viable count estimations of tumor cell suspensions has recently been published by Hoskins *et al* (1956). *In vivo* titrations in mice were compared with four different *in vitro* techniques. None of the five methods was perfectly accurate, and there were discrepancies between results obtained by different ones. The authors, therefore, examined the particular way in which each test was operating, in order to specify conditions which would reduce the discrepancies. This sort of inquiry may be a necessary second step, for many discriminative models are chosen on the basis of detailed knowledge of the replacing model (e.g. of the biochemistry of microorganisms). The latter development is eminently desirable from the humane point of view.

Where such knowledge is lacking, parallel results are still, in themselves, perfectly adequate grounds for choice of a model. Virulence, for instance, is normally a complex property of pathogen-host interaction. But if it can be unfailingly correlated with a property (such as antigenicity) which can be tested, this is all we need ask for practical purposes. Virulence tests are among the least humane encountered in diagnosis. *In vitro* tests of virulence, which usually save cost and frequently time, are specially to be welcomed (cf. King and Frobisher, 1949--diphtheria virulence; Burrows, 1956--pasteurella virulence).

Finally, there are contexts in which ignorance is well recognized to be no barrier to bliss. In producing a vaccine, our aim is so to modify the pathogenic organism concerned that it will retain its antigenic structure (thus conferring active immunity) while losing most, if not all, of its virulence. It does not matter a scrap how this is done. It may be a matter of trial and error, and the modifying system need not have anything particular in common with man. A great many vaccines can now be produced by such modifying systems as tissue cultures and hen's eggs (including, incidentally, that for canine distemper--Scanlon and Fisher, 1951; Cabasso *et al*, 1951--which, as we saw, employed thousands of live dogs in 1952).

Against this background, the high-fidelity argument is seen to lose most of its force. While it is often ignored in practice, it has never been effectively combatted in principle. Such refutation and (more important) effective general progress here depend alike on the development of a completely general theory.

Towards a General Theory of Replacement

Evidently we need precise information on the conditions under which models of poor fidelity will be useful, and on the conditions under which two models of very different degrees of fidelity may be equally good for purposes of discrimination. The commonsense remarks of the last few pages need to be buttressed by more rigorous conceptions, and a really general theory provided. It might supply many more, perhaps unexpected, guiding principles.

Such a general theory, if it has not yet arrived, is on the way (Ashby, 1956a; and cf. Anon., *Nature*, 1956). Rules for the use of models are gradually emerging from an area of mathematical theory of great generality, which is related to "black box theory", long the playground of the engineer. What is important is that the rules can be laid down for any definite degree of ignorance of the insides of the black boxes. A brief account in outline of such a system of rules has been provided in Ashby's admirable text of cybernetics (1956a). In general, it supports most of what we have said. If two models of totally different kinds give regularly correlated results, they may be described as *isomorphic* with each other (see Fig. 5). In such circumstances, it is absolutely indifferent which of them we use. What we described as discriminative models are essentially, in terms of the theory, *homomorphisms*² of the original system. Very rough and imprecisely, this means that if we can simplify the original system by ignoring many of the differences between the states it can take, it then becomes isomorphic with the model, or with the part of the model we are observing. But it might be better not to attempt a description in such vague terms of concepts which Ashby has defined with complete clarity and precision. Suffice it to say that all the materials are by now available for someone with the requisite mathematical equipment to derive a systematic applied theory of replacement. The tools are there, and we commend the job to anyone competent to do it. He will be rendering a considerable service both to experimental biologists and to experimental animals.

Figure 5. The Concept of Isomorphism

(From Ashby, 1956a, Figure 6/8/I)

The two diagrams represent two kinds of machine, in each of which we can distinguish an input and an output.

In the upper machine, the input is represented by the rotation of the axle (I) at the left side of the figure. The position of this axle is shown on the dial (μ). The axle is connected through a spring (S) to a heavy wheel (M), which is rigidly connected to the output shaft (O). The position of (O) is shown on the dial (ν). The two dials thus show the input and the output of the system. The wheel (M) dips into a trough containing a liquid (F), which applies a frictional force to the wheel, proportional to the latter's velocity. This machine is therefore entirely mechanical.

The lower machine is electrical. Its input is a potentiometer (J), which emits a voltage shown on the dial (x). In series with (J) are an inductance (L), a resistance (R) and a capacitance (C). (P) is a current meter, recording the sum of the currents which have passed through it. This sum is shown on the dial (y). The two dials thus show the input and output of the system.

If the values of the components in the two machines are matched in an appropriate way, the two systems can *behave* identically. We can observe their behaviour by reading and comparing the input and output dials in each case. If the above conditions are met, *any* sequences of input which are identical in the two machines will give rise to identical sequences of output in both. If the central parts of the machines are covered, and only the dials are observable, we can observe only the behaviour of the machines. They will now appear to us absolutely identical over an infinite series of observations. We should have no means of deciding which was which. Yet these machines are totally different in respect of the materials of which they are composed, and of the physical properties on which their functions depend. Two such machines, which *behave* identically, however different in other ways, are said to be *isomorphic* to each other. Either could obviously be used as a perfect model of the other. Indeed, if we wish to study the behaviour of, say, the electrical one, it is absolutely indifferent which of the two we actually use. If it is desired to find a model for a third system, these machines would be of exactly equal merit for the purpose.

In the present context, it remains only to add that in physiology and pharamcology all we are ever interested in is the *behaviour* of a system, in this extremely general sense of the term.

¹The influence of the fallacy may, in fact, be important not so much among experimenters as among those who control their work.

²Isomorphism and homomorphism are, of course, old concepts in mathematics and logic; the novelty lies in their use in the theory of machines.