

THE RELIABILITY OF PSYCHOLOGICAL¹ EXPERIMENTS

LYLE H. LANIER
VANDERBILT UNIVERSITY

The purpose of this paper is to discuss some of the problems and methods of experimental psychology, by way of trying to clarify numerous obscure notions as to just what this science is. Psychology, perhaps more than any other science, suffers from charlatanism, mysticism and cheap sensationalism. Students on entering courses in psychology generally feel either that they are about to embark upon an erotic orgy or that they are going to be regaled with hidden secrets concerning the intimate operations of the soul—preferably both. The notion is rather general that psychology deals with a different order of phenomena from that studied in other biological sciences, that we are engaged in some mystical delving into that vague unworldly entity, the human mind. In comparison with these awesome expectations it is not surprising that students, and popular readers generally, are often disappointed with so-called "scientific" psychology. Thus the undergraduate student bodies of two of our largest universities have voted that psychology is the least valuable course in the college curriculum. The students want the psychologist to elucidate on peculiar sexual aberrations, while the obstinate wretch persists in trying to explain why they see things right side up. They want to participate in an exciting hypnotic seance; whereas the dull professor makes them listen to an account of experiments on conditioned reflexes by an unknown Russian. The fact is that only an extremely small minority of people have the tough-mindedness to settle down to learn the facts and methods of an experimental science. They want stimulation rather than education. Psychology disappoints them more, as a rule, than other sciences because they have expected so much of it.

Vague, mystical notions concerning human behavior are not confined to students, nor to the man in the street. The writer once heard a reputable animal ecologist state that he did not believe the psychologist's disposition of mental telepathy to be final. Another professor, not a scientist, once stated to a class of which I was a member that he believed mental telepathy to be an established fact. Such beliefs as these result, in an immediate sense, from loose thinking or from insufficient familiarity with the facts. But their more fundamental cause would seem to lie in the general tendency of present-day humanity to seek explanations of phenomena not in the phenomena themselves but in more unusual, imagined causal agencies lying beyond the pale of natural events. This tendency is a legacy of the Middle Ages, whose mysticism has obscured the tough-minded thinking of Aristotle by a cloud of

¹Presented before The Tennessee Academy of Science, April 27, 1929.

other worldliness. Another effect of this malignant influence is the acceptance of every-day objects as if they were perfectly understood and as if no mystery attached to them. This is sadly to confuse acquaintance with understanding. We have mysteries enough in every common object and in every human or animal act to satisfy anyone not deficient in imagination and not dependent upon sentimental illusions to guard him against what he conceives to be the materialism of science. The fact that I can communicate with a perfect stranger by means of language is just as mysterious and much more interesting to me than the improbable notion that I could sit in one room and reveal something to him by "thinking" it.

This is perhaps enough about what psychology *is not*. I shall now try to say what I think psychology *is*, and to relate my remarks especially to the question of the reliability of experiments, inasmuch as this is really the crucial problem for psychology as it is for every other experimental science. If we can secure consistent measurements of a given human reaction and if we can by the method of concomitant variation determine what antecedent event or group of events produce the reaction, then we are working on a scientific basis and can expect to formulate general laws describing our phenomena. A brief review of some of the experimental methods which have been used in psychology will probably make clear the general lines along which we may hope to see develop a fruitful science of psychology.

Experimental work on the psychology of testimony has shown us that reports of an event observed by several individuals will probably show wide variations in the details which each person claims to have noticed. I recall a recent case in Nashville where one newspaper reporter related that Lindbergh had hunted only bear on what was then his first trip to Mexico. In the other paper it was authoritatively stated that Lindbergh had greatly enjoyed his deer hunt in Mexico, and that he had hunted no other animal in that country. I think we may safely assume that the flyer himself was not responsible for this lack of agreement. Perhaps neither story was accurate, although we may reasonably believe that on this question of bare fact one of the reports was accurate. This sort of discrepancy is not unusual; it would occur whenever any two persons attempted to relate the details of a complex situation without any specific restrictions as to just what elements of it they were to notice. The early stages of experimental psychology were crowded with "introspections," many of which were probably as unreal as the inaccurate hunting story. A man would spend ten minutes writing a description of the sensations he claimed to have experienced during the brief span of one second. It is no wonder that such a method should have been repudiated, for it is practically impossible to check up on the consistency of such reports. In fact, about the only legitimate use of the introspective method is under conditions where the subject has to observe only one stimulus or relationship and to make a specific statement about it, preferably in categorical terms. Thus a man may be instructed to observe the brightness of two lights, presented either

numerical score for each of the two series is 50. Another individual is relatively unpracticed at the beginning of the experiment and gets a score of 40. But in the second series he has improved to the point of earning a score of 60, and hence of surpassing the person who originally outranked him. The score of 60 may be misleading, in that we don't know how much higher his score might be if he practiced the performance to his physiological limit. Even where the subjects are, roughly speaking, equally unpracticed, different learning rates may produce apparent unreliability in the results if we correlate scores based upon cross-sections at two different stages in the learning process.

Another factor which may produce unreliable results, judged by the "reliability coefficient" as explained above, is what we designate as the "attitude" of the subject. This, of course, is a resultant of numerous conditions, but the specific form in which it usually affects an experiment is for the subject to concentrate as completely as possible on the reaction required at one time and to fail to do this at another. In studying the reliability of certain tests of sensitivity to tonal relationships, I found that frequently individuals who made the required discriminations remarkably well at the first trial would make scores relatively much lower on the second trial. Questioning these persons revealed that the monotony of the tones was such that a positive distaste for the experiment was felt at the second sitting. This naturally affected their discriminability. Incidentally, here is a good example of where the verbal report of the subject was valuable in throwing light upon the objective results.

This is perhaps enough to suggest the nature of what we may call our primary reliability problem, namely the development of experimental technique adequate to insure consistent measurements for a given subject. This problem, which is essentially an experimental one, is often confused with the question of whether or not the results from whatever number of subjects have been used are representative of the total population of which these individuals are a part. This we may call our secondary reliability problem, although it is secondary only in the sense that it cannot legitimately be considered until the experimental situation is such that accurate results can be secured. Unfortunately, many psychologists have been too uncritical in their eagerness to apply statistical treatment to any sort of numerical results. In applying statistical methods to observed quantitative data two assumptions must be satisfied if the work is to mean anything. The first is that each individual observation be accurate; statistical method cannot compensate for data which are false. And secondly, there must be a representative sampling. When these two requirements are satisfied then it is pertinent to ask what the probabilities are that the obtained values really represent the total population of which the limited sampling studied is a small part. For nothing is truer than that a few cases, however well measured and however carefully selected, can only show us the general direction in which we expect to find the solution to our problem. Such limited results are not unscientific, necessarily, but they

do not carry us very far in that direction of precise quantitative generalizations which would appear to be the ultimate aim of every natural science.

A further limitation to the use of statistical method might be noted. It has been found that many types of psychological measurements are distributed about their several measures of central tendency in such a manner as to conform approximately to what is known as the normal probability curve. Now this curve is bilaterally symmetrical with respect to its mean, the slope of the curve at the mean is zero, and, of course, measures of variability from the mean include certain fractional amounts of the total area of the curve. It has been a frequent custom in psychological research to calculate the standard errors of the chief statistical values obtained from an empirical distribution and to assume that these standard errors defined limits within which such statistical values would be expected to lie in future experiments. This procedure is only justified if the measurements in question would actually be distributed in the form of the normal curve when an infinite number of cases were observed. Furthermore, it must be assumed that any future sampling of measurements would differ from the present one in no constant manner. "Fluctuations due to bias, due to the absence of random selection in the sampling, due to persistent errors of any sort, quite elude this method of determining probable stability (Mills, F. C., *Statistical Methods*, p. 560). These cautions are not always observed in our zeal to place our science on a quantitative basis and to make rather far-reaching generalizations from relatively few cases. The best way to study the reliability of the results from a given sampling of subjects is to secure a number of similar samplings and to study the *actual* fluctuations, their amount and direction.

In conclusion it would seem that we have a lot of prolonged laboratory work ahead of us in psychology if we want to do anything more than get on the trail of facts. It is, of course, legitimate to get on the trail of facts, for unless we find out where the facts are, we can never expect to measure them. There is precedent enough for this rough preliminary type of hewing out of pathways in our two most exact sciences, physics and chemistry. These sciences did not emerge full-fledged into their present quantitative status. Cavendish, in the eighteenth century, found the resistances of different materials by comparing the shocks he received in his arm when conductors of those materials were used to connect him with a battery. This was not unscientific, although it certainly was not exact measurement. Similarly, the greater part of the experimental work done in psychology during the past eighty years has concerned more the discovery than the measurement of phenomena. Some of this work appears now to have been useless, for reasons which we have already given. But even the rankest sort of introspection has often served the useful purpose of locating and defining new problems. From this viewpoint there is a continuity of development in our field which is often obscured by the numerous theories which have afflicted it. The theories about

psychological phenomena frequently baffle the lay reader who wants a concise general statement of what it's all about. He soon loses his way among pretentious dogmas such as the following: structuralism, functionalism, Freudianism, dynamic psychology, behaviorism, purposive psychology, configurationism. Indeed, the lack of agreement has so befuddled that boisterous representative of America's intellectual adolescence—the American Mercury—that it takes refuge against the necessity of thinking it out by saying that psychologists don't know what they believe. It requires only a little historical perspective to show that these theories are only ripples on the surface of the science. In so far as they stimulate research by suggesting new problems and methods, they are valuable, but 100 years from now they will be as barren as Thales' physics.