MODELS OF CONTROL AND CONTROL OF BIAS

Martin E. Johnson
University of Utrecht

MODEL 1: Lowest degree of intersubjective control.[1]

The experimenter carries out the investigation according to this model under such conditions that no relevant control securing intersubjectivity is possible. In its most extreme form, this is manifested by the omission of raw-data in the published article. The only type of scrutinization that is possible, from the standpoint of the reader or the critic, is to find out, if for instance, the statistics that have been worked out, are reasonably appropriate. In addition, one can try to find out whether the conclusions drawn by the author are really a logical consequence of the presented data. The reliability of the results can not be assessed by way of testing. The strength of the data depends on the reliability of the investigator, but unfortunately there does not exist a good method or indeed instrument, for such a measurement; when we are confronted with such a report. There is no available method, except perhaps independent replication, by which to judge the data as to whether or not there are motivational errors, re-stated objectives and hypotheses, selective reporting etc. An editorial "doctoring" of the data may also have taken place.

Commentary

Most exploratory studies in psychology as well as in parapsychology are carried out according to this model (see Figure 1).

In some studies within the behavioral sciences -including parapsychology, two supposedly independent experimenters take part. This procedure in experimental parapsychology is sometimes referred

1) "Intersubjectivity" is a concept of great methodological importance in science. It controls personal bias and provides an answer to the question: "Do you see what I see?"

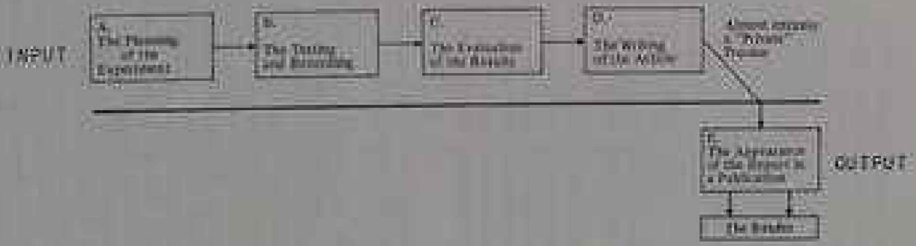

Fig. 1 Model I. Lowest Degree of Information Control in a Publication.

so as the two-experimenter plan (Rhine & Pratt, 1962). This procedure is of course intended to enhance the "intersubjectivity" of the observations and also implicitly to make it a bit harder for the main-experimenter to cheat, or to "doctor" his data. It is however questionable, as to what extent the accuracy will be improved by having two experimenters instead of one. Under certain conditions, for instance if they have opposite expectancies, it seems reasonable to assume that their bias as observers should tend to cancel itself out. If however, the two experimenters share expectancies or are biased in the same direction in other respects, the improvement as regards the intersubjective control is rather questionable. One can also question the gain in intersubjective control with the use of two experimenters if:

a) the main-experimenter has been allowed to select his co-experimenter;

b) the main-experimenter is more senior than his co-experimenter.

Finally however, it should be stressed, that in practical life the use of two more or less independent observers is supposed to enhance the accuracy of the observations they make. Just as an illustration, think of an airline captain and his first officer, how they carry out the checking procedure by reading "against" each other.

MODEL 2: A recording and evaluation device is added to control unintended errors on the part of the experimenter.

This model of control is relevant to the case when the experimenter is aware of the potential risk of making errors of observation and recording. Two types of "personal" or human errors should be considered; one type (Type I) occurring without the awareness of the experimenter (unintentional errors); the second type (Type II) being pertinent to the case when the experimenter intentionally or deliberately manipulates the outcome of the experiment. Model 2 seems to be especially relevant for the Type I category. Its value for the control of Type II error seems to be very limited. In its simplest form, the only difference between Model 1 and Model 2 is, that in the case of Model 2, the experimenter utilizes equipment that automatically records the subject's calls or responses. By way of addition the recorder can be equipped with some kind of memory-unit that makes it possible to evaluate automatically the total number of hits that a subject has obtained during a test-session. A further sophistication of the automated procedure would be to use a programming-unit by means of which targets could be generated and recorded. In the latter case, one could consider using a computer for the generation of targets as well as for the

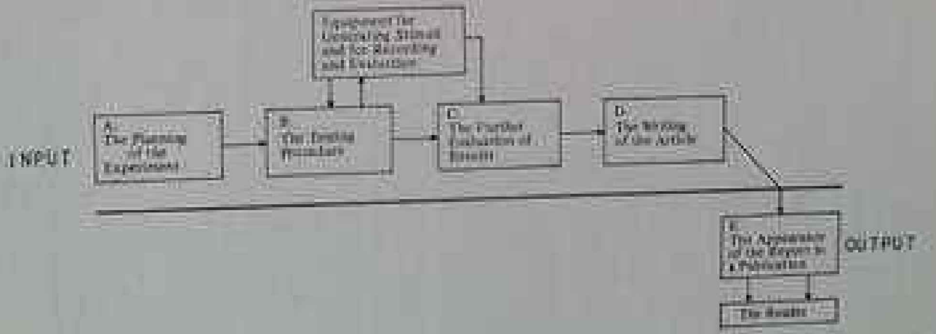

Fig. 2

Model E. A Recording and Evaluation Design is Utilized in order to Control Unintended Motivational Errors

recording of a subject's responses, and for the automated evaluation of hits.

Commentary

This model basically applies to the case when the deception hypothesis is not seriously considered (for instance under very exploratory conditions). The main advantage of the technical devices would be to reduce unintentional human errors in the procedure of recording and evaluating data.

In addition, it would relieve the experimenter from the burden of making certain observations and notations.

This model applies to the type of precautions which have sometimes been taken in more recent and advanced experiments in parapsychology. The model also depicts the types of safeguards against human errors that are frequently used in experimental psychology and other areas of experimental research. An inspection of the "critical" links or "synapses" of this model, will however, inform us, that this outline for the control of human errors does not help us very much in controlling fraudulent behaviour on the part of the experimenter (see Figure 2). This is especially true if the experimenter is also responsible for handling the equipment and for reading the obtained number of hits in his experiment.

MODEL 3: Optimized automated and intersubjective control of a psi-experiment.

As has previously been witnessed there are a number of potential artifacts that may influence the outcome of a scientific study, from its outline up to the stage of reporting the results in a scientific journal. Further, it has also been shown, that deception on the part of the experimenter may not be such an infrequent phenomenon as is usually thought.

Even if I believe that the "fool-proof" experiment is somewhat of an Utopia, I nevertheless think that it is not only possible, but extremely important, to develop methods involving a wide range of precautions against artifacts related to "human errors" or human limitations, for want of better terminology. My position does not suggest that I believe there to be only one scientific method - the one? I am quite aware that all methods rest upon some assumptions as to the nature of "reality" and how to investigate it.

It is a truism to state that scientific methods are invented by man. But according to certain criteria, methods may vary in their sensitivity to artifacts. May I remind you, that deception in science is not a phenomenon restricted to parapsychologists but

most certainly applies also to psychologists as well as scientists in other areas of research. Based upon what has previously been said, I am quite convinced that it is a very important strategy of research to try and do everything we can to control and indeed eighten up the designs of our experiments in order to avoid the potential influences of "human artifacts" of all conceivable types.

I therefore suggest that we should try to analyze and identify the critical events and synapses of our experiments, from the very first outline to the published report.

I believe that in this context, the use of the system approach could be very efficient and appropriate. As a matter of fact, we have already used this method in our discussions and illustrations of what here have been described as Model 1 and Model 2. As regards Model 3, the objectives should be:

a) the model should offer the maximum possible automated control, out of reach of possible manipulations on the part of the experimenter;
b) the model should offer the maximum possible "public" control, or in other words, a maximum of intersubjective testability;
c) according to the philosophy of this model, the experimenter should define his problem, formulate his hypotheses, and outline his experiment, prior to commencing his study. He should write his manuscript, stating at least essential facts, before carrying out his investigation. This manuscript, in principle only lacking data in the tables, presentation of results, and interpretation of results, should be sent to one or more editors, and the experimenter should not initiate his study until at least one of the editors has promised to publish the study, regardless of the outcome of the experiment. In this way we could avoid selective reporting. Furthermore the experimenter will not be given the opportunity to change his hypotheses in such a way that they "fit" the outcome of the experiment;
d) a "neutral" agency should be responsible for generating the targets to which the subjects are supposed to respond. The same agency or institution should record the responses made by the subjects. Finally, the controlling agency should be responsible for the evaluation of raw-data, for instance the evaluation of the total number of hits.
Such a "central" agency could be a computer centre.
At the experimenter's laboratory there should be a computer terminal, through which the subjects are able to "communicate" with the computer centre.
Before the experiment commences, the experimenter should request and specify what kind of assistance he will need from

the head of the computer centre at a particular time (number of random "events", probability ratio, etc.).

Such a request should follow a certain routine. The head of the centre should in such a case, after evaluating the outcome of the experiment, fill in a form or empty spaces planned for tables, sent to him by both the editor and the experimenter when the services were requested. The form, with the evaluated raw-data should be duplicated.

The head of the computer centre should keep one for himself and one copy should be sent to the experimenter, the other to the editor.

Commentary

One can always claim that the experimenter could have "bribed" the head of the computer centre to take part in a well-organized conspiracy, which might also include the co-operation of one or more of the editors. In my opinion the probability of such a hypothesis, from a common sense point of view, should be much smaller than the probability of an experimenter exhibiting deceptive behaviour under conditions when nobody can really control him. Assuming one can state that the hypothesis of a conspiracy is irrefutable, and since that is the case with all the discussed models, what is then the sense of making the experimental procedure more complex, time-consuming, and more expensive? My answer to this question is that one should notice carefully that not all empirically irrefutable hypotheses have the same probability. It should also be stressed, that in principle it should be possible to engage a great number of computer centres which can reserve time for a certain experiment. At the time of the experimenter's commencement, a random procedure, out of the reach or control of the experimenter, could determine which of these computer centres would take care of the requested services.

In this way, the possibility of obtaining positive results by way of a conspiracy becomes still more improbable.

Another type of criticism that I anticipate, as regards experiments carried out according to Model 3, concerns the idea that well-controlled experiments should always exert a devastating influence on the gullibility of the subject or the experimenter-subject dyad, by imposing an "unfavourable" atmosphere. This kind of criticism seems to me to lack both sophistication and justification. If the experimenter himself accepts to work according to Model 3, and if his subject remains ignorant of the location where the targets are generated, what "psychological" impact could that have on the subject's scoring performance? It

should be noted that nothing in the laboratory setting will be allowed to reveal whether or not the targets are produced within the laboratory or externally. Especially in the precognitive experiment (a test of precognition), the above-mentioned type of criticism seems to be quite irrelevant. "Model 3 conditions" actually relieve the experimenter from the rather dull duty of making registrations, while leaving him free to concentrate on creating a good "rapport" with his subjects, often thought, but scarcely proved, to have a veb-cular effect on a subject's paraability. It is very probable that certain subjects would find this type of experiment to be of a very challenging nature.

I would like you to notice that I have never said that all studies in parapsychology should be carried out according to Model 3. Indeed, what I wanted to say is that I find it extremely important to identify the critical points or synapses in our experiments, to be able to control potential bias. I believe however, that some experiments in parapsychology, carried out by naive "verificationists" and thought of as "conclusive", should be subjected to the rigours of the discussed model. We have to confess that we still find ourselves in a "pre-paradigmatic" period of research (to use one of Kuhn's terms), in which we primarily involve ourselves in fact-gathering or at least data-gathering.

That means that there is also place for the use of more exploratory types of studies. It is a pity that there does not exist a scientifically sound formula by which we can decide how to order the priorities as regards application of approach methods. On the other hand, there do exist very general and well-validated research ideals to which we have every reason to adhere: we should always do everything we can to obtain the highest possible degree of intersubjective control in our studies. This also applies to case studies in parapsychology.

The positions taken in this part of the inaugural address are considered as some of the hallmarks of this journal. We will try to avoid selective reporting, that is, the tendency to bury negative results and only to publish studies that "turned out". For one thing we believe that there is a chance to learn something important even from negative findings; for another the policy of exclusively publishing "supporting" findings even if such a policy may have a rationale or at least easily can be rationalized, may probably exert a strong temptation to "doctor" ones data.

At the same time we do not want to make our journal a graveyard for all those studies which did not "turn out". This can be managed by adhering to the editorial policy that the acceptance or

rejection of a manuscript should be done on the basis of the quality of the design and methodology of the planned study and prior to the carrying out and the evaluation of the results of the study.

ABSTRACT

The author discusses how to increase the quality and reliability of the research and reporting process in experimental parapsychology. Three levels of bias and control of bias are discussed. The levels are referred to as Model 1, Model 2 and Model 3 respectively.

Model 1 is characterized by its very low level of interobjective control. The reliability of the results depends to a very great extent upon the reliability of the investigator and the editor.

Model 2 is relevant to the case when the experimenter is aware of the potential risk of making both errors of observation and recording and tries to control this bias. However, this model of control does not make allowances for the case when data are intentionally manipulated.

Model 3 deploys a rather sophisticated system of control. One feature of this model is, that selective reporting will become harder since the editor has to make his decision as regards the acceptance or rejection of an experimental article prior to the results being obtained, and subsequently based upon the quality of the outline of the experiment. However, it should be stressed, that not even this model provides a fool-proof guarantee against deliberate fraud.

It is assumed that the models of bias and control of bias under discussion are relevant to most branches of the behavioural sciences.

ACKNOWLEDGMENT

This article constitutes a part of the author's inaugural speech as professor of parapsychology at the State University of Utrecht, The Netherlands. The speech was delivered 8th November, 1974.

REFERENCES

Rhine, J.B. and Pratt, J.G. Parapsychology. 'Frontier Science of the Mind'. Springfield, Illinois, C. C. Thomas Publisher, 1962.