REPLICATION OF EXPERIMENTS

by

Philip J. Runkel

April 1994

The conception of replicability of an experiment common to physics and chemistry is a conception that cannot fit living things. Consider for example a simple experiment on the bending effect when a force is applied to, say, an aluminum beam. You would support the beam at both ends, put a weight on the middle of the beam, and measure the distance the beam sags. You could ask others to do the same thing and see whether they observe the same sag. They should, of course, examine no more and no less than you did. First, they should make sure they are measuring the behavior of the same "thing." That is, they should use the same beam you used or one very much like it. (To get a beam sufficiently like yours, maybe they could get one made from aluminum taken molten from the same vat and rolled or extruded into a beam by the same machine.) Second, they should make sure that the internal conditions of the two beams are the same. Neither beam should have undergone violent stresses that might have produced cracks from "metal fatigue." The second beam should be at the same temperature as the first. Third, the environments should be as close to identical as possible. The supports for the two beams should be spaced at exactly the same distance. The second weight should have exactly the same poundage as the first. There should be no strong winds that might produce an aerodynamic force on the beam or the weight. As long as the replication is being conducted here on earth, I can think of no other necessary precautions.

Notice how many kinds of events I have omitted. I have left out everything not strictly physical. We do not expect either beam to have any preferences about the amount of weight put on it. We do not expect either beam to give a little and then refuse to sag any further. We do not expect it to slump because of the conversation of the laboratory assistants, the music the janitors played during the night, or the political situation. I also omitted a great many physical events and conditions, because physical theory is very clear about what might make a difference and what would not. I paid no attention to cosmic rays or to the differences in air pressure between stormy and sunny weather. I assumed some effects to be too small to bother about: the differences in gravity from one place to another on the surface of the earth and the effects of differing degrees of corrosion over a period of a few days or even a few weeks, for example.

Replicating this simple experiment is straightforward. One can easily specify steps to make sure (1) that the same kind of "thing" is being examined, (2) that its internal condition is the same, and (3) that the environmental influences are the same. No matter how complicated a physical experiment may become, it is always possible, by the very nature of physical theory, to specify some feasible steps by which the experiment can be replicated. The first assumption in physical theory is that if a certain kind of force can be seen to act upon a certain kind of thing with a certain result, then it can be seen at another place and time. If you observe, in Peoria today, an aluminum beam of certain dimensions sagging by a quarter of an inch under a weight of a hundred pounds, then I can observe, In Sacramento next week, the same thing. This beginning assumption can hold for the linear, sequential, episodic conception of physical reality because it is possible to specify a replicable small set of prior condition (input), a replicable small set of identifying characteristics of the "thing" being observed, and a replicable small set of consequent conditions (output). The prior conditions can be relied upon to hold still while you arrange the experimental input and make measurements, the identifying characteristics will stay substantially the same during the experiment, and the consequent conditions can be relied upon to hold still while you measure again--or you can choose points in the sequence of events where conditions do hold still long enough. The aluminum beam waits, unchanging, while you put the weight on it. Then it waits again, unmoving, while you measure the sag.

In certain respects, in ways of interest to football players and surgeons, among others, influences on living creatures have effects very like the bending force on the aluminum beam. If an automobile slams against a human body, the auto will knock over the person just as it will a fence post. If physical force overcomes the physical ability of the living body to oppose it, the body will act just as a nonliving thing acts. But most of us, most of the time, deal with others in nondestructive ways that do not overwhelm the abilities of other persons' bodies to cope with events. It is true that a great deal of maiming and killing do go on, every day. But I am limiting this discussion to the kinds of behavior and experience that leave the bodies of the participants capable of their full repertoire of behavior. We do often use our bodies for peaceful interaction, without maltreating or coercing the other person, when we join in cooperative activity or in other ways communicate directly. For example, we dance with partners and caress those we love, activities not at all like bending a beam or knocking over a fence post. Despite a great deal of bodily contact, however, most of us deal with one another most of the time by communicating orally or in writing. That is, we make sound waves in air for others to hear or we make letters on paper (or other surfaces) for others to see. We use light to communicate not only through writing and reading, but also when we make sculpture, paintings, and other forms of visual art.

Most influences between humans are not, I repeat, those of brute force--not pushing or pulling or bashing. They are influences to which the other person will "respond" or "react." The sequence of physical events inside a person who hears a sentence is very different from the sequence of physical events inside the aluminum beam when the weight is put on it. What the person does in response to the sentence is not capable of description by Newton's three laws of motion. What the person does depends on whether the sentence disturbs some internal standard--some perceptual variable the person is controlling.

Given the fact that behavior is always motivated by the interaction of the person and the environment--by the resultant of the internal standard and the external disturbance to the perception--how can the internal condition of the person or the external condition of the environment by replicated? With living creatures, neither condition can be replicated. In respect to the variables that affect bending, the internal condition of the aluminum beam stays very much the same during the period of time that we care about. (The condition might change appreciably in a few hundred years; if we were to separate replications by that much, we would protect the beam from the weather, from marauders, and so on.) But the internal condition of the person, in respect to the variables the person cares about, is in constant flux. Think, for example, of the way your mind leaps from one possibility to another as you talk to the clerk in the shoe store about the shoes you look at and try on. Or think of someone at work inviting you to join with others in some venture; what further information do you invite or discourage as your picture of the opportunities and dangers changes? Or think of shaking hands with a person who seems to want to show you that his grip is stronger than yours. Every snippet of our lives is full of reassessments, adjustments, adaptations, revisions, and corrections of course. We are never in the same condition we were in yesterday or even a moment ago.

At every moment, the brain and other parts of the nervous system are busy controlling perceptual variables. As action continues, opportunities arise in the environment for the person to bring other variables nearer to the internal standard. An action to improve the match-to-standard of one variable may disturb the match of another, and that discrepancy will await repair. Though the person may be taking no overt action, the person's condition of vigilance (alertness, sensitivity) will change continuously. The person's internal condition will not necessarily change when some external event occurs that would interest a friend, spouse, employer, teacher, police-person, or experimenter. It will change, rather, as the person deals with any and every disturbance to an internal standard of his or her own. It will change as unforeseen disturbances turn up, as the person acts upon the environment, and as the person alters, internally, his or her understanding of every new situation and every new opportunity. The person's array or constellation of controlled variables that are and are not at match with their internal standards is therefore changing continuously. As a result, no matter with what meticulous, scrupulous, and thorough care an experimenter may work to produce a replication of an earlier environment, the experimenter can be sure only that the effort will not succeed.

The conditions in the environment that will spur the person to action are those that (1) disturb some controlled variable or (2) offer the person a means through which to oppose a disturbance while not worsening a disturbance of some higher-order controlled variable. But an experimenter can never be sure, in the case of a person even minimally free to act in an environment furnished with even a minimal variety of means for action, what variables the person will be controlling <u>this</u> time. Therefore, the experimenter cannot be sure whether the presumably same environment will offer the means for pursuing the purposes the person will this time want to pursue, nor can the experimenter be sure whether the means the environment offers will facilitate acts disturbing to other variables the person is controlling. In short, action arises from the interaction between environment and internal standard, and because the experimenter cannot know the variables being controlled by the person at any given moment, the experimenter cannot know whether some environment (in the sense of a place furnished with objects) is, in any effective way, the same for a person at two separated times.

The fact that a place changes its "meaning" as one has various experiences in it is a fact long celebrated by poets and story-tellers. We behave one way in the house of a stranger and in another way after we have come to know the householder well. We are alert for a certain range of possibilities when walking through a city park for the first time and alert for a different range after having sat on a bench there to eat lunch on several summer days. Our readiness for a certain range of acts is heightened partly by the acts we are able to conceive at any place and time and partly by the kinds of acts of which we are reminded by the place and time in which we find ourselves. And the place or time reminds us of some possibilities and not others to the extent that we find it matching our memory of it. If we knock on a door only to find that our friend no longer lives there, our readiness to take some acts drops and to take others rises. The changing "meaning" of a place as our experience with it changes is a simple idea, but one that seems sometimes to evade the minds of experiments who undertake to set up the same conditions at two separate times or for two or more persons.

Most experimenters, however, are at least sometimes aware of the changing meanings of environments. Experimenters with rats, for example, customarily let a rat become familiar with a maze (to satisfy and thereby reduce its urge to explore a new place) before putting the rat through the test runs. And experimenters who put subjects through two or more experiences will often order the experiences in one sequence for some subjects and in another for other subjects (in a "balanced" design) with the idea of randomizing the experiential effects.

Indeed, books on research method almost always tell the reader to reduce the biases that can arise from environments or sequences of experience by assigning subjects randomly to places and times and to draw conclusions from the averages calculated over the subjects in the separate places and times.

That tactic, it is true, reduces bias between the averages. That is, when persons are assigned to groups randomly, the probability of an average being wrong in one direction becomes equal to the probability of its being wrong in any other direction. For example, the probability of the average of the measures taken from the people in this place being greater than the average taken from people in that other place becomes equal to the probability of its being smaller. That's nice, but biased or not, the probability of an average or a difference between averages being at least to some degree wrong in one way or another is always very high. Experimenters sometimes speak of "equating" two groups by randomizing the assignment of persons to them--as if randomly assigning persons somehow makes the averages come out equal. It does not.

But the unreliability of averages is really beside the point. The point is that assigning persons to groups, randomly or nonrandomly, and calculating averages does not reduce the effects of the internal processes I have mentioned on the behavior of the individual; it does not reduce the continuous changes in readiness, meaning, and the like. It does not reduce the continuous changeability of the person's internal condition. It does not bring us closer to replicating the conditions, internal or external, in which individuals find themselves. It does not enable us to study how individuals cope with those changing conditions.

Taking an average over individuals yields a number uninterpretable in regard to any individual person. Am I to think that the average characterizes any individual in the group? No, I should not think that. If one person measures .26 of a second in reaction time and another .56, the average is .41--which is the reaction time of neither of them. Indeed, averages cover up or neglect information that is often crucial. For example, one person may be running north, another east, another south, and another west, all at the same headlong speed. The average of those vectors is zero--no motion at all. But to report that the average movement of the group is zero is to draw attention away from what, if you had seen the individuals themselves, would have been the most noticeable thing about them; namely, that they were all in headlong, divergent flight. The average over individuals tells nothing about what is happening to any individual, internally or externally.

It is true that it is possible to produce very special and powerful environments. It is possible (1) to force a large discrepancy in an animal between what the animal perceives to be the present state of a vital process such as ingesting food and its standard for that state and (2) to impoverish the animal's environment so that there is only one feature of the environment the animal can use to reduce that very salient discrepancy. A very hungry rat in a Skinner box is the classic example. It takes only a minimal knowledge of animal behavior to predict that if a rat is extremely hungry, it will spend almost all its time hunting for signs of food. It will look at and smell and handle everything it can in search of any clue to food. When it happens to push on the lever sticking out of the wall and a food pellet pops into the little cup, no one will be surprised that the rat before long hangs around the lever instead of prowling elsewhere. The rat will pump out more food until it is no longer hungry. In brief, when you have good reason to believe that one discrepancy from an internal standard is overwhelming all others in the internal condition of the animal, and you have the environment arranged so that there is only one kind of act that will improve matters--produce food, in this case--you can then predict with a high degree of success that once the animal has discovered that kind of act, it won't be long before it performs that act again. (Among humans, actually, there are exceptions even to this.) Of course, the animal may, especially after eating the first few pellets, scratch its leg or defecate or run around the box before pressing the lever the next time, but the rat will continue to press the lever pretty frequently until it is no longer hungry. After its hunger is satisfied, then you are back where you started, unable to predict very well what act the rat will perform next, even in a severely impoverished environment such as a Skinner box.

Notice, however, even with the rat in a Skinner box, the experimenter is not actually predicting particular patterns of neural and muscular action. When the rat stands to the left of the lever, its muscles combine to move the lever in a configuration different from that required when the rat stands to the right. When the rat presses the lever with its chin, the bodily action is not the same as when it presses the lever with its paw. And so on. Rats always perform what we call the "same" acts in various ways. The experimenter is not predicting the complex of muscle action called forth by pulses in the efferent nerves. The experimenter is, instead, predicting purpose--any action directed toward a purpose, in this case any movement that presses a lever that in turn brings food. What all this comes down to is that the experimenter is predicting that when the rat is hungry, it will do what it can to get food. Aristotle knew that. So did Tutankhamen.

In sum, replicating an experiment with living creatures is impossible. We should expect it to be impossible the moment we realize that living creatures must and do maintain a continuing purpose (control a perceived quantity) despite an unpredictably varying environment--a fact that William James (1890, volume 1, pp. 6-8) insisted upon more than a hundred years ago. An experiment can approximate a replication only when the environment is simplified until it can almost be said to contain only one bit of information--lever pressed or unpressed--and only as long as one controlled quantity (such as hunger) overrides all others.

For me, the interesting question is not how we can achieve a proper replication of an experiment, but rather why anyone should have wanted to do so. If we want to find out whether we can construct an actual working model of purposive action, we must do so by following the action of an individual person in holding a perceptual quantity constant despite random environmental disturbances to it, and we must take observations about as fast as nerves and muscles can reverse their action, so that we can be sure whether our data-points are giving us a continuous record or merely a trend that might hide important deviations. If we want to model a hierarchy of control, we must observe the individual person to find out which variable the person holds constant at the expense of another. If we want to study the relative times required to bring a disturbed variable back to the standard value at various levels of the control hierarchy, we must compare the reaction times of the single individual. In every case, we must discover not whether one average over persons is greater than another, but whether any single person is capable of the control we postulate. And then whether any person shows up who does not make use of such control. Averages of groups are beside the point.

And we should not want to test whether a person will do exactly the same thing in exactly the same situation, because our theory tells us that the person is incapable of doing exactly the same "thing" as in any act in the past. When we say a person is doing the same thing as a moment ago or a week ago, we almost always mean that the person seems to us to be pursuing the same immediate purpose: opening a door, trying to persuade somebody of something, eating a pear, and so on. We do not mean that the person is tensing the same muscles to the same degree and applying the same force to the knob of the door; those are bound to differ from an earlier time because of strength added by exercise or subtracted by fatigue, because of a slightly different position in grasping the knob, and so on. We do not mean that the person is saying the same persuasive words or using the same energy to move the mouth in forming the words. The internal condition of nerves, muscles, glandular secretions, and so on must always differ from one moment to another. To test whether we can be confident that we have modeled the functioning of the human creature, we should test whether the model can maintain pursuit of a purpose for some reasonable span of time during unpredictable variations in the environment, since unpredictable variation is the constant condition of the uncontrived environment. We must demand of any model that it perform realistically in an unpredictably changing environment.

I'll say yet another word about the continuously changing environment. You may be thinking yes, it changes a lot, but don't some things stay very much the same for a pretty long time?

Maybe your bank hasn't had a remodeling in 33 years. When you walk in, you use the same number of steps getting from the front door to the teller's counter that you used the last time. You breath the same furniture polish. The same person smiles at you. And so on. Can you not use at least a good many of the same acts today that you used last time? Well, you can use many of the "same" parts of the environment, yes--such as the same path from door to counter and the same teller. But you cannot use the same particular "acts." On your last visit, you came into the door from the west; on this visit, from the east. So the centrifugal force on you as you came through the door is different, and that changes what your muscles must do to get you across the floor to the right teller. And you are not feeling as cheerful as last time, you do not welcome the teller's clarion "hello-o-o," and you do not act toward the teller in the same way. Regardless of whether the bank seems to a presumably objective observer (an experimenter, let's say) to stay the same from day to day, the bank can never stay the same for you. When you act in relation to the bank, you must always act anew.

You can think of these small differences in the manner of recent discussions of chaos. A butterfly's wing in California can effect a tiny change in air currents that combines with a small flow that nudges a cool air mass up this canyon instead of that, the air mass joins with others over the Rockies, and eventually rain falls on Denver that would not have fallen there but for the beat of the butterfly's wing. Similarly, a tiny difference in a curving path as one goes through the door of the bank can bring one to swerve a foot or two off last week's path and require another pound per square inch of force to bring one back toward the intended teller. A living creature must act constantly against sliding off into chaos.

We should want, therefore, to test whether the model can perform properly in a changing environment or in several kinds of environments, not in an unchanging environment. We should test whether we can model several actual individual humans. In brief, the wish to mimic physicists by replicating the subjection of the same or equivalent object or substance (an animal) to the same influences (stimuli) has been a terrible mistake. We should instead be observing how the living creature maintains its perceptual input constant while the environment changes.

And if replication is a vain hope, the physicist's idea of prediction is likewise one to be given up. Living creatures divert and negate the expected physical effects of physical forces. They do not react to one another (or to anything else) according to the laws governing billiard balls. The ordinary concept of prediction has this form: Under the condition X, we expect to see the action Y. But if we can never establish the condition X (except trivially), we should not expect to see the action Y in connection with any specifiable condition. And it is the case that we do not. We should not try to predict particular acts. I mean visible "acts" (observed without any thought of purpose) such as pushing on a doorknob , pouring water into the mouth, putting one foot in front of another, making an "X" in a little printed box on a piece of paper, and so on. We should instead predict that a person will pursue a purpose. We should then try to learn how the person manages to achieve it and what the ranges of success can be with various individuals. We should try to see how an individual relinquishes one purpose and takes up another. What helps or hinders individuals in making use of their environments in pursuing multiple purposes (which is almost always the case)? That is to say, we should learn the capabilities and limitations of human individuals so that we can predict the circumstances under which humans are likely to function competently (regardless of the particular acts they may employ in doing so) and the circumstances in which they will likely be stressed too much and function poorly, become incompetent, or even die. We should learn, too, the ways we can help one another function well and the ways we inadvertently (and advertently, too) hinder one another.