

# What Kind of Empirical Research Should We Publish, Fund, and Reward?

## A Different Perspective

Paul Rozin

University of Pennsylvania

---

**ABSTRACT**—*When evaluating empirical papers for publication, grant proposals, or individual contributions (e.g., awarding tenure), the basic question one should ask is how much the contribution adds to understanding in psychology and not whether the contribution takes a particular form or represents one particular model of how to do empirical studies. Academic psychology has flourished with its mastery of the hypothesis–experiment model of science and its expertise in generating and eliminating alternative hypotheses and isolating causation. These accomplishments are a critical part of psychology, and they are well and appropriately taught by psychologists. However, they are only a part of science and should not comprise the almost exclusive criteria for evaluating research. In particular, discovery of fundamental phenomena, such as functional relations that apply to the real world and have generality, should have a higher priority in psychology. Such findings have been the basis for theoretical advances in other natural sciences.*

---

Proposition 1: The principal aim of academic psychology is to understand how humans and animals behave, think, and feel and how these events influence and are influenced by their material and social environment.

Proposition 2: It follows that contributions to the psychological literature should, as a principal aim, serve the purpose of psychology as indicated in Proposition 1.

Proposition 3: It follows from Propositions 1 and 2 that potential contributions to psychology, as journal articles or grant

applications, should be evaluated principally in terms of the degree to which they advance our understanding. There also must be some acceptable ratio of contribution divided by either journal space or funds required.

These three rather simple and highly linked statements do not describe current practice in the field. Rather, grant support, acceptance for publication, and rewards (such as tenure) are principally dependent on methodological sophistication, clarity of conclusions, and direct advance in understanding of a well-defined laboratory effect (and, most recently, demonstration of activation of an area in the brain). This focus strongly privileges hypothesis testing, experiments, and sophisticated methodologies and statistical analyses.

There are two aspects or types of research in the natural sciences. The first type of research describes phenomena, and the second involves the creation and testing of theories to explain these phenomena (Haig, 2005). Phenomena can be defined as “relatively stable, recurrent, general features of the world” (Haig, 2005, p. 374). Historically, in the mature natural sciences, description of phenomena, including invariant functional relations between variables (such as pressure and temperature in gasses, degree of selection pressure and rapidity of evolutionary change, and time in the dark- and light-detection threshold), precedes and becomes the basis for theory and hypothesis testing. In general, in spite of the example of better developed natural sciences, psychology has demeaned description of phenomena and assessment of their generality and moved directly into hypothesis testing. But, of course, an hypothesis or theory is ultimately only as good as the importance and reliability of the events the theory proposes to explain.

This point has been made before, notably by Solomon Asch (1952/1987; see also Baumeister, Vohs, & Funder, 2007; Haig, 2005; and Rozin, 2001, 2006). Ed Diener presented a view very consistent with this position in his editorial stating the policy of

---

Address correspondence to Paul Rozin, Department of Psychology, University of Pennsylvania, 3720 Walnut Street, Philadelphia, PA 19104-6196; e-mail: rozin@psych.upenn.edu.

*Perspectives in Psychological Science* (Diener, 2006). Asch (1952/1987) made the basic point about the primacy of phenomena particularly well:

Before we inquire into origins and functional relations, it is necessary to know the thing we are trying to explain. (p. 65)

If there must be principles of scientific method, then surely the first to claim our attention is that one should describe phenomena faithfully and allow them to guide the choice of problems and procedures. If social psychology is to make a contribution to human knowledge, if it is to do more than add footnotes to ideas developed in other fields, it must look freely at its phenomena and examine its foundations. (p. xv)

There are abundant examples of important contributions to psychology that were basically descriptions of phenomena. These include the initial descriptions of the tip of the tongue effect, the amnesic syndrome, apparent movement, the dark adaptation function, the psychophysical law, the rewarding value of brain stimulation, the “bug detectors” in the frog visual system, and the laboratory demonstrations of obedience by Milgram. Three theories from outside psychology that have had major influences on the field were based on careful empirical descriptions, not experiment and hypothesis testing. I refer to Darwin’s theory of evolution, Watson and Crick’s (1953) theory on the structure of DNA, and Chomsky’s theory of syntax. Darwin was a genius at describing phenomena and extracting invariances. The four empirical papers cited by Watson and Crick as support for the extraordinary inference of the structure of DNA were not experiments and were not hypothesis driven (Rozin, 2001). They were just attempts to describe the structure of DNA. We do not want to reject a paper that documents a talking horse because the authors cannot identify what part of the particular training procedure was critical in producing the effect or whether the particular horse in question was a genius.

I teach large classes of introductory psychology and have done so on and off for over 30 years. It just came to my attention last year that the course I teach is labeled “Introduction to Experimental Psychology.” The name was adopted before I arrived at the University of Pennsylvania in the early 1960s, as my department was becoming much more “hard science” oriented. In fact, much of what I actually teach in my course does not qualify as being experimental. One of the hallmarks of modern academic psychology is its methodological sophistication: a focus on hypothesis testing, controls, multiple measures, consideration of alternative accounts of results, careful and often sophisticated statistical analysis of results, and distinction between inferences about cause versus correlation. It is my sense that psychology developed these features of good science more so than did any other academic field in the 20th century. These methodologies and rigors are in large part responsible for the deserved success and advances in psychology and for the growth of the field. These same methodologies are poorly ap-

preciated by intelligent laymen (especially the generation of alternative accounts and the distinction between correlation and causation). My introductory psychology course focuses on the development of critical thinking: understanding the nature of evidence and the need for controls and experimental design, learning the value of the generation and testing of alternative accounts, and distinguishing correlation from causation.

It is perhaps true that it is most important for us to teach critical thinking and the type of natural science we have developed to undergraduates in psychology. But it does not follow that this should be either the central focus of graduate study or the principal criterion for judging the quality of research.

In fact, in the course of developing a very sophisticated science of hypothesis testing and experiment, we have almost forgotten the important precursors of these activities. In advanced sciences, and in particular areas of psychology where there is a strong background in phenomena and a history of hypothesis testing and experimentation, it is entirely appropriate to emphasize hypothesis generation and experimental tests. In such cases, as with the  $n$ th study (where  $n > 10$ ) on a particular phenomenon or claim, it is appropriate to determine whether proper controls have been conducted, whether alternative accounts have been dealt with, and whether there are any errors in thinking or experimentation. But first, we have to find out what it is that we will be studying, what its properties are, and its generality outside of the laboratory and across cultures.

Diener (2006) focuses on one of the major problems with our current science: our focus on using faults as a major criteria when rejecting papers. Both Diener and my third proposition hold that although faults detract from the value of a paper, they can be compensated for by novelty, by opening up new problems, or by providing different perspectives.

There is a strong psychological force working to promote an error-detection focus in evaluation. We like to be able to objectify our decisions, and it is much easier to point to methodological errors as a reason for rejection than to matters that may be more disputable, such as judgment and taste. But, as Diener (2006) points out, we have to work to overcome this bias.

Our current obsession with faultless experiments comes with its own scientific shortcomings: We are almost indifferent to the source of samples (usually students at major universities who take introductory psychology). We apply the rules of “scientific hygiene” (which I believe are not actual descriptions of what physicists, biologists, and sensory psychologists actually do) indifferently to the first and 100th studies of the same effect and equally to studies in which there is a precious set of 20 difficult-to-find individuals versus 20 undergraduates.

The best formula for professional success in psychology is to first establish a phenomenon, preferably one with wide generality across populations, and identify the contexts under which it appears. We then can do many studies, and get many grants, to analyze the mechanism of the effect in behavioral, mental, or

neural terms. We are not rewarded for looking at the generality of the effect. Is it a fragile result of a carefully selected set of parameters? Or is it robust and operative across many situations and/or populations? It is often the case that an experiment demonstrating a mechanism or effect is the result of many pilot studies in which a set of optimal parameters are selected to highlight the finding in question. This is good, normal science, but it is not good, normal science to fail to reveal the history of parameter selection (of course, journals would not be receptive to the space required for this), nor is it good, normal science to discourage investigation of the populations, contexts, and parameters that allow an effect to appear at the expense of almost exclusive focus on discovery of the mechanism of a highly selected effect.

I offer an example from my own work: Three journals in the last 5 years rejected a study I was involved with in which we did 2–3-hr structured interviews of 29 American Holocaust survivors in an attempt to describe their current attitude to contemporary Germans. We used a careful binary choice procedure (not criticized in any of the reviews) through which survivors would indicate whether they were comfortable or not in a set of about 30 situations (e.g., hearing the German language spoken, living next door to an American of German origin whose family migrated to the United States before World War II, riding in a Volkswagen). We found a continuous range of response: Some of the survivors had no aversion (discomfort) to anyone but Nazis, and some were averse to anything German, including the grandchildren of Germans who were not active in World War II and the idea of riding in a Mercedes. We thought this enormous range of response to perhaps the greatest trauma of the 20th century was of note, and not easily explained by current psychological theories. (We reported a few correlations that suggested some possible causes and made other causes less likely, but we were limited by the  $N$  of 29. “Degree” of trauma in the camps did not correlate with aversion, but Jewish religiosity and a non-German prewar background enhanced the probability of having a widespread aversion.) Reviewers thought this was interesting, but they wanted a more complete process analysis. One form of the paper was eventually published (Cherfas, Rozin, Cohen, Davidson, & McCauley, 2006). The analysis legitimately desired by reviewers required a bigger sample from a very rare breed of person. We thought the phenomenon was of interest, particularly because it presented a serious problem for psychological theories and dealt with something in the real world.

To me, one of the most valuable and elegant contributions of the last 2 decades in social psychology is the work on the culture of honor by Nisbett and Cohen (1996). These authors started with the idea that some individuals, particularly those in the southern United States, hold personal honor as a particularly valuable characteristic that should be defended aggressively when threatened. In a classic book, Nisbett and Cohen set out to describe the culture of honor and identify the populations that displayed it. Their survey of the literature and recruitment

of relevant statistics from a number of sources, including sociology, history, news reports, and government records, led them to define the focus of the phenomenon: white, American, Southern males primarily outside of major cities. Notice that they did not start with undergraduates, but rather took a broad sociocultural view and defined and localized the phenomenon of interest. The rest of the impressive initial work on culture of honor (Nisbett & Cohen, 1996) brought to bear the many sophistications and the ingenuity often exhibited, and beautifully developed, in experimental social psychological research. Field experiments showed the expected South–North difference in newspaper reporting and hiring. The authors established that undergraduate males at the University of Michigan of Southern or Northern origins displayed the critical difference that allowed them to more easily study and define the manifestations of the culture of honor.

Physiological and behavioral measures in a controlled experimental context defined the phenomenon in more detail. When one finishes reading *Culture of Honor*, one has no doubt that this set of attitudes and behaviors really exists and in only part of the population. Two things stand out to me about this work (other than its cleverness and importance): (a) It began in the real world, and then went to the laboratory, and (b) it is not a search for mechanism (up to the point of the publication of the book), but rather a careful description of the features and extent of an important phenomenon. (Nisbett & Cohen, 1996, offer a cultural-evolutionary account for the culture of honor in their book, but the work stands on its own as description of phenomena whether or not one subscribes to their theory.) *Culture of Honor* fits in with Darwin and Goffman more than with the almost-exclusive type of publication we see in modern social psychology.

I think the following are some reasonable criteria that we should adopt for grants, publications, and awards:

1. There are many ways to directly or indirectly advance our understanding of psychology.
2. Length of contribution is itself not a criterion. The longer the paper, the more it should accomplish.
3. Elegance and clarity are criteria for publication, but there should be a trade-off with novelty and engagement. Elegance and clarity are important in the service of a well-established goal, but one should be pointed in a worthwhile direction first, and it is through phenomena-oriented papers that we attain this direction.
4. The critical criterion, as stated in Proposition 3 above, is a contribution to understanding.

I list below some types of empirical contributions that could substantially increase our understanding and that should be publishable in our journals, supportable by grants, or constitute a good case for tenure. My evidence comes principally from my own experience with hundreds of papers as an author and many hundreds as a reviewer and editor.

I am not including the standard experimental paper (often a multiple experiment paper), which refines our knowledge about a demonstrated laboratory phenomenon by varying parameters, adding methodologies, and/or testing different theories. Nor am I including important methodological papers.

1. “Here’s what happens in the world.” This paper consists of raw description, carefully documented, and motivated by what I will call “informed curiosity” (Rozin, 2001). Ethologists do a lot of this, as did Erving Goffman and Darwin. Much of molecular biology takes this form. In my own history, I had great difficulty publishing a paper reporting that virtually everyone in a Mexican village over 5 or 6 years of age liked the burn of chili pepper, but that none of the animals in the village showed a preference for it, even though they ate the pepper daily as they consumed the leftovers of the day in the garbage (Rozin & Kennel, 1983). This nonobvious and previously unappreciated finding turns out to be important in understanding the conditions under which innate aversions are reversed.
2. “Here is a functional relation between two variables.” How many functional relations do we have in social or developmental psychology (other than performance as a function of age) in which we look at multiple levels of the independent variable? Instead, we find a pair of conditions that produce interesting differences and test for process—a natural consequence of reliance on analyses of variance designs. The U-shaped relation between duration or arousal and many phenomena in psychology can only be discovered by systematically varying the independent variable. Far more often, we reward finding one combination of independent variables that shows a big effect and examining it further with an experimental process analysis. Functional relations, such as Boyle’s law in physics and the dark adaptation function, constitute the core of basic empirical work in the natural sciences.
3. “Here’s something interesting that no one has noticed, and it is not easily susceptible to explanation by the principles available to us.” My paper on the German aversion of Holocaust survivors (discussed earlier in this article) is a good example of this kind of paper.
4. “Here’s something we haven’t studied, but it looks like it can be subsumed under something we already know.” That is, we report, for example, that contrast effects on the skin can be accounted for by lateral inhibition, which was developed in the study of vision, or we can even claim that all of these various contrast phenomena, which we may not yet have a mechanism for, are probably accounted for by the same mechanism. The similarity in nucleotide base ratios across species was one of the most critical observations that led to the Watson–Crick formulation. Generally, this type of paper establishes an analogy between one problem and another.
5. “Hey, someone did this really interesting study decades ago, and no one seems to have noticed it.” This paper would call readers’ attention to something already in the literature that is important and unknown or ignored. One of my favorites is Judson Brown’s (1948) finding in rats that the negativity of negative outcomes grows more rapidly with approach than does the positivity of positive outcomes.
6. “Everyone assumes Effect  $X$ , but is  $X$  robust and generalizable?” We are very attracted to clever experiments with interesting results, but we rarely question whether the results are robust and related to something in the real world. This type of contribution could be a failure to replicate some previously accepted finding, or an indication of the fragility of an effect, or a confirmation of the robustness and generality of an effect. Fragile effects can be indications of important processes, but at a minimum, we should know they are fragile because that makes research more difficult. We undervalue replication and extension of major findings to new organisms, domains, or sets of contexts and cultures. Replication and generalization is fundamental in natural science.
7. “This is a messy, criticizable experiment reporting something new and interesting.” This type of study usually involves an interesting idea, with some admittedly far from conclusive evidence for it. The famous Schachter and Singer (1962) attribution study is an example. Of course, some of these types of studies constitute false alarms, and difficult judgments have to be made about publication or grant awards.

As I have said before (Rozin, 2001, 2006) and as Ed Diener (2006) has suggested, the call is for more diversity in approaches, participants, and the questions we ask. There should be more discrimination between studies on new things that are not ready for sophisticated experiment, and studies on highly developed problems that are appropriately more refined and should meet higher methodological standards. And there should also be more interest in what people actually do (eat, have political views, watch television, choose and wear clothing; Rozin, 2006), more concern about whether the paradigmatic instances we choose for experimental analysis correspond to real-world events and are both robust and generalizable, and less concern with faults. Finally, we need more concern with net progress in the field and how we increase understanding. Negativity dominance (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Rozin & Royzman, 2001) is a part of animal and human nature, but it has to be controlled: A really interesting study with a flaw may be more valuable than a flawless but uninteresting study. In those areas of psychology dealing with whole human beings (e.g., social, developmental, clinical, and some aspects of other areas), we should realize that the more established sciences and the more established areas of psychology (e.g., sensation–perception) began with careful description and establishment of functional relationships (e.g., psychophysical functions).

There is no simple solution to reconciling the desire for rigor with the desire for relevance. In this author's opinion, psychology as an academic discipline has tipped the balance too much in favor of rigor, favoring experiment and hypothesis testing over examination and description of the basic phenomena in the field. It is probably more important to explore something real, important, and general across cultures than it is to do sophisticated experiments on something much less important.

We should not see contributions as flawless monuments that we can be proud of 20 years later. An experiment is just a sampling from an enormous set of possible parameters. In retrospect, the great experiments capture a truth about the world, but it is the problem selection, not the elegance, that primarily determines the greatness. We should just ask one simple question about any paper, a grant, or a psychologist: To what degree is our enterprise advanced by the work in question?

### REFERENCES

- Asch, S.E. (1987). *Social psychology*. New York: Oxford University Press. (Original work published 1952)
- Baumeister, R.F., Bratslavsky, E., Finkenauer, C., & Vohs, K.D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*, 323–370.
- Baumeister, R.F., Vohs, K.D., & Funder, D.C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, *2*, 396–403.
- Brown, J.S. (1948). Gradients of approach and avoidance responses and their relation to level of motivation. *Journal of Comparative and Physiological Psychology*, *41*, 450–465.
- Cherfas, L., Rozin, P., Cohen, A.B., Davidson, A., & McCauley, C.R. (2006). The framing of atrocities: Documenting the wide variation in aversion to Germans and German related activities among Holocaust survivors. *Peace and Conflict: Journal of Peace Psychology*, *12*, 65–80.
- Diener, E. (2006). Editorial. *Perspectives on Psychological Science*, *1*, 1–4.
- Haig, B.D. (2005). An abductive theory of scientific method. *Psychological Methods*, *10*, 371–388.
- Nisbett, R.E., & Cohen, D. (1996). *Culture of honor: The psychology of violence in the south*. Boulder, CO: Westview.
- Rozin, P. (2001). Social psychology and science: Some lessons from Solomon Asch. *Personality and Social Psychology Review*, *5*, 2–14.
- Rozin, P. (2006). Domain denigration and process preference in academic psychology. *Perspectives on Psychological Science*, *1*, 365–376.
- Rozin, P., & Kennel, K. (1983). Acquired preferences for piquant foods by chimpanzees. *Appetite*, *4*, 69–77.
- Rozin, P., & Royzman, E. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*, 296–320.
- Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, *69*, 379–399.
- Watson, J.D., & Crick, F. (1953). A structure for deoxyribose nucleic acid. *Nature*, *171*, 737–738.