# The Premature Demise of the Solo Experiment

**Daniel M. Wegner**
*University of Virginia*

*Papers reporting multiple studies have become common publication practice in our field. This practice has several serious costs that have not been foreseen, and it may be responsible for unfortunate developments such as a reduction in creativity, a paralysis of scientific interaction, a decline in the integration of the field, and an emphasis on microcertainty at the expense of examination of broad possibilities. The paper that reports a solo experiment is suggested as a potential solution to these problems.*

> Our errors are surely not such awfully solemn things. In a world where we are so certain to incur them in spite of all our caution, a certain lightness of heart seems healthier than this excessive nervousness on their behalf.
>
> William James
> (1897/1956, p. 19)

Scientific psychology is best served by the publication of papers that include multiple experiments. At least, this is what we all seem to have decided. Over the course of the past few years in personality and social psychology, we have gradually come to the point that a single experiment no longer contains enough truth of consequence to merit publication in the best journals in our field. Is the abandonment of single-experiment papers really a good idea?

This question has been on my mind recently. Even without the helpful statistics provided by Reis and Stiller (this issue), it has become evident that a tremendous bias against the "solo" experiment exists that guides both editors and reviewers. This alarms me for several reasons, and without knowledge of the present symposium I distributed a critique of the multiexperiment movement on electronic mail to a number of colleagues. The present article is a revised and, I hope, more informed treatment of the topic that results from the responses I received after the original message was loosed on the academic computer networks.

The point I wish to make here is that solo experiments are good science. To develop this point, I want to start by discussing the nature of scientific knowledge and then move to a consideration of the potential personal and scientific costs of the multiple-experiment paper.

BUMBLERS AND POINTERS

I developed my first psychological theory at age 11. As a result of my unintentional involvement in many youthful foul-ups and snafus, I decided that the world contained two types of people, the bumblers and the pointers. The bumblers, you see, go through life trying to do things. They simply bumble along, but they enjoy it and somehow get some things done along the way. The pointers, in contrast, do only one thing: They point out the bumblers' bumbles. Naturally, this leads to humiliation on the part of the bumblers and arrogance and condescension on the part of the pointers. The pointers never do anything themselves, but they certainly know when a bumbler has bumbled, and they announce it so widely and enthusiastically that the typical bumbler is paralyzed in shame for quite some time.

I classed myself among the bumblers, of course, and I took some solace in this theory during adolescence when I bumbled relentlessly and achieved massive levels of pointing by my peers and superiors. Over time, I realized that this was not just a matter of individual differences, but that people could move in and out of bumbling and pointing "modes" and thus assume either condition in proper circumstances. Once in a while, I

Author's Note: Correspondence regarding this article should be addressed to Daniel M. Wegner, Department of Psychology, Gilmer Hall, University of Virginia, Charlottesville, VA 22903. Electronic mail may be sent to dmw2m@virginia.edu.

*PSPB*, Vol. 18 No. 4, August 1992 504-508

even pointed. Fortunately, the inelegant nomenclature of this theory has kept me from ever writing it up for publication, and I apologize (as all real bumblers should) for even bringing it up here. Unfortunately, however, it is highly relevant to the issue of solo experiments.

It turns out William James had a similar theory, and as usual he said it better than I ever could and did so many years before I was born. In his essay "The Will to Believe," James makes this distinction: "*We must know the truth*; and *we must avoid error*,—these are our first and great commandments as would-be knowers; but they are not two ways of stating an identical commandment, they are two separable laws. . . . We may regard the chase for truth as paramount, and the avoidance of error as secondary; or we may, on the other hand, treat the avoidance of error as more imperative, and let truth take its chance" (1897/1956, pp. 17-18). James went on to argue that both activities are fundamental for the establishment of knowledge, both in science and in life. The statistically minded reader will recognize in these orientations the concerns with Type II and Type I errors in the hypotheticodeductive model of science.

The separability of these two enterprises is what interests me. It seems that the truth seeker is bent on the establishment of possibilities. Soft-headed and gullible, the person with this desire might discover many mutually exclusive "truths" that simply happen not to have been counterposed or otherwise tested. One may believe in modern medicine for the treatment of a back ailment, for example, and also believe in the effectiveness of a faith healer—as long as one is not concerned about error. Each "truth" may come forward and be useful in its own element, and the ultimate incompatibility of such possible beliefs may be of little concern. With an emphasis on the avoidance of error, in turn, comes a focus on impossibilities. The error avoider would note that two mutually exclusive possibilities cannot both be right and so discount one. But the hard-headed and suspicious individual motivated to reduce error might also attack the remaining possibility, perhaps to leave us with nothing at all to believe.

Science could not survive without the dialectic between these two characters, often embodied in the same person. We need both bumbling and pointing, grinning credulity and glowering skepticism, if we are ever to establish knowledge. If we go overboard in either direction, though, we risk a field that is not knowledgeable at all. With unbridled credulity, we risk multiple incompatible versions of the truth and an unwieldy overpopulation of ideas with little attachment to empirical referents. This direction leads toward the same problems that occur with multiple religions or multiple ideologies that share no basis for deciding on what is right: We get conflict and little else. With unbridled skepticism, however, the costs may be even more horrible. This direction leads to annihilation of our field.

Tipping the balance toward skepticism can eradicate ideas faster than we can generate them. Eventually, we arrive at a vacuous chasm, with no theory standing and no idea left without serious wounds. We have nothing left to think and nothing left to offer to others when they ask what social and personality psychologists know. This may seem like a caricature, an apocalyptic vision that has few implications for our field. But let us just think a moment about the demise of the solo experiment. Here we have a case in which skepticism has so overcome the love of ideas that we seem to have *squared* the probability of error we are willing to allow. Once, $p < .05$ was enough. Now, however, we must prove things twice. The multiple-experiment ethic has surreptitiously changed alpha to .0025 or below.

## THE SHAME OF THE SOLO

Most of us have probably complied with the multiexperiment ethic merely because, in our "bumbling mode," we are ashamed to do otherwise. We haven't really thought through the arguments, and we fall in line through the uninformed belief that bigger is somehow always better for the field. After all, the statistical reasons for multiple experiments are obvious—what better protection of the truth than that each article contain its own replication? Our editors point out from time to time that piecemeal publication of solo experiments can be confusing (Greenwald, 1976) and that multiple-experiment papers have greater impact than solos (Tesser, 1991). Most of us succumb completely. We chalk up our hesitance to conduct another study (after we've already done one perfectly good one) to personal sloth, and we jettison our appreciation of the solo in a flurry of embarrassment. Like Boxer, the workhorse in *Animal Farm*, we solve this problem by deciding to work harder.

Some of us, on occasion, instead succumb to the beauty of a single experiment and overcome our shame momentarily. It's so elegant, so simple, so compelling (we say to ourselves in a moment of creeping grandiosity) that we wish to give it a try. We write it up, submit it, and then hear back from our reviewers and editors the given wisdom we should have remembered: "This study is interesting and very well done but should not be published unless another experiment is added." Normally, then, several flaws of varying magnitude in the solo experiment are cited as reasons, and although none of these may really be "killers," we develop an unbearable malaise surrounding our laziness and finally go off to conduct Study 2. And sometimes it works wonderfully. We come up with a better package, and we thank the editorial team for acting as cheerleaders to urge us on to better work.

More often than not, however, we find the second experiment is harder to do than the first. Even if we do the exact same experiment again, the principles of statistical reliability indicate that we are less likely to find the same result; after all, the first experiment worked because of a combination of true and error variance that fell toward our hypothesis. Doing it again, we will be less likely to find the same thing even if it is true, because the error variance regresses our effects to the mean. So we must add more subjects right off the bat. The joy of discovery we felt on bumbling into the first study is soon replaced by the strain of collecting an all new and expanded set of data to fend off the pointers.

This is something of a nuisance in light of the reception that our second experiment will likely get. Readers who see us replicate our own findings roll their eyes and say "Sure," and we wonder why we've even gone to the trouble. If someone else had replicated our work, everyone would be far more inclined to weigh that person's effort quite heavily. A replication conducted as a class project by the third-graders at the local elementary school would be better than our own. So we must begin the second experiment with the sad realization that our redoubled efforts are buying us much less in the way of scientific credibility than would the efforts of anyone else. If we were planning science at some central control tower, we probably wouldn't want people to waste much time replicating their own findings in light of this observation —but multiple experiments waste scientists' time in just this sense.

Okay, so the second experiment must be bigger and yet will have less impact. Given that we are shamed into doing it anyway, we then discover that it is also harder to do than the first for key design reasons. It can't be just like the first, and it can't be entirely different. We must engage in a very delicate "tuning" process to dial in a second experiment that is both sufficiently distant from and sufficiently similar to the original. This tuning requires a whole set of considerations and skills that have nothing to do with conducting an experiment. We are not trained in multiexperiment design, only experimental design, and this enterprise is therefore largely one of imitation, inspiration, and luck. Like the bomber pilot who is rewarded not for hitting the target but for creating a nice, tight pattern of blasts, however, we may end up far from the mark when we must aim with multiple weapons.

Some of us use tricks to disguise our solos. We run "two experiments" in the same session with the same subjects and write them up separately. Or we run what should rightfully be one experiment as several parts, analyzing each separately and writing it up in bite-sized pieces as a multiexperiment. Many times, we even hobble the first experiment as a way of making sure there will be something useful to do when we run another.

All this concern with packaging is rapidly remaking our field. As Reis and Stiller document, papers are longer and have more methods, tables, references, experimental subjects, and authors, and they may take longer to publish as well. Is this really what we want? One way to consider this question is to take it to the extreme. Let's consider what might happen if we simply banished all solo experiments from the top journals and replaced them with multiple-study articles. Let's assume for a minute that the pointers win.

## A FIELD WITHOUT SOLOS

What would eliminating the solo experiment do to the field? Are there any hidden effects that we might worry about if we were to know about them? I'd like to suggest a connection between multistudy articles and several changes in scientific processes.

1. *Multistudy articles reduce creativity.* There is almost no space left in personality and social psychology, at least at the top, for the highly compelling, theoretically imaginative, wildly incomplete study. Festinger and Carlsmith (1959) would probably not see the light of day in contemporary journals of quality in social psychology. That article left a tremendous number of loose ends, enough for almost every social psychologist of the sixties and seventies to find one personal end to pick up and study. But such a solo can't pass the "do another study for no reason" germ that has so widely infected us. By requiring that authors invariably trudge through months if not years of labor to produce papers for the best journals, we effectively eliminate flashes of brilliance that authors are not willing to follow up with buckets of toil. This means that many of the most interesting ideas people have will be forsworn in favor of safe yet boring research. When solo studies were permitted, we rejoiced in an innovative result because we knew our colleagues' interest would be piqued. Now, such results are just a small start, because we know we must go and do it again (and maybe again) before we can even begin to tell our story.

2. *Multistudy articles undermine interaction among scientists.* Perhaps the most striking change in our field in the past few years is the relative lack of published cross-talk between researchers. People don't seem to cite each other very much, as they become involved instead in creating their own islands of self-citation. To some degree, this is just careerism; we all recommend that people avoid scattered efforts. But the demise of the solo exper-

iment takes us beyond this to rob us of the scientific process, the give-and-take of consensual validation that seems to happen on a regular basis for our colleagues in biology or physics. Because each paper we write must be a summation of several years' labor, we become reluctant to invest so much time in other people's ideas. So the light bulb that goes on when we read a journal article about someone else's work is not likely to lead us to research. Researchers each tell their own stories as they react to themselves and their own past works in new efforts. But the greater story of the field suffers as we all chatter to ourselves in empty rooms.

3. *Multistudy articles break the rhythm of scientific progress.* What ever happened to the "hot" finding? Is this field so muddled that we can well afford to wait years for everything? If we truly believe that progress in our field comes only through multistudy articles, that must mean we also believe that there is nothing new, nothing "late-breaking" to learn from our colleagues. It scares me to think how near we are to operating on an unspoken assumption that there is no underlying truth out there that we all are discovering together. We don't need to know what our colleagues are discovering now, and we are happy to wait 3 or 4 years until they have their act together. After all, each of us seeks his or her own truth, and we don't need to talk. In the end, this balky and syncopated process of scientific communication is something we try to overcome in other ways—by convention talks, preprints, e-mail interaction, and the like. But we don't solve the basic problem we seem to have accepted as a given: The best is invariably the slowest. One reason *JPSP* is no longer interesting is that we've already seen or heard its contents years before they are published.

4. *Multistudy articles exaggerate the value of publication.* It is important that we recognize what our journals really are. They are not collections of facts but, rather, collections of findings. A journal report is a possible truth—a hypothesis and some evidence—not a cold, hard, etched-in-granite, universal and incontrovertible law. Yes, even in *JPSP*. If we so much as suspect otherwise, we are taking scientific publication far too seriously and are missing the fluid, sometimes meandering, but always self-correcting nature of science. If an article is wrong, it will either (a) be exposed publicly by some number of researchers who can't get the effect or (b) be ignored because those researchers who tried to get it and failed just didn't publish. In either case, the research line will die. Science doesn't abide by wrong ideas for long. The solo experiment, tossed into the scientific grinders, may survive slightly less often than the multiple experiment. But we must learn again to trust those grinders, to have some respect for a system that has got us this far. In science, all papers are "working papers" and should not be mistaken for more.

5. *Multistudy articles promote "microcertainty."* If a multistudy paper does its job, it circumscribes the findings of the first study. Seldom do we find second or third studies in a series that really add something important and new to the original notion. (In fact, authors are unhappy about "spending" such new-idea studies on the second or third position in a paper.) What this means, then, is that the typical multistudy paper begins the work of the skeptic before the paper is even circulated. It slashes and burns any tall grass around the idea so we can see just what is left standing when the idea has been attacked from several vantages. This is to be admired in some sense, as it involves the development of knowledge through both of James's processes—the creation of the possibility and the elimination of error. However, the end result of this approach is that ideas emanating from the multistudy paper are small ones. They have been pruned of overextensions in advance and so allow us a prelimited view of what could be true. Microcertainty is science in baby steps. Rather than each study exciting people about what it could portend, it typically constrains the original advance and so cuts short our interest in what might have been.

All this is to say that I'm not nearly as optimistic as Reis and Stiller about the direction our field is taking. As it happens, most of my colleagues seem to agree. Of the 33 replies I've received on e-mail since I sent out a version of this article, all but 3 were resoundingly positive regarding a return to solo experiments. They ranged from "Amen" to "You've hit the nail on the head"—although one respondent did suggest that my essay would have been far more convincing if it were accompanied by another essay or two.

The three contrary responses reasserted the given wisdom and expressed a bit of righteous indignation at my suggestion that the publication process might have a problem. They all agreed on the central reason for multiple experiments: Unbelievable results. Each respondent noted that as a reviewer or editor, he or she was not willing to go out on a limb and accept results that looked to be unlikely or difficult to explain. Still, the relatively weak reply I received from the proponents of multiple experiments was in its own way astounding. I think many of us have assumed that *everyone* (other than us) must be in favor of increasing rigor, especially in the form of self-replication, and it is something of an eye opener to find that this is not true. The cause of scientific skepticism is difficult to argue against in public, and many of us have simply stood by to watch as the multiple-study ethic has crept into our lives.

What to do? Some colleagues suggested doing nothing special. They said authors should send solo-study papers to other journals and let *JPSP* do what it wants. Others suggested stopgap measures such as adding a "brief reports" section to *JPSP*. Another (obviously an editor) said that allowing the journal to add more pages would reduce the emphasis on eliminating papers from the publication queue. And one respondent recommended spanking and bed without supper for any reviewer or editor who ever said, "run another study," without explaining exactly what problem the new study would solve. I don't like any of these solutions (although the spanking makes an interesting image). I'd rather that, as a field, we work to achieve a more balanced view of the benefits *and* costs that multistudy packages can promote.

This essay has been decidedly imbalanced because I perceived that the weight of prevailing publication practice was clearly on the side of the multiple-study paper and that this needed redress. So I am personally not nearly as unreasonable about this topic as the essay would suggest. Indeed, I believe strongly that we must guard against error in our field, and I believe added studies are justified by significant lapses in an initial experiment. If there is one thing I would like this essay to do, it is to encourage people who take on the role of professional pointers—our editors and reviewers—to trust their own judgment a bit more. If we all had the nerve to judge papers on quality as we see it, rather than on quantity as measured by the mindless precept of self-replication, the scientific study of personality and social psychology could not help but surge ahead.

## REFERENCES

Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology*, *58*, 203-211.

Greenwald, A. G. (1976). An editorial. *Journal of Personality and social Psychology*, *33*, 1-7.

James, W. (1956). *The will to believe and other essays in popular philosophy*. New York: Dover. (Original work published 1897)

Tesser, A. (1991). Editorial. *Journal of Personality and Social Psychology*, *61*, 349.