2016-09

Charting The Future Of Social Psychology On Stormy Seas: Winners, Losers, And Recommendations

Roy F. Baumeister

The publisher's version of record is available at https://doi.org/10.1016/j.jesp.2016.02.003

Charting the Future of Social Psychology on Stormy Seas:

Winners, Losers, and Recommendations

Roy F. Baumeister

Abstract

Social psychology’s current crisis has prompted calls for larger samples and more replications. Building on Sakaluk’s (this issue) distinction between exploration and confirmation, I argue that this shift will increase correctness of findings, but at the expense of exploration and discovery. The likely effects on the field include aversion to risk, increased difficulty in building careers and hence more capricious hiring and promotion policies, loss of interdisciplinary influence, and rising interest in small, weak findings. Winners (who stand to gain from the mooted changes) include researchers with the patience and requisite resources to assemble large samples; incompetent experimenters; destructive iconoclasts; competing subfields of psychology; and lower-ranked journals, insofar as they publish creative work with small samples. The losers are young researchers; writers of literature reviews and textbooks; flamboyant, creative researchers with lesser levels of patience; and researchers at small colleges. My position is that the field has actually done quite well in recent decades, and improvement should be undertaken as further refinement of a successful approach, in contrast to the Cassandrian view that the field’s body of knowledge is hopelessly flawed and radical, revolutionary change is needed. I recommend we retain the exploratory research approach alongside the new, large-sample confirmatory work.

These are tough times for social psychology. The current crisis of confidence began with scattered revelations of scientific fraud, but it has built up through new data mining techniques into questioning the field's standard best practices for research — thereby even casting serious doubt on the entire body of knowledge that has been built up over many decades of work by thousands of researchers. Stroebe (this issue) finds it necessary to defend social psychology against the sweeping claim by Ioannidis (1997) that most published research findings are false. Ioannidis was not referring specifically to social psychology and thus was impugning the value of all research in all scientific fields. Still, social psychologists have felt themselves to be more vulnerable than most other scientists to such accusations of pervasive wrongness. It is a sad commentary on the state of the field that social psychologists need to assert that the field has in fact made progress and accumulated some valid knowledge.

The special issue has presented an impressive assortment of contributions addressing various aspects of the field's problems. Some offer creative, valuable, insightful suggestions for enabling us to do better science, such as by improving the quality of survey data (Berinsky, this issue), incorporating more field studies into our research programs (Maner, this issue), and improving our research designs so as to study mediation and mechanism more effectively (Pirlott, this issue). These should be circulated among laboratories, to help researchers get ideas for how to raise the quality of our research. Other papers, meanwhile, grapple with how the replication crisis will and should affect our discipline. My comments will be focused on the latter issue.

It is clear that the field is changing. I shall argue that the optimal approach for now is not a matter of replacing an obsolete, ineffective system with a shiny new one. Rather, we should continue business as usual while adding new approaches that can correct deficiencies and improve the aggregate quality of the published literature. In other words, keep the old model alongside the new model(s), rather than replacing it. I will particularly elaborate Sakaluk's (this issue) proposal that the optimal model is to explore small, confirm big.

EXPLORATION VERSUS CONFIRMATION

Many writers have called for sweeping changes in how research is conducted. To me, the more reasonable approach involves adding new approaches rather than deleting ones that have worked well. Social psychology has flourished by exploring all manner of phenomena. Now our field seems abashed that some ostensible discoveries might not be easily replicable, and so we want to shift the emphasis to confirmation.

My take is quite different. I think the amount of fraud is quite small. I think the field has done very well with the methods and standards it has developed over recent decades. It can improve further, and I support that. But to me the optimal policy prescription is for moderate change building on a successful general model, rather than radical restructuring. Social psychology has discovered many important phenomena and created a broad basic understanding of human nature and human social behavior that has been interesting and useful to other disciplines and to the general public. Have we made mistakes? Of course, but probably no more than is to be expected in a young science. We can tweak what we do to improve our output. What is needed in my view, however, is not revolutionary upheaval, but rather further refinement. Put another way, we need to find how to continue building on success, rather than starting over or facing up to a failed general paradigm.

Some writers have called for the field to abandon null hypothesis significance testing. Sakaluk (this issue) argues persuasively against doing that. He concedes the validity of some criticisms of that method, but he also recognizes problems with the alternatives. His solution is to reconceptualize the research process as having both an exploration (discovery) stage and a confirmation stage. Small samples with null hypothesis significance testing could serve the former. Sakaluk is quite right to suggest that respecting both steps would be a healthier approach to creating a strong discipline than a one-sided emphasis on either to the detriment of the other. From that perspective, the current crisis arose because psychologists have put too much emphasis on exploration and too little on confirmation — whereas the danger is now that the field will make the opposite mistake, emphasizing confirmation so thoroughly that it undercuts the exploration part.

Exploration is a crucial process that is often overlooked in the arguments about how research should be done and whether it is legitimate to conduct a series of small studies. Critics of the practice of running a series of small studies seem to think researchers are simply conducting multiple tests of the same hypothesis, and so they argue that it would be better to conduct one large test. Perhaps they have a point: One big study could be arguably better than a series of small ones. But they also miss the crucial point that the series of small studies is typically designed to elaborate the idea in different directions, such as by identifying boundary conditions, mediators, moderators, and extensions. The typical Study 4 is not simply another test of the same hypothesis as in Studies 1-3. Rather, each one is different. And yes, I suspect the published report may leave out a few other studies that failed. Again, though, those studies' purpose was not primarily to provide yet another test of the same hypothesis. Instead, they sought to test another variation, such as a different manipulation, or a different possible boundary condition, or a different mediator. Indeed, often the idea that motivated Study 1 has changed so much by the time Study 5 is run that it is scarcely recognizable.

A series of small studies can build and refine a hypothesis much more thoroughly than a single large study. That is Sakaluk's point, and I agree. The field will be better off to continue using that method to let ideas emerge and improve over the course of the data collection process.

In particular, the new fetish for large samples will constrain the discovery process. I find no way to argue against the value of replicating findings with large samples. Such studies will improve the field — but perhaps they will be better understood as *complementing* rather than *replacing* the way the field has operated for several decades. In my humble and biased view, social psychology has actually done quite well.

EFFECTS ON FIELD: MEET THE NEW SOCIAL PSYCHOLOGY

It is presumably uncontroversial to assert that the praxis of social psychology research is changing, and the changes are likely to last for the next decade or more. The question for the field is whether the new contingencies will be the best for making more scientific progress in our field. Here, I discuss some likely changes to how the field operates. Subsequent sections will comment on how these changes will benefit or hamper various groups.

Most of the changes will be driven by the replication crisis. Larger samples will be required. When I was in graduate school in the 1970s, n=10 was the norm, and people who went to n=20 were suspected of relying on flimsy effects and wasting precious research participants. Over the years the norm crept up to about n=20. Now it seems set to leap to n=50 or more. Replication and other forms of solid proof are to be demanded. What will this mean for practicing social psychology?

Enabling New Discoveries. Large samples increase statistical power. Therefore, if social psychology changes to insist on large samples, many weak effects will be significant that would have failed with the traditional and smaller samples. Some of these will be important effects that only became apparent with larger samples because of the constraints on experiments. Other findings will however make a host of weak effects significant, so more minor and trivial effects will enter into the body of knowledge. Researchers speak informally about finding “low-hanging fruit,” which means discoveries that are novel and easy to find. With suddenly much larger samples, there will be new low-hanging fruit.

Taken together, this means that plenty of new discoveries are now available. They are a mix of important findings that were previously hard to produce and now easier — and of trivial findings of real but objectively minor patterns that can be reliably produced in the laboratory but have hardly any impact on human social life in society and everyday life. Put another way, the new low hanging fruit for young scientists is full of weak effects that can reach significance in large samples.

Effect Sizes. To avoid the problem of filling our journals with weak effects, some recommend abandoning null hypothesis significance testing and focusing instead on effect sizes. Indeed, many argue that the somewhat arbitrary fetish of $p<.05$ is silly and obsolete and that it fails to do justice to the data. I respect these arguments. But the downside to that shift needs to be asserted. Ultimately, I think that would be a short-sighted and destructive change.

The laboratory experiment is not a good method for establishing effect sizes. The size of a laboratory effect is not a meaningful indicator of what happens outside the laboratory. *Laboratory effects are both artificially inflated and artificially deflated*, and so the true size of an effect in the so-called real world could be much larger or much smaller than what happens in the laboratory. It is important to recognize both possible distortions in laboratory effect size.

The inflation occurs because a carefully designed laboratory experiment sets up optimal conditions for producing the effect, such as by deliberately screening out (or holding constant) many factors that might operate outside the lab. The observed effect of X on Y will be larger in a perfectly designed, controlled, constrained laboratory experiment than it will be in everyday life. For example, the effect of a fleeting verbal cue is likely much greater in a laboratory context, in which the participant’s attention is focused on the stimuli being presented, no distracters are salient, the participant is not beset by emotion or other concerns, and so forth — as compared to that same verbal cue briefly passing through the environment while the person is engaged in some absorbing activity such as playing tennis, driving through a downpour, or arguing with a romantic partner.

Meanwhile, though, laboratory effect sizes are often much smaller than what might happen in the so-called real word. Practical and ethical constraints reduce what one can do in the laboratory. For example, my research program on interpersonal rejection and belongingness has operated by manipulating rejection experiences, such as by having strangers meet and get to know each other, and then randomly telling participants that the others have accepted or rejected them. Yet there is no way to make these rejections commensurate to what happens in real life. Being rejected by a couple of randomly chosen strangers in the laboratory, based on a contrived artificial interaction, is almost certainly less impactful than being divorced by a spouse or fired from a job, or even than being snubbed by an important acquaintance.

What the laboratory experiment does better than any other method is to establish causal relationships. It surpasses the alternatives at answering the question of whether X causes Y. This is a binary outcome: either it does or it does not. Null hypothesis significance testing is the best way to draw such a conclusion, again with the proviso that null findings are inherently ambiguous.

Some advocates of abandoning significance testing point out that the data exist on a continuum rather than a dichotomy, so to focus on effect sizes is more realistic than focusing on a yes-or-no conclusion based on $p<.05$. But as Swets, Dawes, and Monahan (2000) and others have argued, much of life is about drawing dichotomous conclusions based on continuous data. Either the person has cancer or he does not. The accused defendant committed the crime or did not. The suitcase contains a bomb or it does not. The woman is pregnant or she is not. The group will be awarded the contract or it will not. The student will be admitted to the program or not. The data on which such conclusions are drawn and decisions are made may be on a continuum, but the underlying reality is yes or no. The pure basic scientist's question is whether X causes Y or does not. Null hypothesis significance testing based on laboratory experiments is still the very best way of establishing whether X causes Y, under reasonably optimal conditions.

To be sure, a large sample furnishes a stronger basis than a small one, including for making dichotomous decisions. My argument here is not that we should be content with small samples. It is simply that the scientist's test of a hypothesis is inherently dichotomous, and if social psychology continues to emphasize laboratory experiments under artificial conditions as its primary method, then it must recognize that its goal is still to answer the yes-or-no questions about causality rather than to furnish a meaningful estimate of the size of such effects. The size of an effect in the laboratory is of no general or practical interest, insofar as the laboratory effect is independent of the power of that same causal effect in actual social life.

To illustrate this point with my research on rejection and belongingness, one question we addressed was whether being rejected increases aggression. The answer was yes (Twenge et al., 2001). We added further refinements, such as which potential targets of aggression are affected. I assume and hope that the causal relationship is true outside my laboratory, so that rejected people become more aggressive, at least under some circumstances and toward certain targets. But I see no way of estimating the size of the effect of rejection on aggression in the real world based on the size of the effects in the laboratory. Put another way, the lab study can say whether rejection increases aggression or does not — but it is hopelessly useless as to estimating by how much (also how often) rejection increases aggression.

Boring the Intellectual Community. Another danger is that if we get caught up into a highly rigorous but narrowly focused style of research, we will cease to produce work of interest to colleagues in other fields. Note what happened to personality psychology. Unquestionably the study of personality is far more rigorous today than it was half a century ago. Yet back then personality psychology captured the interest of people in many different fields, influencing anthropology, literary criticism, philosophy, plus lots of other subfields in psychology. The gain in rigor was accomplished by a loss in interest value. The Big Five, in particular, is far more solidly based in data than, say, Freudian psychoanalytic theory — but its influence on thinkers in other fields is far less, and indeed it has failed to capture the imagination of the intellectual community.

We have certainly taken steps in that direction already. The broader intellectual community is still impressed with some of social psychology's early classic studies, such as the Milgram obedience studies, the Darley and Latane bystander intervention studies, Schachter's creative work on emotion and on eating. Hardly any of those studies would be published today. But where would our field be without those? If one considers the amount of labor required to conduct such studies, it is impractical to demand that the researchers should have done them with n=50 per cell. Are we shutting the door on what our field's most influential work has been and could be?

Building Careers, Ignoring Behavior. For young researchers seeking to establish their career (e.g., earn tenure at a research-oriented university), the practical approach would be to eschew labor-intensive studies. Instead, the rational careerist approach will be to find convenient, efficient ways to get large amounts of data quickly.

One can envision social psychology in twenty years time in this way. The emphasis on large samples and rigorous replicating has produced a field in which the main practical way for research to get done is to have large samples of people sitting at computers making ratings. To caricature, social psychology will become the study of how people make ratings and judge hypothetical scenarios.

Social psychology in the 1960s and 1970s featured elaborate experimental setups and procedures, with plenty of actual behavior. This style has been in decline for decades, and some have worried that psychology will become the science of self-reports, finger movements, and hypothetical ratings rather than the science of behavior (Baumeister, Vohs, & Funder, 2007). Staging a 1970-style experiment was labor-intensive, often requiring an experimenter and a confederate to be both present for half or even a full hour for every research subject, one at a time. Doing this for 40 participants spread over four cells was a big investment in time. Doing the same thing for 200 participants for the same design may be prohibitively difficult. Hence whole categories of investigation will be abandoned.

A social psychology consisting mainly of people making ratings while seated at computers would not cover all the topics in the field. Much will get lost, including behavior. To be sure, uneven coverage is the rule, not the exception. Social psychologists do not study all aspects of social psychology, but rather they tend to cluster in “hot” areas with effective, easily replicated methods and exciting ideas (as well as social and political appeal; see Jussim, this issue). Thus, if the new rules change the focus of the field so that some topics become hot whereas others are ignored or downplayed, this would hardly be unusual.

Hence it is not necessarily a disaster if the field’s new methodological standards drive research away from some areas and toward others. But there still needs to be a place for those others. In the history of social psychology, a huge step occurred in the 1980s, with the recognition that relationships could not be studied as rigorously as first-time interactions among strangers. It is after all not possible to assign people randomly to be married vs. single. Yet a social psychology that studied only stranger interactions and ignored relationships would be deficient: it would miss out on explaining the majority of human social interactions, which occur among people who have an ongoing relationship extending into past and future. And so methodological standards were relaxed for relationships research, even while they became more rigorous for other topics (e.g., social cognition).

To me, this was optimal, and it made social psychology an intellectually rich, diverse, exciting area. We study many things, and not all lend themselves equally well to social psychology’s laboratories and methods. The optimal solution was to use different criteria for publication, so as to publish the best available work on each topic — rather than hold one uniform standard for all topics. Ideally our field would publish new and slowly improving work on many different topics. Some of that work will be more rigorous than others, because of the variation in inherent difficulty of studying the topic (like marriage vs first impressions). If we hold up the same high standard for all different topics, we will foreclose many important areas of study.

Risk Aversion. A last forecast is that the new rules will make researchers avoid risky work. Part of that comes from the numbers. If you have an allotment of 500 hours of research participants in a semester, you can gamble 50 on a risky, creative idea. But if you have to allocate 200 hours for the same study, that seriously hampers what else you can get done that semester, and so it is probably not worth taking the chance. A risk-averse field will be less likely to make exciting new discoveries. This will potentially curtail the field’s progress and influence.

The impetus to avoid risks extends beyond squandering research participants. Exploratory research is now disdained. Running a small experiment to test a boundary condition is risky because there is increasing pressure to report all failed experiments and all measures in all experiments. I have long advised my trainees to add an extra measure or two to a study, just to see what happens. Lately they push back, afraid to try something out because a failed attempt to explore might weaken the credibility for what they mainly seek to show.

THE WINNERS ….

Who stands to gain from these changes? All social and cultural changes create winners and losers. It is instructive to pause to consider who will flourish under the new rules and who will suffer. I begin with the likely winners.

Clearly, the much larger samples required by the new social psychology will require levels of diligence and patience far beyond what previous generations needed. When I ran my own experiments as a graduate student and young professor, I struggled to stay motivated to deliver the same instructions and manipulations through four cells of n=10 each. I do not know how I would have managed to reach n=50. Patient, diligent researchers will gain, relative to others.

Patience and diligence may be rewarded, but competence may matter less than in the past. Getting a significant result with n=10 often required having an intuitive flair for how to set up the most conducive situation and produce a highly impactful procedure. Flair, intuition, and related skills matter much less with n=50.

In fact, one effect of the replication crisis can even be seen as rewarding incompetence. These days, many journals make a point of publishing replication studies, especially failures to replicate. The intent is no doubt a valuable corrective, so as to expose conclusions that were published but have not held up.

But in that process, we have created a career niche for bad experimenters. This is an underappreciated fact about the current push for publishing failed replications. I submit that some experimenters are incompetent. In the past their careers would have stalled and failed. But today, a broadly incompetent experimenter can amass a series of impressive publications simply by failing to replicate other work and thereby publishing a series of papers that will achieve little beyond undermining our field's ability to claim that it has accomplished anything.

Having mentored several dozen budding researchers as graduate students and postdocs, I have seen ample evidence that people's ability to achieve success in social psychology varies. My laboratory has been working on self-regulation and ego depletion for a couple decades. Most of my advisees have been able to produce such effects, though not always on the first try. A few of them have not been able to replicate the basic effect after several tries. These failures are not evenly distributed across the group. Rather, some people simply seem to lack whatever skills and talents are needed. Their failures do not mean that the theory is wrong.

I admit there is an opposing ideal, articulated by the editors of this special issue, which is that in a serious science, causes may be precisely delineated and anyone should be able to replicate any effect. My reasons for thinking otherwise include the following. First, the view that anyone can replicate any finding in a serious science is patently wrong. In many sciences, years of specialized training and skill cultivation are necessary before competence is reached. I can read and possibly understand great experiments in nuclear physics, molecular biochemistry, astronomy, or neuroscience, but I do not begin to think I could replicate those findings. Second, because psychology deals with people amid complex social reality, replicating a finding may be more than setting up X and observing how Y inevitably ensues. It may also involve ensuring that C, D, and E do not interfere. An annoyed or disengaged participant may not act in ways not predicted by the hypothesis, not because the hypothesis is wrong, but because the angry or unmotivated state takes precedence and prevents the normal causal process from operating. The Open Science Collaboration (2015) "Many labs" replication project that brought such negative publicity to social psychology failed to make clear how many of the studies yielded nonsignificant results on the manipulation check. If the manipulation check indicates no difference, the study did not test the hypothesis, because the independent variable was not successfully manipulated.

More generally, it is naïve and perhaps absurd to assume that the same procedures will yield the same results regardless of context. Researchers need to calibrate their procedures to their participant population. Only if the independent variable is effectively manipulated and the dependent variable is properly, sensitively measured is the hypothesis actually being tested.

Social psychology seems to be forgetting the basic point, presumably taught in every methodology class, that null effects are inherently more ambiguous than significant findings. A failed study can mean the hypothesis is wrong, but it can mean many other things too.

Another important category of winner is the destructive iconoclast. Although my attitude toward the field is positive, I have some sympathy for these people. Many of these researchers are angry young men who want to make their mark. They see much published literature and fear that in order to persuade journals to publish their work, they must first discredit all that has gone before. My impression is that this is perennial, and I had some of the same feelings myself once. Colleagues persuaded me to present my work as building on prior work rather than destroying and replacing it, and this has served me well. But I do recall the feeling that I need to slash and burn in order to make a place for my own work.

Today, these angry young men (and perhaps women, though anecdotally this seems to me more a male pattern) have many more tools with which to slash and burn. Indeed, one could caricature the replication crisis as a way of making an excuse for young researchers not to bother reading the extant literature. The body of published work can be condemned almost wholesale, leaving room wide open for new work to be published on everything. If all the field's prior work is misleading, underpowered, or even fraudulent, there is no need to pay attention to it.

This is unfortunate for the long-term collegiality and progress in the field. I would prefer to assume that most social psychologists are competent, sincere, careful individuals who are doing their best to produce good research. If we could all treat each other with some degree of such respect, and as a result if we can resurrect the attitude that we are largely working together to build a common body of knowledge, the field is likely to do better than if we indulge the current preference for dismissing large areas of work and suspecting many of our colleagues of fraud.

Another set of potential winners are researchers in nearby fields who compete with social psychologists for grant funds, awards, faculty lines, graduate student support, and pages in broad-readership journals. As social psychologists persist in discrediting our field's work, we lend ammunition to other areas who argue that the precious and scarce research funds should be diverted away from social psychology and into their own areas.

There is one more potential group of winners, should they step up to capitalize on the opportunity. This would include some of the minor journals in the field. If they were willing to stay with the traditional approach of reporting series of smaller-sample studies as exploratory work, they might get some of the best, most creative new work, thereby raising their status and influence. As Schaller (this issue) noted, some findings will not be able to be done with large samples, but they may be among the most interesting and creative work in the field. Retaining this work in our published body of knowledge would benefit the field and could draw more attention and citation-impact benefit to the journals that publish it.

… AND THE LOSERS

What kinds of researchers are likely to be disfavored by the new rules? These people will have a harder time getting hired and earning tenure. Even beyond that, these people may be rejected for publication with work that previously would have been published, and so their contributions will be lost. Before we embrace rules that will turn these people into losers, we might consider their value to the field. Perhaps the system can be adjusted so as to continue to benefit from their contributions.

Young Researchers. The first group of losers consists of young researchers who are in training and trying to build their careers. This encompasses graduate students, postdoctoral fellows, and assistant professors. The future of the field depends on these people. Getting started on a research career is always difficult, but the new changes vastly increase their problems. They are seeking to master a craft while the criteria for success are being revised and made more difficult in ways that are unclear.

By analogy, imagine some young athletes seeking to succeed at a sport for which multiple sets of rules exist. The rules are not clearly spelled out. Each contest will have referees, but there is no way of knowing which set of rules these referees will use, nor exactly what the rules are. Compounding the difficulty, each contest will be adjudicated by different referees, who may differ as to which rules they use. It is hard enough to master a sport with clear, explicit, and consistent rules. Mastering one with unclear, unstated, and inconsistent rules is considerably more difficult.

Moreover — and this is crucial — the requirements of large-sample replications will certainly mean that it takes longer to prove oneself. These days it takes a lot more data to make the same point, as compared to fifteen years ago. New PhDs compete for jobs after five years in graduate school. In the past that was enough time to establish one's competence. Today that may not be enough. Many fine, talented researchers find that some of their best ideas are proven wrong, and there is no way to know in advance which ones those will be. (That is after all why science insists on testing them!) One has to be lucky in choosing which hypotheses to pursue. In the past, a promising but wrong idea could be tested and discarded in fairly short order, but now that the testing takes three or more times as long, the impact of bad luck is far more devastating. If a second-year graduate student has two novel hypotheses and happens to pick the wrong one to pursue, his or her career may be derailed.

Furthermore, the learning process may be compromised by the new standards. The recent model is that students run experiments and then reassess. If the study did not work as planned, they meet with their advisor to ponder the reasons for the failure and then try something different. By doing a series of medium-sized experiments, they develop skills and refine ideas. If we now start insisting that n=50 per cell, each experiment will take much longer than in the past. Feedback will be much rarer, and so the opportunities for learning will be greatly reduced. Skills cannot be cultivated as rapidly, and ideas cannot be refined as quickly.

If students cannot progress as rapidly and luck plays a bigger role in dictating the degree of early success than in the past, the hiring and promotion processes will be less reliable, so that in effect jobs will be allocated more randomly. Having sat through many tenure review meetings, I have long suspected that the five-year pre-tenure period is already too short to enable a university to make a reliable judgment about whether mid-range candidates will likely be productive over the long haul. If it takes twice as long as in the past to get a publication, then on average people coming up for tenure will have have only half as many publications, and tenure decisions will have to be made based on more limited, less reliable evidence.

Stroebe (this issue) notes a common problem with running a series of small, "underpowered" studies. Even if the hypothesis is correct, the underpowered one might fail to reach significance. "The temptation then is to disregard that study and to conduct another one, in the hope that it will work out." Instead of that, he advocates meta-analyzing the various studies together, so that the failed study is not swept under the rug. Stroebe's suggestion is sensible, but it may not be as straightforward to implement as one might think. When most of us conduct a series of small studies, these are not identical tests of the same hypothesis. Rather, each one seeks to try something new. Even with a fully correct and true theory, not all possible extensions, variations, and operationalizations will be successful. Counting the failures against the successes will bias the data against the theory, and so the discovery will not be made. To illustrate, suppose a theory enabled a researcher to make eight predictions, three of which are central but the other five indicate possible extensions. If the core three are falsified, the theory is impugned, but the core theory could be true even if all five of the extensions fail — especially if three of the failures reflect operational problems, such as failure to use sensitive measures or failure to put the participant into the appropriate mental state (e.g., anxious, conflicted, confident, motivated) needed to test the hypothesis.

As I write this, the Society for Personality and Social Psychology held its 2016 convention, and one issue up for debate is the replicability of ego depletion effects on self-regulation. Fishbach (2016) reported that her lab group failed to replicate ego depletion effects with online samples, despite finding them with other sorts of samples. Failures to replicate convey a sense that the phenomenon may not be real. However, Fishbach and her group explored reasons for the online failure other than wrongness of theory. They found that many online (Mechanical Turk) participants simply want a quick, easy, and somewhat pleasant experience. In the ego depletion condition, they must complete a somewhat arduous or unpleasant task at first — so large numbers drop out. MTurk does inform researchers of participant dropout rates until later in the process, and few studies report them. But she found that participants dropped out of the depletion condition much more than out of the control condition. As a result, the final sample of the depleted condition consisted mainly of highly motivated, self-disciplined participants, whereas the control condition contained the full range. The self-selection difference canceled out the effect of the manipulation (i.e., people with better self-regulation being overrepresented in the condition in which self-regulation was predicted by theory to be worse).

If we do not train the next generation of researchers as well, the quality of our field's research will suffer. And if we end up allocating jobs and tenure more randomly (i.e., on the basis of less information and less reliable information), again, the quality of our field will suffer.

Literature Reviewers and Textbook Writers. The second group of losers consists of people whose work relies heavily on trusting the field's body of knowledge. This includes people (like me) who write literature reviews. Being myself an avid thinker but only a middling experimentalist, I have built much of my career on writing reviews of empirical literatures to which I had contributed little or nothing. Literature reviews can add integrative theorizing beyond what single investigations can do, but they generally require faith that the work one is integrating contains truth.

This category of people who must trust the field's body of knowledge also includes people (also like me) who write textbooks. A textbook covers a large amount of published information in the field's literature. A textbook author must trust (mostly) the body of work and believe that it contains mostly truth.

To write a literature review or textbook is to integrate a large amount of information coming from the research findings of a large number of colleagues, including many whom one has never even met. In the current climate, one is supposed to be suspicious of all these other researchers, who may have p-hacked or even fabricated their data. This renders the literature reviewer's task almost impossible.

I genuinely believe that most social psychologists and most researchers in general are honest, diligent, competent practitioners of our craft, and that most of what they find is linked to genuine truth (albeit perhaps misinterpreted or especially overinterpreted, and subject to other biases).

Everything means something: That has been one guiding principle by which my students and I try to understand research findings. That arguably includes every observation in every study. It certainly includes every significant finding. Maybe it only means that demand characteristics or some confound produced the effects, so the data do not establish what the researchers want to claim it means. But it means something. And more often than not, it does indeed point toward some genuine fact about human cognition, motivation, emotion, and/or behavior.

In the current climate of suspicion and wide-ranging dismissal, how can one write a textbook or literature review? Would instructors adopt a social psychology textbook that omitted the Milgram studies, or the classic bystander intervention ones, simply because they would not be publishable by today's standards? And what remains of our field — would that even be an interesting and viable textbook?

Researchers who emphasize creative, exploratory work. A third group of losers is based on the "two kinds of scientists" taxonomy that has been reasserted in various forms. I do not know the original source and I suspect it has been reinvented many times. The two types are sometimes called the careful and the interesting, the sticklers and the creative ones, or those who strive to never be wrong and those who strive to never be boring. The distinction may also capture the differential emphasis on exploration/discovery versus confirmation (Sakaluk, this issue). To avoid being boring, one has to increase one's chances of being wrong, and vice versa. That is, the two scientific styles may have a tradeoff, such that greater emphasis on never being boring increases the risk of being wrong (especially since counterintuitive ideas are generally more interesting than merely confirming conventional wisdom), and vice versa. A thriving field benefits from having both types of researchers. As our field collectively raises the emphasis on never being wrong, it will disfavor the restless, creative types who emphasize being interesting even at the risk of occasionally being wrong.

I mentioned the rise in rigor corresponding to the decline in interest value and influence of personality psychology. Crudely put, shifting the dominant conceptual paradigm from Freudian psychoanalytic theory to Big Five research has reduced the chances of being wrong but palpably increased the fact of being boring. In making that transition, personality psychology became more accurate but less broadly interesting. I do not wish to offend my colleagues in personality, and certainly there is plenty of interesting work. Still, it seems undeniable that the interdisciplinary intellectual excitement that psychoanalysis and related theories once stimulated has not been sustained by the new, more rigorous programs of research. Social psychology might think carefully about how much to follow in personality psychology's footsteps. Our entire field might end up being one of the losers.

Researchers at Small Colleges. Another group of losers consists of researchers at small schools, where amassing giant samples is simply impractical. I myself moved in 2003 from a small private university to a large public one, partly for the sake of the big subject pool. Those of us at such institutions can adapt reasonably well and conduct large studies. At small schools, it will now take multiple semesters, perhaps multiple years, to do a proper, publishable study.

For them, the confirmation part of research has become impractical. If they are to participate in the research enterprise and contribute to advancing the field, the small-study plan is the only option. I submit that the field can ill afford to lose their contributions.

RECOMMENDATIONS AND SUGGESTIONS FOR MOVING FORWARD

My main recommendation is that social psychology as a field should embrace Sakaluk's (this issue) suggestion to explore small and confirm big — and should separate those into different parts of the grand research process, probably to be done by separate research groups and laboratories, and published separately. The long-term trend in almost all cultural evolution is toward greater specialization and division of labor. Social psychology should embrace this.

The field has long had a fairly clear pecking order of journals, and the difference among the journals has been mainly in terms of prestige and quality of research. Perhaps at this point we should institutionalize the distinction between exploration and confirmation, by encouraging different journals to focus on different steps. The most prestigious journals may now want to emphasize confirmation. But given the plethora of journals, perhaps some could focus on exploration. If our top journals will now refuse to publish articles that report series of small studies to develop hypotheses, perhaps some other journals might step into that gap and publish those papers. As noted above, that could elevate the prominence of those journals.

Uhlmann (this issue) proposed conducting non-adversarial large-scale replications. This seems a highly reasonable approach. But again, perhaps young researchers should not have to defer publication until those work out. The confirmation could come as a separate step and be published separately.

My recommendation, therefore, is to add rather than reduce ways of doing publishable social psychology. By all means, we should give special attention and rewards to people who do the best possible research, including with large samples, full mediation, and replications. But let us also continue to publish the sort of work that has brought so much success to our field. The traditional sort of paper, with a series of smallish studies that develop and refine a theory through multiple iterations, still has an important place in our field, alongside its new big-sample companion.

REFERENCES

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. doi: 10.1111/j.1745-6916.2007.00051.x

Berinsky, A. (this issue). Can we turn shirkers into workers? *Journal of Experimental Social Psychology*.

Fishbach, A. (2016: personal communication). Differential self-selection into conditions confounds and impairs online studies.

Ioannidis, J.P. (1997) Why most published research findings are false. *PLoS Med., 2*, e124.

Jussim, L., Crawford, J.T., Anglin, S.M., Stevens, S.T., & Duarte, J.L. (this issue). Interpretations and methods: Towards a more effectively self-correcting social psychology. *Journal of Experimental Social Psychology*.

Maner, J.K. (this issue). Into the wild: Field studies can increase both replicability and real-world impact. *Journal of Experimental Social Psychology*.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. Science, 349, #6251. DOI: 10.1126/science.aac4716

Pirlott, A., & MacKinnon, D.P. (this issue). Design approaches to experimental mediation. *Journal of Experimental Social Psychology*.

Sakaluk, J.K. (this issue). Exploring small, confirming big: An alternative system to the new statistics for advancing cumulative and replicable psychological research. *Journal of Experimental Social Psychology*.

Schaller, M. (this issue). The empirical benefits of conceptual rigor: Systematic articulation of conceptual hypotheses can reduce the risk of non-replicable results (and Facilitate Novel Discoveries Too). *Journal of Experimental Social Psychology*.

Stroebe, W. (this issue) Are most published social psychological findings false? *Journal of Experimental Social Psychology*.

Swets, J.A., Dawes, R.M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest, 1*, 1-26.

Twenge, J. M., Baumeister, R. F., Tice, D. M., & Stucke, T. S. (2001). If you can't join them, beat them: Effects of social exclusion on aggressive behavior. *Journal of Personality and Social Psychology, 81*, 1058-1069. doi: 10.1037/0022-3514.81.6.1058.

Uhlmann, E.L. (this issue) The pipeline project: Pre-publication independent replication of a single laboratory's research pipeline. *Journal of Experimental Social Psychology*.