

***Psychological Science* Pilot Project on Disclosure Statements**

Note: The gist of this letter was presented as part of the Cross-Cutting Theme Program on *Building a Better Psychological Science: Good Data Practices and Replicability*. Association for Psychological Science, 25th Annual Convention, Washington DC, May 2013.

Eric Eich
PSCI Editor-in-Chief ee@psych.ubc.ca
17 May 2013

Dear Fellow Editors: In the 2012 PSCI Annual Report, which you received last month, I alluded to a pilot project of mine that's aimed at improving the publication standards and practices of *Psychological Science* (PSCI). The project has concluded and I now write to share with you its methods, results, and potential implications for our journal's policies and procedures.

Background

The project has its origins in the oft-cited article on by Simmons, Nelson, and Simonsohn (PSCI, 2011, 22, 1359-1366) on the role of experimenter degrees of freedom in the emergence of false-positive effects. (Their article was the focus of a spirited discussion among several PSCI editors, and other interested parties, in October 2011.) The more proximal motivations for the project are a follow-up paper by Simmons et al. (2012) titled "A 21 word solution"—see http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2160588—and *PsychDisclosure*, a recent initiative by Etienne LeBel and his associates—see <http://psychdisclosure.org/>. The latter's platform "... provides authors of recently published articles in psychology an opportunity to publicly disclose four categories of important methodological details [Exclusions, Conditions, Measures, and Sample Size] that are not required to be disclosed under current reporting standards, but which are essential for interpreting research findings."¹

In my project, authors of recent submissions to PSCI were asked whether they had reported (1) the total number of observations that were excluded (if any) and the reasons for doing so, (2) all independent variables or manipulations, whether successful or failed, (3) all dependent variables or measures, and (4) how they determined their sample size. Authors were told that these four items constituted the **Disclosure Statement**, and that their replies should cover "... all studies in your recently submitted manuscript to *Psychological Science*".

¹ I am indebted to Etienne LeBel for letting me borrow the questions and examples he developed for *PsychDisclosure*, with some minor modifications to his wording.

Authors were also informed that (a) their participation in the survey is both voluntary and independent of the ongoing PsychDisclosure project, (b) whether or not they complete and return the survey would have no bearing whatsoever on the fate of their submission in initial or extended review at PSCI, (c) I would play no role whatsoever in any editorial decisions regarding their submission, and (d) their responses would be kept confidential.

In addition to the four core items, authors were asked three ancillary questions: (5) how long it took to complete the Disclosure Statement, (6) how informative their responses to the Disclosure Statement would have been to readers, reviewers, or editors, had they been able to see them, and (7) whether they have any suggestions or feedback—positive or negative—on any aspect of the Disclosure Statement.

Over a six-week period (mid February through March, 2013), I emailed the attached invitation and questionnaire to the corresponding authors of an opportunity sample of 243 PSCI manuscripts (Research Articles, Research Reports, or Short Reports), typically within two days of manuscript submission. To date I've received 145 completed surveys, representing a 60% response rate.

Disclosure Statement Items

Each of the graphs shown below corresponds to one of the four items in the Disclosure Statement; specifically:

(1) We reported the total number of observations that were excluded (if any) and the reasons for doing so. (If no observations excluded, mark Yes.)

Yes: ___ No: ___

If No, please report this information:

(2) We reported all independent variables or manipulations, whether successful or failed.

Yes: ___ No: ___

If No, please provide brief explanation for not reporting this information:

(3) We reported all dependent variables or measures.

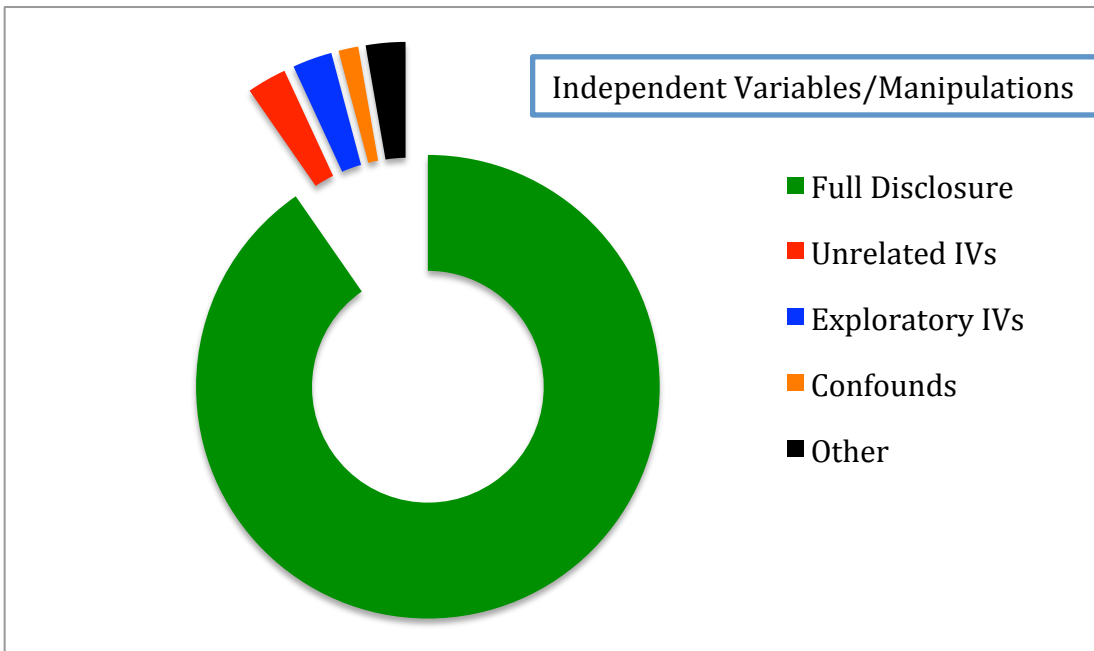
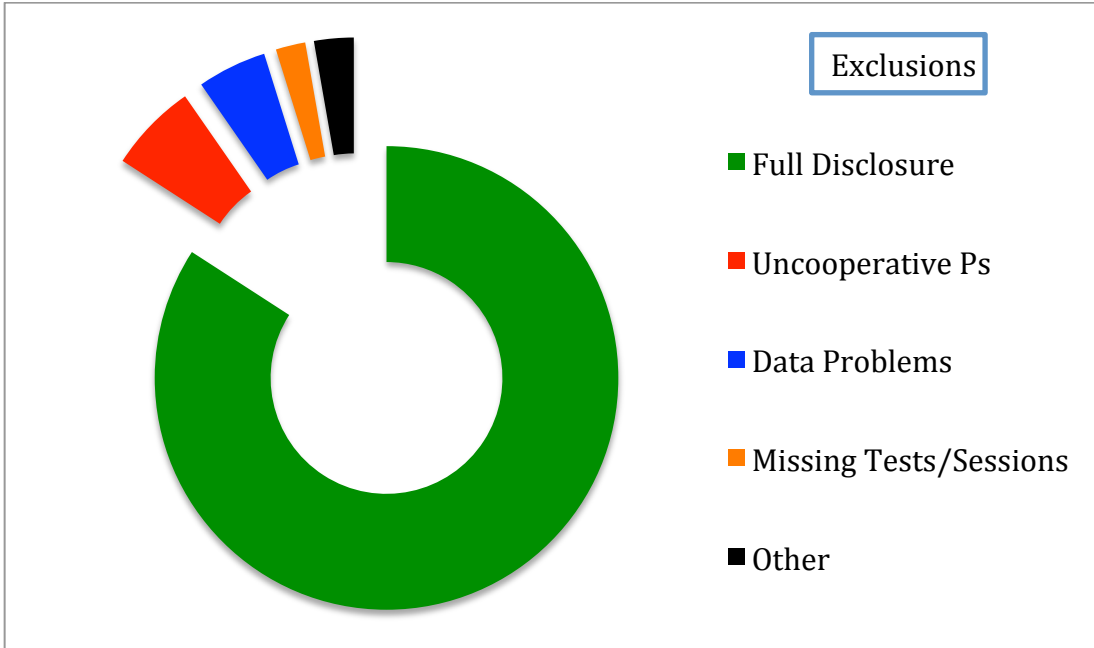
Yes: ___ No: ___

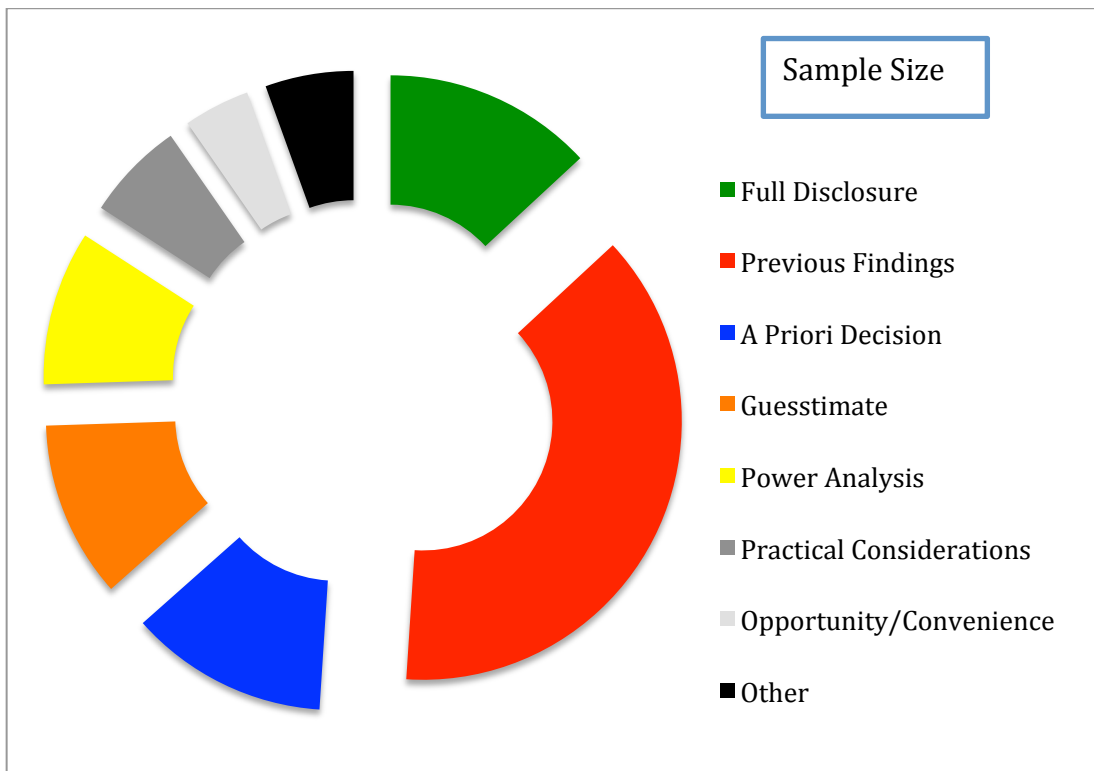
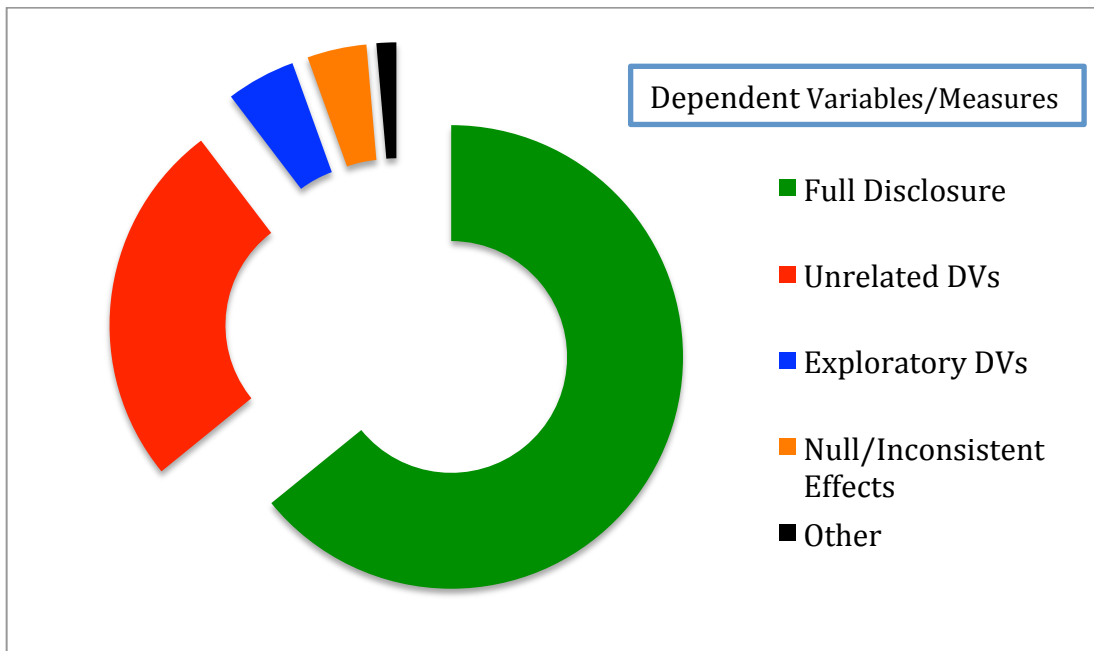
If No, please provide brief explanation for not reporting this information:

(4) We reported how we determined our sample size.

Yes: ___ No: ___

If No, please describe (a) the rationale for the sample sizes used and (b) how you decided to stop collecting data:





The green area in each graph represents “yes” or full disclosure responses (maximum = 145). The other areas reflect “no” responses, color-coded to indicate the primary reason given by the corresponding author. Not infrequently, authors provided one or more secondary reasons as well, which we’ll discuss later.

The full-disclosure rates for Exclusions and Independent Variables/Manipulations (henceforth IVs) were both robust (84% and 90%, respectively). Exclusions not reported in the main text, or in Reviewed Supplemental Online Material (SOM-R), were chiefly due to uncooperative (inattentive, unfocused) participants or to various data problems (computer malfunctions; calibration errors) or to incomplete protocols (e.g., the research assistant administered the wrong task or a participant missed the final test session). For the most part, unreported IVs were manipulations that authors deemed exploratory in nature or judged to be unrelated to the key issues at stake in their paper. Occasionally a condition was dropped when a confound was discovered while the experiment was underway. Note that a sizeable proportion of papers were non-experimental (correlational; observational) in nature; by default, such papers were coded as “full disclosure” with respect to IVs.

Only 64% of the sample submissions reported all Dependent Variables/Measures in the text (or SOM-R). As was the case for unreported IVs, the majority of unreported DVs were described as being “unrelated” or “exploratory”. The former type was very common in longitudinal studies of cognitive development, healthy aging, personality processes, and other topics that invite a multimethod research approach entailing, for example, behavioral, physiological, neuroimaging, and genetic measures. In these cases, authors reported only the DVs that related directly to the issues at hand—a perfectly reasonable strategy. (Another commendable practice: Many of these same authors expressed their willingness to provide a complete, or at least more fulsome, description of omitted measures to other researchers, upon request.) Exploratory measures (demographic surveys, personality questionnaires, mood ratings, etc.) were common in many areas of research. In most instances, authors pointed out that such measures were administered at the end of the study rather than between the principal IVs and DVs (another good practice).

In a small fraction (5%) of submissions, measures went unreported because they yielded null or inconsistent results as main or interactive effects, or as potential moderator or process variables. Such omissions may or may not reflect the researchers’ questionable—but not malicious—decision to take advantage of a situation that makes their results look better than they are. As noted by Simmons et al. (2012, p. 1363), “one reason researchers exploit researcher degrees of freedom is the unreasonable expectation we often impose as reviewers [and editors] for every data pattern to be (significantly) as predicted.” To counter this expectation, reviewers and editors “should be more tolerant of imperfections in results” (Ibid.) and prioritize “transparency over

tidiness; if a wonderful study is partially marred by a peculiar exclusion or an inconsistent condition [or, I would add, a misfiring measure], those imperfections should be retained. If reviewers [and editors] require authors to follow these requirements, they will” (Ibid.).

Regarding sample size, all submissions reported how many participants were tested (no surprise there), but only a small fraction (13%) revealed why that number was chosen or what decision rule governed termination of data collection. The Disclosure Statements offered up many different answers, including: reliance on previous findings reported in the literature or obtained in one’s own lab (sometimes expressed as “rules of thumb”), a priori decisions to achieve a certain, minimum sample size, power analysis (usually without details on target alpha levels or effect sizes), guesstimates (common in studies with novel, hybrid methods), opportunity or convenience samples, and practical considerations (e.g., test as many Ps as possible with the time and funds available).

The Other category under Sample Size included a few cases in which more data were collected to determine whether results that looked promising might indeed be significant. As Simmons et al. (2012, p.2) remark, “Some think dropping conditions is fine, others do not. Some think collecting 10 subjects at a time is fine, others do not. Some think dropping measures is fine, others do not. And so on.”

The deeper issue at stake here is trust. Continuing their argument, Simmons et al. (2012) claim that:

For trust to exist, people must agree on what it is they are trusting. We cannot “trust” to run and report their studies “properly” if there is no shared understanding of what “properly” is. There is no shared understanding of what “properly” is. ... Asking authors to disclose does not take trust out of our scholarly exchanges, it creates a framework for trust to meaningfully exist (p.2).

Ancillary Questions

On average, Disclosure Statements were completed in 7.9 minutes (range 1 to 60 min; n=142 owing to 3 missing or uncodable responses). Give a pronounced positive skew to the data, the median completion time—5 minutes—is a better measure. 83% (118/142) of the Disclosure Statements took 10 minutes or less to complete, suggesting that the task is reasonably straightforward and practicable.

The second ancillary question was: “How informative do you think your responses to the Disclosure Statement would have been to readers, reviewers, or editors, had they been able to see them?” Two authors didn’t respond and six said they were unsure, leaving 137 codable answers. Of these, 17 (12%) were

negative, one passionately so. Most respondents didn't provide specific reasons, but among those who did, the most common concern was that readers can't tell if Disclosure Statements are true--another side of the "trust" issue raised earlier.

The remaining 120 codable responses (83%) were positive to varying degrees (ranging from "slightly useful" to "immensely informative") and for varying reasons (e.g., potential to improve publication practices or to enhance replication efforts). In roughly half the cases, respondents qualified their positive opinions by, for instance:

- identifying some Disclosure Statement questions as being more important than others (though, in the end, every question had its fair share of champions),
- viewing the Disclosure Statement as valuable, provided it wouldn't count toward word limits (we'll return to this point later on), or
- claiming that Disclosure Statements are beneficial in general, but not in their specific case. Twenty-two respondents made this proviso, many who had answered "yes" to most or all of the Disclosure Statement questions. As a consequence, hypothetical readers wouldn't need to rely on the Disclosure Statement for answers: they could find them in the manuscript itself!

The final ancillary question sought feedback on any aspect of the Disclosure Statement. Three recommendations predominated: (1) expand the Disclosure Statement to cover analyses performed but not presented in the main text, (2) ask about additional studies, including pilot studies, that explored the same research question but that were excluded from the main text, and (3) while the Disclosure Statement is a step in the right direction, PSCI should take the lead in publishing all data, codes, and materials. Let's examine the first two recommendations now, and save the third for later.

Some of you may recall that, back on 8 October 2012, I sent a letter to PSCI's Senior and Associate Editors outlining three initiatives for the journal. One was to require authors to complete a **Research Process Statement (RPS)** at the time of manuscript submission. The goal of the RPS—just like the streamlined Disclosure Statement that replaced it—was to increase transparency and completeness in reporting the work. Also like the Disclosure Statement, the RPS was inspired in part by Simmons et al. (2011) experimenter-degrees-of-freedom paper, cited at the outset of this letter. The five open-ended items in the RPS were:

(1) Describe the rationale for the sample size in each study including whether preliminary analysis of data influenced whether to stop or continue data collection.

- (2) Describe any experimental conditions or other independent variables collected in each study but not mentioned in the main text.
- (3) Describe other measures or data collected for each study but not mentioned or included in the main text. If description is not feasible, explain why.
- (4) Describe analyses performed other than those reported in the main text.
- (5) Describe additional studies, including pilot studies, not mentioned in the main text but that test the same research question.

In a subsequent letter, dated 13 January 2013, I mentioned that the feedback received from PSCI editors and other folks on the RPS was mostly positive, except for the last two items.

Regarding #4, several people, especially those working in large research teams, remarked that they don't keep track of every analysis they run and that doing so would disrupt the team's workflow. Others observed that #4 is presumably intended to get at tested but ignored alternative specifications (e.g., exclude outliers vs. log transform vs. square-root transform, etc.). They weren't convinced that #4 would help matters, and I came to agree: instead of reporting all of their analyses, authors should provide their data. I'll return to this point momentarily.

Regarding #5, many felt it would open a large can of worms. The major concern, mentioned by Leif Nelson (among others), was that:

... it is all too easy for a researcher to think that an excluded study doesn't count. Furthermore, this actually puts a meaningful burden on the "full disclosure" researcher. #1 - #3 are equally easy for everyone to answer; either that information is in text or they write it down right now. But #5 is different. The researcher who convinces herself that the excluded study doesn't count has now saved herself the hours it might take to write it up for this query. File-drawering studies is damaging, but I am not convinced that this will solve that problem.

To be sure, not everyone within or outside our group would agree with this view. But I do think it's a valid point and that other, more effective remedies are available—a point we'll return to shortly. My immediate goal is to explain why the Disclosure Statement retains the spirit, if not the exact letter, of the first three items in the RPS (concerning sample size, manipulations, and measures), but leaves out the last two items (concerning unreported analyses and related studies).

Moving Ahead

Inspired by Simmons et al.'s (2011, 2012) approach to the problem of experimenter-degrees-of-freedom, and impressed by Etienne LeBel's *PsychDisclosure* initiative, the pilot project reported here sought to (1) assess authors' willingness to disclose methodological information that is not normally reported, under current publication guidelines, and (2) develop a clear picture of what "Disclosure Statements" would look like, in terms of both their composition and length, should my fellow editors and I decide to make them a fixture of future PSCI articles.

On balance, I view the results of the pilot project as promising: most respondents completed the Disclosure Statement in a timely manner (10 minutes or less) and with apparently little trouble (there were few, if any, complaints about the questions being unclear or unfair) and positive evaluations of the Statement's utility outnumbered negative evaluations by a large margin. That said, the limitations of the project must be acknowledged: both the sample size (145 respondents) and the response rate (60%) are modest, and most obviously, the sample is self-selected: anyone who objected to the Disclosure Statement, for principled or practical, reasons, could have (and doubtless did) hit "delete" when my email invitation arrived.

What I find most striking about the results is the combination of two things. One is that the questions included in the Disclosure Statement speak to four very basic elements of scientific method: were there any data exclusions, dropped manipulations, or dropped measures, and how was sample size determined? None of these are abstruse bits of methodological arcana. And yet, for the second observation, in less than half of the sample (42% or 61/145) were any three of these elements already covered in the main text. Further, the percentage of manuscripts containing all four elements (10%) was comparable to the percentage that contained none of them (4%).

The key point is this: Had authors included answers to the four Disclosure Statement questions, either in the main text or in supplemental online material, they could have added the following, 15-word statement (used by the Open Science Framework (<http://openscienceframework.org/>) to their manuscripts: "We report all data exclusions, manipulations, and measures, and how we determined our sample sizes."

The brevity of this statement belies its importance: this is the sort of information authors should share with readers, reviewers, editors, and potential replicators. But, they seldom do, under current reporting standards.

Moreover, even if authors cannot make the 15-word statement (e.g., they are not reporting all of their measures), then it is in everyone's interests for that explanation to appear right in the text, ideally right next to the modified statement

(e.g., "Although we report how we determined our sample size and all of our experimental conditions, we do not report all measures for the following reasons..."). The PsychDisclosure website contains many examples where respondents answered "No" to one or more questions and explained why in a clear and concise manner. Most of the answers provided in my pilot study of the Disclosure Statement were similarly straightforward.

I think we all agree that it would be a good thing to create a simple public norm for reporting what should be requisite information. My sense is that Simmons, LeBel, and their associates are on the right track and that PSCI is well positioned to promote the cause. So, I suggest that, beginning in January 2014, we require authors to complete a Disclosure Statement, similar if not identical to the one used here, when they submit their manuscripts. (The scope and substance of the initiative should be widely advertised in the meantime, so that authors are well prepared in advance.) If the paper is published, the completed Disclosure Statement could be posted online as reviewed supplemental material (SOM-R). Alternatively, during revision (which almost all PSCI articles undergo), the key contents of the Disclosure Statement could be moved into the main text, but not counted toward word limits. Or even a hybrid model might work in some cases, with a short synopsis in the text and (if needed) a longer, detailed discussion of specific issues (e.g., unreported measures) posted online. Whatever floats your boat.

While I'm cautiously optimistic that the benefits of Disclosure Statements will outweigh their costs (to authors, reviewers, and editors alike), there's more that PSCI can and should do to raise our game.

In particular, I would like us to include a statement, in the *Submission Guidelines*, that PSCI strongly encourages authors to (a) register their research projects (using the Open Science Framework, for example) and (b) make their data publicly available (via OSF, Dataverse, Figshare, or equivalent data-hosting provider). Of course, I would need to give good reasons for these recommendations, but that's easy to do. For instance, study registration is a potent remedy for the pernicious file-drawer problem, whereas public data posting enables more researchers to examine, and thus learn from, a given body of evidence (see Wicherts & Bakker, *Intelligence*, 2012, 40, 73-76).

PSCI articles containing links to registered studies and/or posted data could carry special symbols or icons—"gold stars", in a manner of speaking—as tokens of acknowledgement. But I think the most we can do, at least for now, is nudge PSCI authors toward study registration and data posting. Prior attempts by psychology journals to require data posting proved unsuccessful, and while the practice is now commonplace in economics, political science, and several other disciplines, the impetus mainly came not from the top down (e.g., by journal decree), but from the bottom up (i.e., by leading researchers taking the initiative). It's too soon to say whether the same will hold for study registration, but both the

ongoing Reproducibility Project and the forthcoming Replication Reports in PoPS seem certain to raise awareness of the advantages of registration and, in turn, promote its practice.

Thanks for your attention and I welcome hearing from you at your earliest convenience.

Best,
Eric