## REPLY TO FRANCIS (2014)

Francis's previous critiques have been expertly criticized on both methodological and conceptual grounds (e.g., Galak & Mervis, 2012 Simonsohn, 2012; 2013). As Simonsohn (2013) noted:

> "In numerous one-off critique-articles, Francis presents evidence that individual psychology papers suffer from publication bias, and concludes that the results from these papers ought to be fully ignored. I recently argued the critiques themselves suffer from publication bias and, more importantly, that the recommendation to throw out all data does not follow from the presence of publication bias."

First, it's important to be clear about what "publication bias" means. It doesn't mean that anyone did anything wrong, improper, misleading, unethical, inappropriate, or illegal. Rather it refers to the well-known fact that scientists in every field publish studies whose results tell them something interesting about the world, and don't publish studies whose results tell them nothing. Francis uses sophisticated statistical tools to discover what everyone already knew—and what he could easily have discovered simply by asking us. Yes, *of course* we ran some studies on "consuming experience" that failed to show interesting effects and are not reported in our JESP paper.

Let us be clear: We did *not* run the same study over and over again until it yielded significant results and then report only the study that "worked." Doing so would be clearly unethical. Instead, like most researchers who are developing new methods, we did some preliminary studies that used different stimuli and different procedures and that showed no interesting effects. Why didn't these studies show interesting effects? We'll never know. Failed studies are often (though not always) inconclusive, which is why they are often (but not always) unpublishable. So yes, we had to mess around for a while to establish a paradigm that was sensitive and powerful enough to observe the effects that we had hypothesized. In one study we might have used foods that didn't differ sufficiently in quality, in another we might have made the metronome tick too fast for people to chew along. Exactly how good a potato chip should be and exactly how fast a person can chew it are the kinds of mundane things that scientists have to figure out in preliminary testing, and they are the kinds of mundane things that scientists do not normally report in journals (but that they informally share with other scientists who work on similar phenomenon). Looking back at our old data files, it appears that in some cases we went hunting for potentially interesting mediators of our effect (i.e., variables that might make it larger or smaller) and although we replicated the effect, we didn't succeed in making it larger or smaller. We don't know why, which is why we don't describe these blind alleys in our paper.

All of this is the hum-drum ordinary stuff of day-to-day science—and all of this is very different than running exactly the same experiment over and over again and then publishing the 5% that worked by chance. That's clearly bad practice, we don't know anyone who does this, we have never done this, and we are glad that Francis doesn't suggest we did. But Francis does suggest that our results should be ignored, and in so doing, he conveniently ignores the fact that our effects have now been conceptually replicated in at least three published papers: Buechel, Zhang, Morewedge, & Vosgerau (2014); Ebert & Meyvis (in press); and Novemsky & Ratner (2003). Francis's methods have a host of other problems that have been described so well and so thoroughly by others that we won't reiterate them here.

So yes, Francis demonstrates what everyone already knew. Simonsohn (2013) said this in his most recent critique of Francis, but he also said something else of equal importance:

> "What are we learning from the (8 so far) Francis critiques? We are not learning that publication bias happens, we already knew that. We are not learning that the critiqued studies ought to be ignored, because that just does not logically follow from them

> containing publication bias. We are not learning that the critiqued studies have more severe publication bias than others, because Francis' selective reporting of results, and non-representative selection of studies to analyze in the first place, prevents us from making such inference. What do we learn then? We learn that new statistical tools, perhaps especially those that provide potential critics with access to easy publications, can be misused. We learn, then, that developers of new tools ought to include in their papers safeguards to prevent their misuse… For instance, one could require critics to disclose how they selected the target of their critique..."

The fact that Francis's critiques are the result of "cherry picking" not only invalidates their conclusions, but also represents a potential misuse of statistical tools. Is that what happened here?

We asked Francis why he picked our paper to criticize, and after just a little squirming he admitted that he had picked our paper because two of its authors—Gilbert and Wilson—had spoken out publically about problems in the so-called "replication revolution." We initially found this so difficult to believe that we wrote to Francis to make sure we hadn't misunderstood. Gilbert wrote:

> "Okay, I think I understand now… In the paper you sent us as an attachment, you specifically targeted our published paper in JESP because Wilson and I are its authors and we have spoken out on our blogs about problems with the so-called 'replication revolution.'"

Francis confirmed this statement with a one-sentence reply: "Yes, that's the situation."

We have included our email correspondence with Francis below the references. Readers can see for themselves that Francis's critiques are not just trivial, but are the result of targeting particular scientists because of their public speech. We believe that such targeting should be unacceptable to the scientific community.

*Carey Morewedge, Daniel Gilbert, & Timothy Wilson, 7-22-14*

**References**


Buechel, E. C., Zhang, J., Morewedge, C. K., & Vosgerau, J. (2014). More intense experiences, less intense forecasts: Why affective forecasters overweight probability specifications. *Journal of Personality and Social Psychology, 106,* 20-36.

Ebert, J.E.J. & Meyvis, T. (in press). Reading fictional stories and winning delayed prizes: The surprising emotional impact of distant events. *Journal of Consumer Research*.

Galak, J., & Meyvis, T. (2012). You could have just asked: Reply to Francis (2012). *Perspectives on Psychological Science*, *7*, 595–596.

Novemsky, N., & Ratner, R. (2003). The time course and impact of consumers' erroneous beliefs about hedonic contrast effects. *Journal of Consumer Research, 29,* 507–516.

Simonsohn, U. (2012). It does not follow: Evaluating the one-off publication bias critiques by Francis (2012a, 2012b, 2012c, 2012d, 2012e, in press). *Perspectives on Psychological Science, 7*, 597–599.

Simonsohn, U. (2013). It really just does not follow: Comments on Francis (2013). *Journal of Mathematical Psychology, 57*, 174–176.

From: Gregory Francis <gfrancis@purdue.edu> Sent: Fri 06-27-14 9:13 AM
To: morewedge@gmail.com; Gilbert, Daniel; tdw@virginia.edu
Cc:
Subject: Morewedge et al. (2010)

Message | Paper.docx (479 KB) | Exp1.R (4 KB) | Exp2.R (4 KB) | Exp3.R (4 KB) | Exp4.R (8 KB) | Exp5.R (7 KB)

Dear Drs. Morewedge, Gilbert, and Wilson,

I am writing about an article you published in 2010 in the Journal of Experimental Social Psychology titled "Consuming experience: Why affective forecasters overestimate comparative value." I have read with interest the blog postings by Gilbert and Wilson about replications and negative psychology, so I thought it would be interesting to share with you some observations about this paper.

As you can see in the attached Word document, the statistical findings in the paper fail the Test for Excess Success that has been used elsewhere to identify publication bias or questionable research practices. My intention is to submit a short note to JESP describing this analysis, but given your interests on these topics I thought we might collaborate to write a broader paper.

I share many of the concerns expressed by Gilbert and Wilson about how a failure to replicate a finding is sometimes improperly interpreted as an ethical shortcoming of the original authors. I support replication efforts, but my view is that the field does not have too few successful replications but rather too many. As the analysis in the enclosed document demonstrates, if the effects reported in Morewedge et al. (2010) are valid, then a successful outcome across every experiment should only occur with an estimated probability of 0.003. That is, across the five experiments there should have been several unsuccessful hypothesis tests. The absence of these failures indicates something went wrong in data collection, data analysis, or reporting.

I would like to collaborate with you in a discussion about what went wrong in Morewedge et al. (2010). Rather than simply speculate about publication bias or questionable research practices, I think it would be valuable to consider what exactly happened in this case. In particular, my impression is that what are often called questionable research practices are only "questionable" with regard to standard techniques in hypothesis testing. Their presence does not reflect ethical shortcomings by researchers but rather proper scientific methods that undermine hypothesis testing techniques. For example, from a scientific perspective it makes sense to add more data in order to get a clear result, but this approach is incompatible with standard hypothesis tests. This aspect of the so-called crisis in psychology has been under appreciated and I think it deserves a proper treatment with a particular example.

I imagine it can be uncomfortable to receive a critique of your published work, but I hope you agree that it would be interesting and valuable to collaborate in order to fully discuss these issues in a manuscript. It would simultaneously avoid a continuation of misunderstandings about your work in particular and would push the field toward a more open and reasoned discussion about research practices.

Best wishes,

Greg Francis

Professor of Psychological Sciences
Purdue University

On 27 Jun 2014, at 04:33 pm, Gilbert, Daniel <gilbert@wjh.harvard.edu> wrote:

Hi Greg,

Thanks for your note. We are scientists and we are all quite used to being criticized, so no worries on that score. I'll chat with my co-authors on Monday and we'll do our best to get back to you next week. In the meantime, we have three questions:

(a) Are you saying you are going to submit the paper you attached and that you hope we might collaborate on a different paper, or are you saying that rather than submitting the version of the paper you attached, you'd prefer to submit a collaborative version? In either case, I will let you know that I find the title terribly offensive. "Too good to be true" is a stock phrase in English that clearly suggests chicanery (it is from the proverb "If something seems too good to be true, then it probably is"). Most journalists will assume that what follows the title are accusations of impropriety, and most journalists do not have the technical skills to understand your argument, so the title will be about as far as they get. If you don't intend to accuse us of unethical behavior (and I hear you saying you don't) then I'd like to suggest changing it. I am usually a big fan of glib titles, but not when one is dealing with people's professional reputations.

(b) Of the many papers you could have chosen, why did you choose this one?

(c) How many other papers that were written by any or all of us have you analyzed?

Thanks,

-d.

**Daniel Gilbert**
**Edgar Pierce Professor of Psychology**

Harvard University | Department of Psychology
33 Kirkland Street, William James Hall 254, Cambridge, MA 02138 USA

617.495.3892 | gilbert@wjh.harvard.edu
danielgilbert.com | @DanTGilbert

From: Gregory Francis <gfrancis@purdue.edu> Sent: Fri 06-27-14 10:50 AM
To: Gilbert, Daniel
Cc: morewedge@gmail.com; tdw@virginia.edu
Subject: Re: Morewedge et al. (2010)

Hi Dan,

Thanks for the quick response.

(a) The document I sent is only a draft (or really just a starting point of a draft, thanks for the comments on the title). My preference would be for all of us to work together on a paper that includes this analysis with much more discussion. If there is no interest in working together than I would submit some variation of the draft manuscript. (That sounds a bit like a threat, but I really don't mean it that way. I think there is a serious need for a discussion about how scientists acting ethically can end up with results that don't match up with their conclusions. I cannot write that paper myself, except in an abstract sense, because it requires input from the original authors.)

(b) It has four or more experiments and provides enough statistical information to compute the probability estimates. I did deliberately choose an article co-authored by Dan and Tim because I saw their blog posts about replications and negative psychology, and I thought it might be interesting to work together to discuss broader issues.

(c) One. That analysis is part of a manuscript that investigates a large set of studies in a single journal over a certain time span. The article was not "selected" just the journal and the time span. My manuscript is currently under review.

I will be out of town much of next week, so if I do not quickly get back to you it does not mean I am not interested.

Best wishes,

Greg Francis

Professor of Psychological Sciences
Purdue University

On 27 Jun 2014, at 04:58 pm, Gilbert, Daniel <gilbert@wjh.harvard.edu> wrote:

Thanks, Greg.

I'm sorry that I'm misunderstanding you and it is probably my fault. You say you deliberately chose this paper because Tim and I are among its authors, but then you say that it is part of a larger analysis of all the papers published in JESP over some time span. I don't see how these two statements are compatible? Can you clarify?

And BTW, I don't take your explanation below as a threat. It is perfectly reasonable for you proceed on your own if we decide not to accept your kind invitation.

Best,

-d.

DANIEL GILBERT
Edgar Pierce Professor of Psychology

Harvard University | Department of Psychology
33 Kirkland Street, William James Hall 254, Cambridge, MA 02138 USA

617.495.3892 | gilbert@wjh.harvard.edu
danielgilbert.com | @DanTGilbert

From: Gregory Francis <gfrancis@purdue.edu> Sent: Fri 06-27-14 11:03 AM
To: Gilbert, Daniel
Cc: morewedge@gmail.com; tdw@virginia.edu
Subject: Re: Morewedge et al. (2010)

The answer in (c) was in response to your question labeled (c). You asked about other papers I have analysed. My response (c) was to that question only. To rephrase my answer, there is one other paper by one of you that I have analysed. That analysis is part of a larger set for a journal (not JESP, as it turns out).

Greg Francis

Professor of Psychological Sciences
Purdue University

On 27 Jun 2014, at 05:10 pm, Gilbert, Daniel <gilbert@wjh.harvard.edu> wrote:

Okay, I think I understand now: (a) You analyzed one of our published papers for a paper that you wrote that is now "under review." In your "under review" paper you chose a large set of targets by selecting a certain time period of a certain journal, and our paper just happened to be among them. (b) In the paper you sent us as an attachment, you specifically targeted our published paper in JESP because Wilson and I are its authors and we have spoken out on our blogs about problems with the so-called "replication revolution."

If that's correct then I do indeed understand you, and I appreciate your having taken the time to clarify.

-d.

DANIEL GILBERT
Edgar Pierce Professor of Psychology

Harvard University | Department of Psychology
33 Kirkland Street, William James Hall 254, Cambridge, MA 02138 USA

617.495.3892 | gilbert@wjh.harvard.edu
danielgilbert.com | @DanTGilbert

Morewedge Paper

From: Gregory Francis <gfrancis@purdue.edu> Sent: Fri 06-27-14 11:11 AM
To: Gilbert, Daniel
Cc: morewedge@gmail.com; tdw@virginia.edu
Subject: Re: Morewedge et al. (2010)

Yes, that's the situation.

Greg Francis

Professor of Psychological Sciences
Purdue University