The Research Reproducibility Crisis and Economics of Science

Zacharias Maniadis and Fabio Tufano[1]

Abstract: We provide a brief summary of two areas where cross-fertilization across economics and other disciplines is likely to have far-reaching benefits. The increasing concern about research reproducibility entails that economic design has much to contribute to the discussion of possible reforms in science, while the empirical discipline of meta-research can inform practices to assess the validity of the economics literature. A mutual investment in investigating possible synergies may be costly but could benefit the scientific endeavour as a whole.

[1] Maniadis: Economics Department, School of Social Sciences, University of Southampton, Southampton SO17 1BJ, UK (email: Z.Maniadis@soton.ac.uk). Tufano: CeDEx, School of Economics, University of Nottingham, University Park, Nottingham NG7 2RD, UK (e-mail: fabio.tufano@nottingham.ac.uk).

We thank the participants to the 'Credibility of empirical research' conference (Southampton, July 2015), as well as Colin F. Camerer, Alfredo Di Tillio, Thomas Gall, Chris Doucouliagos, John P.A. Ioannidis, David K. Levine, John A. List, Marco Ottaviani, Constantine Sedikides, Peter N. Sorensen, Tom D. Stanley, Patrick Sturgis and our colleagues at the University of Southampton and the University of Nottingham for the insightful interactions that over the years have helped us to shape our thinking on the subject matter. The usual disclaimer applies.

1. The Reproducibility Crisis and Resulting Initiatives

> Just as chemical engineers are called upon not merely to understand the principles that govern chemical plants, but to design them, and just as physicians aim not merely to understand the biological causes of disease, but their treatment and prevention, a measure of the success of microeconomics will be the extent to which it becomes the source of practical advice, solidly grounded in well tested theory, on designing the institutions through which we interact with one another.
>
> (Roth, 1991, p. 113)

Increasing concern about the reproducibility of scientific findings has been raised in the recent years. The relevant discussion has been particularly heated in academia – especially in biomedical and psychological disciplines (e.g., Ioannidis, 2005, 2012; Prinz *et al.*, 2011; Open Science Initiative, 2015; Simmons *et al.*, 2011) – and the popular press has followed up, with articles in outlets such as the *Economist*, the *New Yorker* and the *New York Times*. It should be emphasized that the aforementioned disciplines have a long empirical tradition in assessing how knowledge accumulates, having been pioneers in the field of meta-analysis (e.g., Cochran, 1937; Glass, 1976; Bero and Rennie, 1995). Hence, it is not clear that more discussion of the reproducibility of findings in these disciplines entails the existence of a larger problem relative to other fields. It could simply reflect a different research focus.

The accumulated empirical evidence has led researchers in these disciplines to pay increasing attention to researcher incentives (e.g., Bakker *et al*., 2012; Nosek *et al*., 2012; Ioannidis, 2014), with a focus on how to improve research practices. A deeper analysis of incentives was left aside, however, given the urgency of the problem, and plausible intuitive solutions were proposed in order to enhance transparency and reduce researchers' bias and degrees-of-freedom (e.g., Simmons *et al*., 2011; Landis *et al*., 2012; Nosek *et al*., 2012; Fanelli, 2013; Miguel *et al*., 2014). A series of reforms was accordingly implemented, with the most important ones being the imposition of reporting guidelines, open data, and

preregistration at the journal and funder level (e.g., De Angelis *et al*., 2004; Simer *et al*., 2010; Eich, 2014; Collins *et al*., 2014; McNutt, 2014). As far as we know, rigorous policy assessment did not take place, although ex-post studies have found some evidence of success (e.g., Kane *et al*., 2007; Kaplan and Irving, 2015; Kindwell *et al*., 2016).

In several other disciplines there is an active discussion whether similar policies should be imposed (e.g., Nyhan, 2015; Maniadis *et al*., 2014, 2015; Cook, 2016; Byington and Felps, 2017), but the relative deficit in quantitative research synthesis makes this assessment difficult (but, in social sciences, note the seminal work of Franco *et al*., 2014). In economics, there is empirical work (particularly recent) on the reproducibility of economic findings, pointing in different directions, and with some subfields faring better than others. In pioneering work, Dewald *et al*. (1986), and Delong and Lang (1992) estimated both 'methods reproducibility'[2] and the false negatives rate in economics. However, only recently have economists returned to the issue. Chang and Li (2015) found that reproducibility has improved since 1986, but with plenty of room for further improvement. Brodeur *et al*. (2016) analysed statistical tests reported in top economic journals and found a pattern of p-values indicating considerable bias. On the other hand, the replication initiative by Camerer *et al*. (2016) provides evidence that experimental economics fares relatively well regarding replicability. In addition, there have been already new important policy initiatives: for instance, top economics journals require posting data and code for accepted papers and the Journal of the Economics Science Association (devoted to the use of experimental methods in economics) was launched also to promote reproducibility (Nikiforakis and Slonim, 2015).

We posit that this state of affairs generates a double opportunity for cross-fertilization between economics and other empirical disciplines. Two major challenges are common to all disciplines: 1) how to measure the degree of reproducibility; 2) how to understand

[2] This concept has been recently defined as '… the provision of enough detail about study procedures and data so the same procedures could, in theory or in actuality, be exactly repeated'. (Goodman *et al.*, 2016, p.2)

researchers' incentives and their possible responses to reforms in scientific practices. As we argued before, considerable resources have been put into policy reforms that lack rigorous evaluation. This is where economics can help: researchers from afflicted disciplines should be interested in the insights offered by rigorous economic tools, especially design economics. The contribution by Di Tillio *et al*. (forthcoming) published in this journal issue illustrates this approach from a theory viewpoint. On the other hand, by applying (whenever disciplinary differences allow) empirical methods from the emerging discipline of meta-research (Ioannidis *et al*., 2015) economists can systematically assess how the field of economics fares in terms of reproducibility. In this same journal issue, Ioannidis *et al*. (forthcoming) and Maniadis *et al*. (forthcoming) contribute in this aspect by examining whether and how economics can benefit from meta-research methods.

2. Economics of Science and Design Economics

Economic methodology can bring clarity to the examination of mechanisms behind the reproducibility crisis and to the systematic evaluation of possible policy solutions. The study of behavioural responses to incentives embodied in institutions is not the main focus of most scientific disciplines. Hence, there is a clear role for economic methods in assessing reform proposals by examining systematically the trade-offs and interdependencies of researcher behaviour in the scientific environment. In particular, we believe that there is considerable promise in economic design (Roth, 2002), that is, the combination of mathematical modelling and experiments used in assessing the effects of policy and institutional changes. Gall *et al*. (2017) forcefully argue this point to a biomedical audience by reviewing relevant work in economics. We shall now go over some of the key contributions briefly, delving in more detail into the contribution of Di Tillio *et al*. (forthcoming).

A key challenge is to elicit the information necessary for reproducibility from researchers in an incentive-compatible and cost-effective manner. A useful class of games for analysing this type of interaction is persuasion games (Milgrom, 2008). A sender provides verifiable information to a receiver, having an incentive to affect her behaviour. If everybody knows the sender's set of possible messages, the (rational) receiver anticipates the sender's exaggeration and infers the truth. Thus, there is no need for external intervention to improve knowledge accumulation. Henry and Ottaviani (2017) study the possible policies of receivers, such as medical authorities in drug approval procedures, who may commit to approve any drug if the success rate in trials exceeds a threshold level chosen by the authority. Again, rational authorities will expect the sender to choose the number of trials strategically and choose a reasonably conservative approval threshold. Henry (2009) examines a setup where the sender first determines how much research to perform and then what to disclose to the receiver, yielding incentives to conduct excessively many trials and to selectively report results. If the receiver understands these strategic considerations, it follows that in any equilibrium of the game the sender conducts an excessive number of trials and, perhaps slightly surprisingly, reveals all trial results.

Di Tillio *et al*. (forthcoming) attack directly the problem of information revelation in science. They consider the case of a researcher who aims at persuading an evaluator to grant approval of a treatment. The researcher implements an experiment with two conditions: a treatment and a baseline condition. The evaluator observes the experimental findings and, based on how convincing the evidence is, decides whether or not to approve the treatment. While the researcher always benefits from acceptance, the evaluator does so only when the benefit of acceptance outweighs its cost. The researcher has access to pre-existing private information, based on which they can take unobservable actions to make the evidence more convincing. In particular, they can do one of the following: (i) sample subjects conditional on

their treatment effect; (ii) assign subjects to conditions based on their baseline levels; (iii) selectively report the findings of the experiment. By taking any of these unobservable actions, the researcher manipulates the experiment (or its report) and induces a form of persuasion bias. The welfare implications of such bias are subtle. Despite the bias, the researcher's strategic manipulation introduces additional information to the system and, thus, the evaluator may be better off relatively to the case with no bias. However, the evaluator responds strategically to the researcher's behaviour by raising their standard of acceptance and, in equilibrium the researcher may be worse off relatively to the case with no bias. In fact, the researcher may find themselves in a confidence trap in which, if possible, they would rather prefer to commit themselves to no manipulation given the scepticism of the evaluator, who places very low confidence in the provided experimental evidence.

Lacetera and Zirulia (2011) find that policies purported to reduce the cost of monitoring or increase the rewards of successful publication can have different implications, which depend on the exact parameters of the model. Park *et al*. (2014) focus on the fact that learning by observation in the research arena can result to fads and herding. Accordingly, keeping the element of subjectivity in reviewers' recommendation can be beneficial because it safeguards against these problematic phenomena. Gall and Maniadis (2015) directly model the effect of journals imposing reporting guidelines in a tournament environment and robustly find that this policy supresses the overall rate of misrepresentation of research information. Felgenhauer and Schulte (2014) focus on the costs and benefits of experimentation. One conclusion from their analysis is that empirical evidence provided across different journals and disciplines should be assigned a different informational weight. For instance, in disciplines where generating new evidence is inexpensive and for articles submitted in elite journals, a much more conservative standard needs to be imposed.

Plott (1994) illustrates the principle that theoretical methods for assessing policies need to be complemented by empirical evidence, and in particular that experiments can be used as a 'testbed' to address the effects of institutional change. This approach often involves 'horse races' among several alternative institutions and it has been used to assess various real-life institutional challenges. It is well known that the design approach has had considerable success in several domains, including auction and market design and environmental policy (e.g., Ledyard *et al*., 1997; Plott, 1997; Kagel and Roth, 2000; Cason *et al*., 2003). The 'testbed approach' can be used to attack the problem of optimal reforms of research practices, allowing for a wide spectrum of experimentation ranging from laboratory validations of policy changes to randomised control trials in the field.

3. **What Can Economics Learn from Meta-Research?**

From an empirical perspective, one might ask whether economics itself suffers from the reproducibility problem. Ioannidis and Doucouliagos (2013) argue that the problem may affect economics, but much remains unknown. First of all, are there similarities between economics and experimental biomedical and behavioural sciences? It has been argued that with the evolution of experimental economics and recent advances in microeconometrics, empirical microeconomics now hinges much more on the experimental method (Angrist and Pischke, 2010). If so, it is worth examining whether economics can learn from the study of factors that affect the reproducibility of empirical – mainly experimental – research, which constitutes a discipline in its own right: meta-research.

This discipline has developed a varied methodology: metrics of how biased the published scientific evidence is (Ioannidis and Trikalinos 2007; Simmons *et al*. 2013), quantitative analysis of how knowledge accumulates over time (Cooper *et al*., 2009), and assessments of the fraction of findings that correspond to false positives (Wacholder *et al*,

2004; Ioannidis 2005). To illustrate further, let us focus on psychology. Cohen (1962) analysed the 1960 volume of the 'Journal of Abnormal and Social Psychology', and found a median power equal to 0.48. Bakker *et al*. (2012) provided a more general power estimate equal to 0.35. A similar tradition exists in quantifying the publication bias, the file-drawer problem, and the rate of 'results reproducibility.'[3] Bakker *et al*. (2012) provided quantitative evidence regarding the degree of selection and publication bias, finding indications of biases and excess significant results in 7 out of 13 meta-analyses they considered. Moreover, Cooper *et al*. (1997) found that the fraction of studies that are not submitted for publication following data analysis exceeds 60%. Makel *et al*. (2012) examined the 100 most cited psychology journals and estimated the percentage of them that represent replications. They found that about 1% are replications, most of which are successful; about 80% of the total replications are conceptual rather than exact replications.

In economics, analogous evidence is scarce. Zhang and Ortmann (2013) estimated a median power of 0.25 for experimental studies of the dictator game included in Engel's (2011) meta-analysis. Duvendack *et al*. (2015) review replications in economics – but in non-random design that makes it difficult to understand how common replication is. These are important first steps, but much more needs to be done in terms of examining whether meta-research methods for retrospectively estimating power and assessing results reproducibility can be successfully employed in economics. This is where the studies of Ioannidis *et al*. (forthcoming) and Maniadis *et al*. (forthcoming) contribute to fill important gaps.

Ioannidis *et al*. (forthcoming) examine quantitatively 159 meta-analyses relating to different empirical areas of economics research (e.g., labour economics, macroeconomics, etc.) to calculate the proportion of reported findings that are adequately powered. The authors employ four alternative meta-analytical approaches in order to obtain conservative and robust

[3] This is defined as 'obtaining the same results from the conduct of an independent study whose procedures are as closely matched to the original experiment as possible'. (Goodman *et al*., 2016, pp. 2-3).

estimations of the true statistical power of empirical studies in economics research. They find that empirical research in economics is generally underpowered with half of the economics' areas surveyed having 10.5% or less of their studies adequately powered. The median statistical power in empirical economics research turns out to be not larger than 18%. An important implication of the overall inadequate power of empirical research in economics is that a sizable majority of its studies have less than 50% probability of detecting the phenomenon under investigation. The authors also introduce a novel meta-analytical approach to discount for the existence of a possible bias in empirical economics research. Contrary to the standard practice in meta-analyses of including each and every available study, they estimate the empirical effect in a given area (or sub-area) of research by including only adequately powered studies, while discarding all the others. Arguably, such a meta-analytical approach is likely to reduce the research bias leading to the stark finding that almost 80% of the empirical effects reported are importantly inflated.

Maniadis *et al*. (forthcoming) apply both theoretical and empirical meta-research methods to economics research, focusing on replication. They first extend the Ioannidis (2005) framework and examine how independent replications can affect the relevant Post-Study Probability (PSP) that a result is true and how potential bias in the conduct of replications affects inference. Adversarial (sympathetic) bias means that a fixed fraction of time in which a replication study should be declared positive (negative), is declared negative (positive) instead. Suppose a positive result has recently been discovered, and there are several replication attempts. The model finds that in disciplines with adversarial regimes, society's updating on the basis of a fixed number of replications should be larger relative to a discipline with neutral researchers, and much larger than in disciplines with sympathetic replicators. Moreover, when one looks at mixed evidence *ex post*, higher average power does not necessarily increase our confidence in a phenomenon being true. The authors then

conduct a pilot empirical study focusing on experimental economics. By reviewing a large sample of studies, they estimate the prevalence of replications in the literature and study the determinants of 'replication success'. In doing so, they reveal the challenges of doing systematic meta-research in economics also because of the scarce use of the term 'replication'. Overall, the results are compatible with the psychological meta-research article of Makel *et al*. (2012). However, standardization of terminology and systematization is needed for future systematic meta-research attempts. Since these are costly activities, it seems that further investment is needed to convince the economic audience for the added value of meta-research.

4. **Conclusions**

Often disciplinary boundaries artificially restrict interaction and cross-fertilization that may be greatly beneficial. This is especially the case for scientific issues that interest many disciplines. The alleged 'crisis of reproducibility' is a case in point. We delimitate two lines of possibly fruitful interdisciplinary interaction: first, design economics can potentially be useful to help assess how to induce researchers to follow practices that enhance reproducibility. Second, empirical methods from the emerging discipline of meta-research have the promise of enhancing our understanding of the reproducibility of economics.

The articles of the current issue illustrate these two areas of potential cross-fertilization. However, the empirical work of Maniadis *et al*. (forthcoming) offers a cautionary tale, showing that we should not be overly confident that methods seamlessly translate across disciplines.

University of Southampton

University of Nottingham

References

Angrist, J. and Pischke, J-S. (2010). The credibility revolution in empirical economics: how better research design is taking the con out of econometrics', Working Paper No. w15794. National Bureau of Economic Research.

Bakker, M., van Dijk, A. and Wicherts, J.M. (2012). 'The rules of the game called psychological science', *Perspectives on Psychological Science*, vol. 7(6), pp. 543–554.

Bero, L. and Rennie, D. (1995). 'The Cochrane Collaboration: preparing, maintaining, and disseminating systematic reviews of the effects of health care', *JAMA*, vol. 274(24), pp. 1935–1938.

Brodeur, A., Le, M., Sangier, M. and Zylberberg, Y. (2016). 'Star wars: the empirics strike back', *American Economic Journal: Applied Economics*, vol. 8(1), pp. 1–32.

Byington, E., and Felps, W. (2017). 'Solutions to the credibility crisis in management science'. *Academy of Management Learning & Education*, vol. 16(1), pp. 142–162.

Camerer, C.F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M. and Wu, H. (2016). 'Evaluating replicability of laboratory experiments in economics', *Science*, vol. 351(6280), pp. 1433–1436.

Cason, T.N., Gangadharan, L. and Duke, C. (2003). 'Market power in tradable emission markets: a laboratory testbed for emission trading in Port Phillip Bay, Victoria', *Ecological Economics*, vol. 46(3), pp. 469–491.

Chang, A.C. and Li, P. (2015). 'Is economics research replicable? Sixty published papers from thirteen journals say 'usually not'', Finance and Economics Discussion Series 2015-083. Washington: Board of Governors of the Federal Reserve System, http://dx.doi.org/10.17016/FEDS.2015.083

Cochran, W.G. (1937). ‘Problems arising in the analysis of a series of similar experiments’, *Supplement to the Journal of the Royal Statistical Society*, vol. 4(1), pp. 102–118.

Cohen, J. (1962). ‘The statistical power of abnormal-social psychological research: a review’, *Journal of Abnormal and Social Psychology*, vol. 65(3), pp. 145–153.

Collins, F.S., and Tabak, L.A. (2014). ‘NIH plans to enhance reproducibility’, *Nature*, vol. 505.7485, p. 612.

Cook, B. G. (2016). ‘Reforms in academic publishing: should behavioral disorders and special education journals embrace them?’, *Behavioral Disorders*, vol. 41(3), pp. 161–172.

Cooper, H., DeNeve, K. and Charlton, K. (1997). ‘Finding the missing science: the fate of studies submitted for review by a human subjects committee’, *Psychological Methods* vol. 2(4), pp. 447–452.

Cooper, H., Hedges, L.V. and Valentine, J.C. (2009). *Handbook of Research Synthesis and Meta-Analysis*. Russell Sage Foundation.

De Angelis, C., Drazen, J.M., Frizelle, F.A., Haug, C., Hoey, J., Horton, R., Kotzin, S., Laine, C., Marusic, A., Overbeke, A.J.P. and Schroeder, T.V. (2004). ‘Clinical trial registration: a statement from the International Committee of Medical Journal Editors.’, *New England Journal of Medicine*, vol. 351, pp. 1250-1251

De Long, J.B., and Lang, K. (1992). ‘Are all economic hypotheses false?’, *Journal of Political Economy*, vol. 100(6), pp. 1257–1257.

Dewald, W.G., Thursby, J.G. and Anderson, R.G. (1986). ‘Replication in empirical economics: the journal of money, credit and banking project’, *The American Economic Review*, vol. 76(4), pp. 587–603.

Di Tillio, A., Ottaviani, M. and Sorensen, P. (Forthcoming). ‘Persuasion bias in science: can economics help?’, *Economic Journal.*

Duvendack, M., Palmer-Jones, R.W. and Reed, W.R. (2015). 'Replications in economics: a progress report', *Econ Journal Watch*, vol. 12(2), pp. 164–191.

Eich, E. (2014). 'Business not as usual.' *Psychological Science*, vol. 25(1), pp. 3–6.

Engel, C. (2011). 'Dictator games: a meta study.' Experimental Economics, vol. 14(4), pp. 583–10.

Fanelli, D. (2013). 'Redefine misconduct as distorted reporting', *Nature*, vol. 494.7436, p. 149.

Felgenhauer, M. and Schulte, E. (2014). 'Strategic private experimentation', *American Economic Journal: Microeconomics*, vol. 6(4), pp. 74–105.

Franco, A., Malhotra, N. and Simonovits, G. (2014). 'Publication bias in the social sciences: unlocking the file drawer', *Science*, vol. 345(6203), pp. 1502–1505.

Gall, T., Ioannidis, J. P. A. and Maniadis, Z. (2017). 'The credibility crisis in research: can economics tools help?', *PLoS Biology*, *15*(4), p. e2001846.

Gall, T., and Maniadis, Z. (2015). 'Evaluating solutions to the problem of false positives', University of Southampton Discussion Paper in Economics and Econometrics 1504.

Glass, G. V. (1976). 'Primary, secondary, and meta-analysis of research', *Educational Researcher*, vol. 5(10), pp. 3–8.

Goodman, S. N., Fanelli, D. and Ioannidis, J.P.A. (2016). 'What does research reproducibility mean?', *Science Translational Medicine*, vol. 8(341), pp. 341ps12-341ps12.

Henry, E. (2009). 'Strategic disclosure of research results: the cost of proving your honesty', *The Economic Journal*, vol. 119(539), pp.1036–1064.

Henry, E., and Ottaviani, M. (2017). 'Research and the approval process: the organization of persuasion', London, Centre for Economic Policy Research, http://www.cepr.org/active/publications/discussion_papers/dp.php?dpno=11939 (last accessed: 3 May 2017).

Ioannidis, J.P.A. (2005). 'Why most published research findings are false', *PLoS medicine* vol. 2(8), p. e124.

Ioannidis, J.P.A. (2012). 'Why science is not necessarily self-correcting', *Perspectives on Psychological Science*, vol. 7(6), pp. 645–654.

Ioannidis, J.P.A. (2014). 'How to make more published research true', *PLoS Medicine*, vol. 11.10, pp. e1001747.

Ioannidis, J.P.A. and Doucouliagos, C. (2013). 'What's to know about the credibility of empirical economics?', *Journal of Economic Surveys*, vol. 27(5), pp. 997–1004.

Ioannidis J.P.A., Fanelli D., Dunne D.D. and Goodman S.N. (2015). 'Meta-research: evaluation and improvement of research methods and practices', *PLoS Biol*, vol. 13(10), pp. e1002264.

Ioannidis, J.P.A., Stanley, T. D. and Doucouliagos H. (forthcoming). 'The power of bias in economics research', *Economic Journal*.

Ioannidis, J.P.A. and Trikalinos, T.A. (2007). 'An exploratory test for an excess of significant findings', *Clinical Trials*, vol. 4(3), pp. 245–253.

Kagel, J. H. and Roth, A. E. (2000). 'The dynamics of reorganization in matching markets: a laboratory experiment motivated by a natural experiment', *Quarterly Journal of Economics*, vol. 115, pp. 201–235.

Kane, R. L., Wang, J. and Garrard, J. (2007). 'Reporting in randomized clinical trials improved after adoption of the CONSORT statement', *Journal of Clinical Epidemiology*, vol. 60, pp. 241–249. doi:10.1016/j.jclinepi.2006.06.016

Kaplan, R. M. and Irvin, V. L. (2015). 'Likelihood of null effects of large NHLBI clinical trials has increased over time', *PLoS One*, vol. 10(8), p. e0132382. doi:10.1371/journal.pone.0132382

Kidwell, M.C., Lazarević, L.B., Baranski, E., Hardwicke, T.E., Piechowski, S., Falkenberg, L.-S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C., Errington, T.M., Fiedler, S. and Nosek B.A. (2016). 'Badges to acknowledge open practices: a simple, low-cost, effective method for increasing transparency', *PLoS Biology*, vol. 14(5), p. e1002456.

Lacetera, N., and Zirulia, L. (2011). 'The economics of scientific misconduct', *Journal of Law, Economics, and Organization*, vol. 27(3), pp. 568–603.

S. C. Landis, Amara, S.G, Asadulah, K., Austin, C.P, Blumenstein, R., Bradley, E.W., Crystal, R.G., Darnell, R.B., Ferrante, R.J., Fillit, H., Finkelstein, R., Fisher, M. Gendelman, H.E., Golub, Goudreau, J.L., Gross, R.A., Gubitz, A.K., Hesterlee, S.E., Howells, D.W., Huguenard, J., Kelner, K., Koroshetz, W., Krainc, D., Lazic, S.E., and Levine M.S. (2012). 'A call for transparent reporting to optimize the predictive value of preclinical research', *Nature*, vol. 490(7419), pp. 187–191.

Ledyard, J., Porter, D. and Rangel, A. (1997), 'Experiments testing multi object allocation mechanisms', *Journal of Economics and Management Strategy,* vol. 6, pp. 639–675.

Makel, M., Plucker, J. and Hegarty, B. (2012). 'Replications in psychology research: how often do they really occur?', *Perspectives on Psychological Science,* vol. 7(6), pp. 537–541.

Maniadis, Z., Tufano, F. and List J.A. (2014). 'One swallow doesn't make a summer: new evidence on anchoring effects', *American Economic Review*, vol. 104(1), pp. 277–290.

Maniadis, Z., Tufano, F. and List J.A. (2015). 'How to make experimental economics research more reproducible: lessons from other disciplines and a new proposal', in (Eds. C. Deck, E. Fatas, E., and T. Rosenblat, eds.), *Research in Experimental Economics*, vol. 18: Replication in Economic Experiments. pp. 215–230, Emerald Group Publishing.

Maniadis, Z., Tufano, F. and List J.A. (forthcoming). 'To replicate, or not to replicate? Exploring reproducibility in economics through the lens of a model and pilot study', *Economic Journal.*

McNutt, M. (2014). 'Journals unite for reproducibility', *Science,* vol. 346(6210), pp. 679–679.

Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K.M., Gerber, A., Glennerster, R., Green, D.P., Humphreys, M., Imbens, G., Laitin, D., Madon, T., Nelson, L., Nosek, B. A., Petersen, M., Sedlmayr, R., Simmons, J.P., Simonsohn, U. and Van der Laan M. (2014). 'Promoting Transparency in Social Science Research', *Science*, vol. 343(6166), pp. 30–31.

Milgrom, P. (2008). 'What the seller won't tell you: persuasion and disclosure in markets', *The Journal of Economic Perspectives*, vol. 22(2), pp. 115–131.

Nikiforakis, N., and Slonim, R. (2015). 'Editors' preface: introducing JESA', *Journal of the Economic Science Association,* vol. 1(1), pp. 1–7.

Nosek, B.A., Spies, J.R. and Motyl, M. (2012). 'Scientific utopia II. Restructuring incentives and practices to promote truth over publishability', *Perspectives on Psychological Science*, vol. 7(6), pp. 615–631.

Nyhan, B. (2015). 'Increasing the credibility of political science research: a proposal for journal reforms', *PS: Political Science & Politics,* vol. 48(S1), pp. 78–83.

Open Science Initiative (2015). 'Estimating the reproducibility of psychological science', *Science*, vol. 349 (6251), pp. aac4716–8.

Park, I-U., Peacey, M.W. and Munafò, M.R. (2014). 'Modelling the effects of subjective and objective decision making in scientific peer review', *Nature*, vol. 506(7486), pp. 93–96.

Plott, C.R. (1994). 'Market architectures, institutional landscapes and testbed experiments', *Economic Theory,* vol. 4(1), pp. 3–10.

Plott, C.R. (1997). 'Laboratory experimental testbeds: application to the PCS auction', *Journal of Economics & Management Strategy,* vol. 6(3), pp. 605–638.

Prinz, F., Schlange, T. and Asadullah, K. (2011). 'Believe it or not: how much can we rely on published data on potential drug targets?', *Nature Reviews Drug Discovery,* vol. 10(9), pp. 712–712.

Roth, A.E. (1991). 'Game theory as a part of empirical economics', *The Economic Journal,* vol. 101(404), pp. 107–114.

Roth, A.E. (2002). 'The economist as engineer: game theory, experimentation, and computation as tools for design economics', *Econometrica,* vol. 70(4), pp. 1341–1378.

Simer, I., Moher, D., Hirst, A., Hoey, J., Schulz K.F. and Altman D.G. (2010). 'Transparent and accurate reporting increases reliability, utility, and impact of your research: reporting guidelines and the EQUATOR Network', *BMC Medicine,* vol. 8(24), pp. 1–6.

Simmons, J. P., Nelson, L.D. and Simonsohn, U. (2011). 'False-positive psychology undisclosed flexibility in data collection and analysis allows presenting anything as significant', *Psychological Science,* vol. 22(11), pp. 1359–1366.

Simmons, J. P., Nelson, L.D. and Simonsohn, U. (2013). 'P-curve: a key to the file drawer', *Journal of Experimental Psychology: General*, vol. 143(2), pp. 534–547.

Wacholder, S., Chanock, S., Garcia-Closas, M., El-ghormli, L. and Rothman, N. (2004). 'Assessing the probability that a positive report is false: an approach for molecular epidemiology studies', *Journal of the National Cancer Institute,* vol. 96(6), pp. 434–442.

Zhang, L. and Ortmann, A. (2013). 'Exploring the meaning of significance in experimental economics', UNSW Australian School of Business Research Paper 2013-32.