

Crisis in Cognitive Science? Rise of the Undead Theories

Moritz Heene (heene@psy.lmu.de)¹

James Coyne (jcoyne@mail.med.upenn.edu)²

¹Department of Psychology, Ludwig Maximilian University of Munich, Germany

²Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, USA

Greg Francis (gfrancis@purdue.edu)

Department of Psychological Sciences, Purdue University, Indiana, USA

Phil Maguire (pmaguire@cs.nuim.ie)¹

Rebecca Maguire (rebecca.maguire@ncirl.ie)²

¹Department of Computer Science, NUI Maynooth, Ireland

²School of Business, National College of Ireland, IFSC, Dublin 1, Ireland

Abstract

The controversy that followed the publication of Bem's (2011) surprising results had the merit of focusing discussions on the validity of the current paradigm in psychological science. It has been argued that a tendency of journals to avoid publishing null results, in addition to the further extinguishing of null results through questionable researcher practices, is leading to the promulgation of a multitude of 'undead' theories that have little basis in fact (Ferguson & Heene, 2012; Francis, 2012). At the same time, the field of AI has met with success precisely by abandoning theory, prompting Noam Chomsky to remark that it is deviating from anything previously seen in the history of science (see Cristianini, 2014). In this symposium the speakers consider whether current theories offered by psychological science are valuable, whether there might be inherent obstacles which are preventing the identification of valuable theories, and whether psychological theory even matters at all.

Keywords: Publication bias, falsification, fail-safe number, excess significance, overfitting, theory resistance.

Overview

The speakers of this symposium will discuss the reliability and value of psychological theory from multiple perspectives, pooling expertise from a range of disciplines including psychology, artificial intelligence, and algorithmic information theory. **Moritz Heene** is Professor for Learning Sciences Research Methodologies at Ludwig-Maximilians University, and **James Coyne** is Professor of Psychology in Psychiatry at the University of Pennsylvania. They will argue that the Fail-Safe-Number used in psychological meta-analyses is too weak to provide a reliable mechanism for falsification. **Greg Francis** is Professor of Psychological Sciences at Purdue University, with a research interest in replication, statistical consistency and publication bias. He will discuss his finding that the large majority of publications in empirical psychology include statistics which are "too good to be true". Invited speaker **Nello Cristianini** is Professor of Artificial Intelligence at the University of Bristol and a current holder of the Royal Society-Wolfson Merit Award. He is well known in the field of A.I. for his research contributions to the statistical analysis of learning algorithms, with application to support vector machines and kernel methods. Cristianini will provide a review of recent developments in A.I., high-

lighting a paradigm shift away from theoretical understanding towards prediction based on the statistical analysis of big data. **Phil Maguire**, Director of Computational Thinking at NUI Maynooth, will conclude the symposium by offering an algorithmic information theory perspective on why theories describing complex phenomena have lower generalizability than those in the hard sciences.

Heene and Coyne: The failure of meta-analysis as a weapon against undead theories

Publication bias arises whenever the publication probability depends on the statistical significance of the results. Recent research (Fanelli, 2010) has shown that theory supportive results are far more prevalent in psychology and psychiatry than in the hard sciences (91.5% versus 70.2% in the space sciences, for instance). The resulting tendency of psychological science to avoid publishing null results not only severely limits the replicability requirement of science, but also the validity of meta-analyses (Ferguson & Heene, 2012).

Rosenthal's Fail-Safe-Number (FSN) (Rosenthal & Rubin, 1978) aims to estimate the number of unpublished studies in a meta-analysis required to bring the meta-analytic mean effect size down to a statistically insignificant level. We conducted a quantitative literature survey regarding the application of the FSN within psychology ranging from 1979 until 2013. Despite early proofs of its invalidity (e.g. Iyengar & Greenhouse, 1988), we found that its application grew exponentially over the decades, illustrating the persistent shielding of patently false claims from timely correction. We argue that questionable research practices severely reduce psychological science's capability to sustain a proper mechanism for theory falsification, resulting in numerous 'undead' theories that have little basis in fact (Ferguson & Heene, 2012).

Francis: Cargo cult thinking in psychology

In a commencement address at Caltech in 1974, physicist Richard Feynman (1974) described "cargo cult science", where investigations appear to follow scientific methods but do not actually produce scientific results. The term refers to societies in the South Pacific that received aid from the allied

forces during World War II. Once the war ended, the local residents unsuccessfully attempted to bring airplanes loaded with supplies back to the island by building elaborate props and emulating the air traffic control and landing crews. Of course, such efforts were futile because the residents were merely acting out the behavior of the airport staff. Feynman (1974) suggested that some scientists likewise emulate the scientific method without actually understanding and following the principles of science.

Exploring this claim, I applied the Test for Excess Significance (Ioannidis & Trikalinos, 2007; Francis 2012) to detect the presence of bias in the journal *Psychological Science*. The probability of experimental success (power) can be estimated from published experiments, and used to estimate the probability of success for a set of experiments. When findings across experiments are “too good to be true”, readers should suspect that null findings were suppressed, the experiments or analyses were improper, or that the theory does not properly follow from the data. For the 44 articles where the analysis could be applied, over 80% of the articles appear to be biased. Consistent with Feynman’s claim, psychological scientists may fully believe they are following good scientific practice while simultaneously producing fallacious arguments.

Cristianini: On the current paradigm in AI

The field of Artificial Intelligence (AI) has undergone many transformations, most recently the emergence of data-driven approaches centred on machine learning technology. The present talk examines that paradigm shift by analysing the contents of the longest running conference series in the field using the conceptual tools developed by Thomas Kuhn.

A paradigm shift occurs when a new set of assumptions and values replaces the previous one within a given scientific community. These are often conveyed implicitly, by the choice of success stories that exemplify and define what a given field of research is about, demonstrating what kind of questions and answers are appropriate, or: “what success looks like”. The replacement of these exemplar stories corresponds to a shift in goals, methods, and expectations. The most recent such transition in the field of AI will be discussed, taking into account the appropriateness of analogous approaches to different sciences and domains of society (see Cristianini, 2014).

A debate has recently arisen when Noam Chomsky pointed out that the current approach to AI corresponds to adopting a new and different criterion of success than that traditionally accepted in science. Similar changes have been seen in genomics and other fields: predictions can sometimes be obtained more easily than theoretical understanding. The question we face involves the role of science: is it to provide predictions or a better understanding of reality?

Maguire: Theory resistance in complex phenomena

We explore the idea that the more complex a phenomenon, the greater its resistance to theoretical modelling. Results in algorithmic information theory have shown that the accuracy

of prediction of a theory is linked to the level of compression it achieves (see Li & Vitányi, 2008). Inanimate physical phenomena, featuring high levels of redundancy which is easily extracted, are particularly well suited to theoretical modelling. In contrast, the mind is a complex phenomenon which is highly integrated (Maguire et al., 2014). Because of this integration, it is much more difficult to generalize from a small set of observations. A large quantity of theory must be invested into neutralizing observer bias, minimizing the level of compression that can be achieved.

The dearth of null results in psychological journals may be symptomatic of the discipline’s complexity rather than reflecting a problem with review processes per se. When theoretical models tend to have low generalizability, failed replication studies hold less value than those that establish novel effects. Journals adapt by applying a shotgun strategy, presenting as many theories as possible in the hope of ‘getting lucky’. We consider the possibility that the scientific method is ill suited to the modelling of cognition, and that cognitive science might meet with greater success by following the paradigm shift in AI (Cristianini, 2014) and abandoning the restrictive goal of seeking to understand the mind.

References

- Bem, D. J. (2011). Feeling the future: experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of personality and social psych.*, 100(3), 407.
- Cristianini, N. (2014). On the current paradigm in artificial intelligence. *AI Communications*, 27(1), 37–43.
- Fanelli, D. (2010). “Positive” results increase down the hierarchy of the sciences. *PLoS One*, 5(4), e10068.
- Ferguson, C. J., & Heene, M. (2012). A vast graveyard of undead theories: publication bias and psychological sciences’ aversion to the null. *Perspectives on Psychological Science*, 7(6), 555–561.
- Feynman, R. P. (1974). Cargo cult science. *Engineering and Science*, 37(7), 10–13.
- Francis, G. (2012). Too good to be true: Publication bias in two prominent studies from experimental psychology. *Psychonomic Bulletin & Review*, 19(2), 151–156.
- Ioannidis, J. P., & Trikalinos, T. A. (2007). An exploratory test for an excess of significant findings. *Clinical Trials*, 4(3), 245–253.
- Iyengar, S., & Greenhouse, J. B. (1988). Selection models and the file drawer problem. *Stat. Science*, 3(1), 109–117.
- Li, M., & Vitányi, P. (2008). *An introduction to kolmogorov complexity and its applications*. Springer.
- Maguire, P., Moser, P., Maguire, R., & Griffith, V. (2014). Is consciousness computable? Quantifying integrated information using algorithmic information theory. In *Proceedings of the 36th annual conference of the cognitive science society*. Austin, TX: Cognitive Science Society.
- Rosenthal, R., & Rubin, D. B. (1978). Interpersonal expectancy effects: The first 345 studies. *Behavioral and Brain Sciences*, 1(03), 377–386.