OVERCOMING THE EXPERIMENTERS' REGRESS IN
BIOMEDICAL RESEARCH

Romina Zuppone
LOGOS, Universidad Nacional de Quilmes

Abstract: In his book Changing Order, Replication and Induction in Natural Sciences, Harry Collins coined the expression ‘experimenters’ regress’. This refers to an infinite regress that takes place in experimental practice that would prevent us from settling disagreements about experimental results by relying exclusively on scientific considerations. The goal of the paper is to evaluate the impact of the experimenters’ regress in biomedical research and an epistemic way of breaking out of it. In doing so, I will consider Teira’s contractarian approach to the use of debiasing procedures (2013) and argue that this constitutes a necessary condition for an internal answer to the way out of the experimenters’ regress. I will claim that the consensus regarding the implementation of debiasing procedures together with theoretical considerations are conjointly sufficient to justify a choice in the context of an experimental disagreement. The structure of the paper is as follows: in the first section I will explain why the study of the justification of experimental evidence has become so relevant in recent decades; in the second section I will present and reconstruct Collins’ argument; the third section will introduce Teira’s contribution and my proposal.

Keywords: Experimenter’s Regress; Biomedical Research; Mechanistic Reasoning.

Date Submitted: 26/09/2014
Date Accepted: 02/11/2014
OVERCOMING THE EXPERIMENTERS’ REGRESS IN BIOMEDICAL RESEARCH

1. Introduction

We usually agree in considering that the justification of empirical theories comes from their being tested against empirical evidence, and in that, much of the time, evidence is obtained by performing experiments. These considerations lead us to evaluate another question: how is experimental knowledge established? When is an experimental result justified?

Traditionally, it was thought that scientific evidence was judged and validated in several and different ways. To begin with, experimental results have to be reproduced until the chances of getting a spurious result are very low. Then the formal communication of results takes place by means of publication in specialized journals. Therefore, peer review is a different kind of constraint by which the cogency of the experimental design, the methodological strategies deployed, the robustness of the data obtained, and the plausibility of the interpretation of the data are judged. Finally, once the results of an experimental investigation become available to the scientific community, different research teams may be interested in testing them either by replicating the original experiment or by running a different procedure that, in turn, may or may not rely on the same theoretical knowledge assumed in the original experiment. This is how the official story goes. But as soon as we move from the abstract description of the scientific enterprise to the complexities of its actual practice, two problems become apparent.

The first is practical. There is evidence of peer review failure: Papers with serious flaws get published. Sokal’s hoax is a case in point with regard to testing standards for publication in the social sciences. In 1996, Sokal submitted a paper to Social Text, a leading journal for social sciences. Entitled “Transgressing the Boundaries: Towards a Transformative Hermeneutics of Quantum Gravity”, the paper was accepted for publication despite its lack of scientific rigour. Soon after the publication of the paper, Sokal made the hoax public, to the embarrassment of Social Text’s editors. In 1998, Fiona Godlee, editor of the British Medical Journal, sent an article to the journal’s referees in which she deliberately introduced 8 flaws. None of the referees were able to pick out all the errors. None of them was able to detect more than 2 mistakes out of the 8. Indeed, several referees were unable to detect all
of them even when they were told that there were 8 flaws and were invited to find them.

That the poor performance of referees impoverishes, in turn, the quality of specialized journals is, of course, a practical concern that does not prevent scientists from practicing good science. This fact would just show us that a more cautious review process is required. And even in the face of peer review failure, there is a different mechanism that is supposed to prevent us from considering those kinds of productions to be good scientific knowledge. I will consider it next.

The second problem is the theoretical classic picture which also tells us that even if an erroneous experimental report gets published it will not become accepted scientific knowledge because its mistakes will be neutralized by several different mechanisms. Among these mechanisms, the reproduction of experimental results by peers has traditionally been regarded as of the utmost importance in enabling the inter-subjectivity of scientific practice. Reproduction of an experiment can be done in several ways; however, according to Collins, disconfirming someone else’s findings requires performing a replication. So, let us suppose that a research group reproduced an experiment but failed to confirm the original findings (i.e. they got discordant results). Then, if we consider some studies done regarding replication of clinical trials we find that, for instance, Amgen, a well known company devoted to developing cancer drugs, managed to replicate successfully just 6 out of 53 drug studies in cancer research. Bayer managed to replicate successfully only 15 out of 67 studies. So, at least the correctness of 80% of the results published is deeply debatable. This, again, can be due to a practical problem (arguably, by repeating the studies, it could be determined which result was correct).¹ But Collins suggests that replication is doomed to failure as a way of settling disputes concerning experimental evidence because of an in principle reason. I will consider his argument in the next section.

2. The Experimenters’ Regress

In this paper I will be interested in the following: in cases of disagreement in biomedical research, how do we reach consensus? And when we reach consensus,

¹ And furthermore, these examples would show that many of the treatments that are currently on the market are not as efficacious as it was thought. In fact, many of them could even do more harm than the absence of treatment. (Cf. Lexchin, J. 2012).
are we scientifically justified in doing so? Which elements help researchers to reach consensus or to overcome disagreement or, if consensus among researcher is not reached, what helps the scientific community to decide that one experiment was correct while another was not?

This is where Harry Collins’s argument comes into the picture. The British sociologist provides, after doing extensive field work, an argument according to which the overcoming of an experimental disagreement cannot be achieved by providing scientific reasons, or it can only be so in the absence of committed critics. Accordingly, he gives an explanation of how disagreement is overcome by appealing to external means. Collins’ strategy of argumentation consists, on the one hand, in showing that the only epistemic admissible reproduction method for judging the correctness of an experimental result is the replication of the original experiment that lead to the discovery; on the other, it points out that differences with respect to the results obtained can be attributed to an unsuccessful replication rather than to having obtained an incorrect result. The consequence is that checking the adequacy of the replication would lead to an infinite regress. Since, however, it is an empirical fact that science has a way out of this regress, an explanation of how this is achieved is required. The explanation that Collins provides relies on the role of external factors in reaching consensus and settling the controversy. Given the alleged occurrence of the experimenter’s regress in experimental activity, Collins claims, it would be normal to find scientists arguing endlessly against each other about the quality of their findings. But, at the same time, from the fact that scientists eventually reach agreement about which result is correct and which experiment was properly performed, Collins concludes that scientific resources alone are not sufficient to establish the correctness of an experimental result, and that external, sociological factors, are required in order to break the circle. I take it that the challenge with which Collins confronts us is to offer an alternative and epistemic explanation of how disputes regarding discordant results can be overcome.

There are indeed good reasons to believe that many of the problems regarding the reproduction of results are due to the fact that the trials they arise from do not obey – if I may say - the “logic of science” but instead, the “logic of the market”. For example, Dr. Lexchin, has shown extensively how the pharmaceutical industry manages to present results in a way such as to maximize its profits, regardless of the efficacy of the drugs in question and despite their adverse effects, even omitting data that could dramatically modify the conclusions of the studies. (Cf. Lexchin, J. 2012). Bearing these problems in mind, I will circumscribe this paper to bona fide biomedical trials.

By external factors, Collins understands non-scientific reasons. The decision that one result is correct rather than another has to do with, according to him, the persuasive skills of the actors involved, their influence and renown in the scientific community, etc., but at all not with scientific reasons. (Cf. Collins, 1985, cps. 2 and 6).
explanation for the case of biomedical research will be the main purpose of my paper. Let us reconstruct Collins’ argument:

- Replication requires tacit knowledge transference. (Cf. Collins, 1992, pp. 73-74).
- The evaluation of the proper functioning of the experimental device and the experimenter’s expertise are determined by their ability to produce the correct experimental result. (Cf. Collins, 1992, p. 75).
- The only way to determine whether an experiment is being properly performed is by obtaining a correct result; but what the correct result is can only be known if the experiment is properly performed. (Cf. Collins, p.129).

The conclusion Collins extracts from the above premises is the following:

- Replication cannot settle an experimental disagreement.

But, as a matter of fact, even in the face of disagreement between research groups, eventually scientists reach consensus. Given that this is the case, if there were a regress, it would be necessary to explain how it is overcome. The explanation that Collins provides is that the way out from the experimenters’ regress is achieved by means of applying extra-scientific strategies. He assumes the following, without further argument:

- To scientifically resolve an experimental disagreement is to experimentally resolve it.

And he then concludes the following:

- (C1) Non-scientific tactics are responsible for overcoming an experimental disagreement.

Jumping from the difficulties of determining what would count as a good replication, to the impossibility of overcoming disagreement by means of scientific resources, he concludes that the determination of an experimental result cannot be achieved via scientific means. His conclusion is expressed clearly in the following non-sequitur: Some non-scientific tactics must be employed because the resources of the experiment alone are insufficient. (Collins, 1992, p. 143).
Let’s now summarize my reading of Collins’ dialectics: disconfirming an experimental claim requires replication, but since replication requires the successful transference of tacit knowledge, and judging what counts as a successful replication requires knowing in advance which the correct result is, a discordant result can always be contested by pointing out that the replication was not achieved. As a consequence, the dispute would be endless (therefore the regress), but in reality it actually comes to an end (hence, the need for an explanation of the way out). And, for Collins, it is the impact of social, economic and political considerations, and not scientific reasons, which stops the regress. I would like to offer an alternative explanation. In order to do so, in the next section I will consider David Teira’s contribution to the discussion and what he named the contractarian solution to the experimenters’ regress (2013). I will supplement Teira’s consideration with my own, so as to propose an alternative way out of the regress that relies only on scientific reasons.

3. Debiasing procedures and theoretical reasons

In his paper, A Contractarian Solution to the Experimenters’ Regress, David Teira addresses the problem of the experimenters’ regress in biomedical research, in particular, in drug testing. Teira proposes a contractarian way out of the regress, in which the involved parties agree on the application of a set of debiasing procedures, so as to guarantee that, even if their experiments are biased in some way, they will not be biased in such a way as to favour either of their respective hypotheses. There are two ways in which to read Teira’s paper: (i) as justifying the relevance of randomization, even if this cannot always avoid unbalanced distributions, or (ii) as an answer to the experimenters’ regress. Under the first interpretation, Teira would be arguing against Peter Urbach (1985) and John Worrall (2007), amongst others. If that were the case, he might have a point; but then it would be misleading to present his arguments, as he does, as an answer to the experimenters’ regress. If we take the second interpretation, however, I think his answer fails to provide a “solution”; for it would count as a solution to the regress only if the only sources of error that can lead to disagreement between researchers were biases, which is not the case. I will opt for the weakest reading of his proposal, hence attributing him just (i); and taking account of the contractarian approach he suggests, I will claim that agreement in the

---

4 I take this to be a minimal and reasonable reading of the problem Collins is trying to present. It is also what Collins (1984) himself claims in his discussion of Howson and Franklin’s (1984) paper.
use of debiasing procedures is a necessary condition for overcoming in an internal and scientific way the experimenters’ regress. (Nevertheless, I will not commit to randomization being necessary, just the application of a set of debiasing procedures). I will also suggest another condition, which together with Teira’s condition, would be conjointly sufficient to internally justify the acceptance of an experimental result in conflictive cases. Let us now explore some of the peculiarities of biomedical research.

In this kind of study, the experimenter usually postulates a causal claim in order to test it. There are several theoretical and practical difficulties associated with clinical trials; I will mention just a few of them. These experiments are sensitive to several types of bias which operate mainly unconsciously and can dramatically alter the results obtained, causing, for instance, the obtaining of positive results even when administrating a physiologically inactive substance. There is, in addition, a lack of unifying theories to guide the research. The commercial interests associated with the research generate, moreover, conflicts of interest and what can be thought of as a tug of war between pharmaceutical companies, who usually fund most of this research, on the one hand, and scientific imperatives, on the other (Cf. Lexchin, 2012). As has been mentioned before, however, here I will be concerned only with bona fide research.

Given the impact of bias in clinical trials, Teira argues that consensus among researchers would be easier to reach if debiasing methods were used. The contractarian approach rests on the fact that the more reasonable strategy for researchers to follow is to agree on accepting some methodological rules that establish what would count as a proper and well conducted experiment. The prerequisite of these rules is that they are impartial; they have to give each hypothesis the same probability of being correct. (Teira, 2013, pp. 714-716). Notice that while the debiasing strategy of the contractarian approach deals with the methodology of experimental practice, it is silent regarding the evaluation of the outcomes of an experiment. However, Collins’ contention is concerned precisely with how to determine the correctness of an experimental result. Despite this, Teira seems to believe that debiasing procedures generate consensus about experimental outcomes. Thus, for example, he says:

5 In this sense, he would be claiming that debiasing methods are subject to the experimenters’ regress, and even if it is not possible to judge in an objective way whether or not a debiasing procedure has worked, and even if it might be the case that applying debiasing procedures can still yield a biased result, their use is still justified, since at least they prevent the data being biased so as to favour each researcher’s hypothesis.
The contractarian solution to the experimenters’ regress is to implement debiasing procedures that make sure that the experimenter is impartial, even if the outcome sometimes is not. In contexts in which no objectivist alternative is available, we do not need more than a precommitment to these procedures to make an experimental result epistemically acceptable. (2013, p. 720, emphasis added).

This passage leads me to think that Teira may be making the more robust claim (ii). If that is so, then I disagree with him in this respect. I consider that the use of debiasing procedures cannot, on its own, make an experimental result epistemically acceptable. That would suffice if biases were the only sources of error, but this is not the case. That is why, even granting the use of debiasing procedures as a necessary condition for the epistemic acceptability of an experimental result - and hence, as a necessary condition for an internal explanation regarding how an experimental controversy is solved - there will still be room for disagreement regarding the quality of an experiment. I take it that a proper epistemic answer to how the experimenters’ regress is overcome is one that provides an independent but scientific criterion to determine what could count as correct result. On the one hand, by definition, a biased result cannot be considered a correct one. So a necessary condition for obtaining a correct result is to apply any methodologically appropriate procedure to prevent bias. The use of debiasing procedures, then, could indeed help to make it easier to reach consensus amongst researchers. But, given that there are several other causes of discordant results, this does not suffice as an explanation of the way out of the regress by internal considerations: the regress could persist if the researchers argue that factors other than biases have altered the other group’s results. For instance, the failure of tacit knowledge transference could account for the discrepancy between results and this is something that cannot be taken care of just by agreeing on the use of a set of debiasing procedures. My stance on this discussion is that (4) is the premise that should be resisted, and it seems that this cannot be done merely by applying debiasing procedures. I will therefore suggest a second condition that helps us to deny the truth of (4):

*Theoretical Calibration Principle:*

- (TCP) To determine the plausibility of the correctness of an experimental claim, seek for compatibility with independently known mechanisms that can explain the result.

Let me now show how the application of these two principles may provide an epistemic explanation of the way out of the experimenters’ regress. I will do so by
briefly considering an episode that Collins and Pinch described (Cf. 2005, cp. 4), and to which Teira himself refers. It concerns the testing of vitamin C and its capacity to control the growth of cancerous cells. A Scottish physician, Edward Cameron, wanted to test the hypothesis put forward by Linus Pauling, and in order to do so, developed a clinical trial, which according to Cameron’s reports, yielded positive results. The trial was soon replicated in the Sloan-Kettering Cancer Institute, but with opposite results, and later a double blind clinical trial ran at the Mayo Clinic by Dr. Charles Moertel, again, with negative results. Both parties defended the correctness of their own results and accused each other of committing selection bias. Cameron argued that, unlike his own patients, the patients Moertel selected were treated with chemotherapy and underwent radiotherapy. In response, Moertel performed a second trial, this time with patients who received neither chemotherapy nor radiotherapy and he got, once again, non-significant results. After this, Moertel accused Cameron of committing case selection bias. Cameron and Pauling then showed how the protocols in the different trials diverged and how this fact can explain the discordant results obtained. In spite of this, however, the scientific community judged that it was Moertel who had obtained the correct results.

Given this, we could claim that the failure to apply debiasing procedures can explain the disagreement, and hence, that applying Teira’s proposal would have sufficed to reach consensus by appealing to internal/scientific resources, hence overcoming the regress by epistemic resources. This, however, is not the case. For, in this episode, selection bias was not the only reason the researchers provided to account for the discordant results, thus the contractarian proposal cannot provide a complete, internalist explanation of how the decision regarding which was the correct experiment was reached. In fact, there were several differences in the protocols implemented that could account for the differences in the results. For instance, while Cameron’s patients were hospitalized, one of the replications in Mayo Clinic tested vitamin C with ambulatory patients. There was also evidence that the Mayo Clinic patients were taking vitamin C of their own volition, because, although they did not – and could not – know whether they were part of the control group, they wanted to be sure that they were taking the active principle just in case 1) they were indeed part of the control group and 2) that the active principle worked. This clearly invalidates the results of the study as there is no longer a control group and the dosage of the drug is no longer the one that the researchers stated in the protocol.

On the other hand, Pauling and Cameron lacked a solid mechanistic story about how the vitamin C could alter the growth of cancer cells. I believe that it is this
last fact that was decisive in the dispute and settled the issue in favour of the orthodoxy, represented by the researchers in the Mayo Clinic and by the hypothesis according to which vitamin C has no effect in the growth of cancer cells. If that is the case, then, we have an alternative answer to the experimenters’ regress. This does not need to appeal to any social, economical or political factor to explain the overcoming of the regress: there can be scientific reasons (even if it is not purely experimental) to break out of it. If the protocols are identical, then the application of debiasing procedures guarantees the impartiality of the results and a mechanistic story provides an explanation of how the treatment works. Together they jointly constitute sufficient conditions to epistemically justify an experimental result and hence, to explain internally and scientifically the way out of the regress.

Let me summarize the claims in this section: experimental evidence can be obscured by bias and by experimental errors; bias can be neutralized by applying debiasing methods while errors can be detected by testing the results against what is expected to happen given the best mechanistic knowledge we posses.

4. Concluding Remarks

In this paper, I have analysed Collins’ experimenters’ regress and offered an alternative explanation of how disagreements regarding experimental results can be settled, one which appeals to scientific resources. Even if we can never be absolutely certain about the correctness of an experimental result, this does not mean that we are not epistemically justified in believing it to be correct and therefore, making a scientifically informed decision in cases of discordant results arising from experimental replications. That this is so arises from the following considerations. To begin with, as David Teira claims, the use of debiasing procedures (in _bona fide_ research) guarantees that experimental results are neutral with respect to the interests of the actors. Moreover, in the absence of systematic theories, mechanistic reasoning is a relevant tool for deciding whether or not a result can be correct, hence closing the debate on the basis of scientific reasons.

If my claims are correct, there are two corollaries to be drawn. First, concerning biomedical research, we will only have to deal with the practical aspects of

---

6 This remark is quite important. For I also think that in this episode, Pauling and Cameron were right in claiming that the differences in the implementation of the protocols could have accounted for the differences in the results Moertel obtained.
implementing health policies that can properly regulate clinical trials and how and which data is presented to the scientific community for its evaluation. In other words, we can apply policies that can prevent non-bona fide research being considered as a candidate for scientific knowledge. Second, my analysis of the way out of the experimenters' regress may also shed some light on the debate concerning the golden standard of evidence in the context of evidence-based medicine. For according to the classic grading of evidence, randomized clinical trials are the gold standard, superior to observational studies and mechanistic reasoning (Cf. Howick, 2011, cp. 3). However, if what I claim is accurate, then mechanistic reasoning plays a crucial part in obtaining reliable evidence in clinical trials and it follows that even if mechanistic reasoning alone cannot justify the implementation of a particular treatment, its status should be reconsidered.

5. Acknowledgements

"I would like to thank the members of the LOGOS GRG for discussing with me some preliminary theses that are worked out in this paper. Also the attendees of the First LOGOS-EXRE-EIDOS workshop held in Fribourg, specially to Maria Scarpati, Aurelien Darbellay and Ljubomir Stevanovic where I discussed a previous version of this paper. Finally, I would like to thank the attendees of the BPPA Conference, specially to my respondent, James Fraser and to the editors of this volume: Elizabeth Watkins and Nahuel Sznajderhaus."

Romina Zuppina
rzuppone@unq.edu.ar

REFERENCES

Collins, H. and Pinch, T. (2005), Dr. Golem, How to Think about Medicine, Chicago, The University of Chicago Press.
