A Contractarian Solution to the Experimenter's Regress

David Teira*†

Debiasing procedures are experimental methods aimed at correcting errors arising from the cognitive biases of the experimenter. We discuss two of these methods, the predesignation rule and randomization, showing to what extent they are open to the experimenter's regress: there is no metarule to prove that, after implementing the procedure, the experimental data are actually free from biases. We claim that, from a contractarian perspective, these procedures are nonetheless defensible since they provide a warrant of the impartiality of the experiment: we only need proof that the result has not been intentionally manipulated for prima facie acceptance.

1. Debiasing Procedures and the Experimenter's Regress. The epistemology of experimental error often draws on repertoires of error-correction methods used across different disciplines (e.g., Franklin 2002, 6). In this article, we discuss a class of methods used in different experimental settings in order to correct errors arising from the experimenters' cognitive biases. We name them *debiasing procedures*.

The so-called confirmation bias provides an instance of the sort of error these procedures aim at correcting: this bias refers to unwitting selectivity in the acquisition and use of evidence; that is, the agent selects evidence in order to confirm one particular belief or hypothesis, although she does not intend to do it in a biased way and does not realize that she is doing so when it happens. Confirmation biases have been well documented by psychologists in a diversity of contexts (Nickerson 1998), including scientific research. Kevin Dunbar and his team, for instance, have shown how lab-

*To contact the author, please write to: Departamento de Lógica, Historia y Filosofía de la ciencia, UNED, Paseo de Senda del rey 7, 28040 Madrid, Spain; e-mail: dteira@fsof.uned.es.

†This article has been funded by the Spanish Ministry of Science research grant FFI2011-28835. María Jiménez-Buedo and Jesús Zamora made very valuable suggestions at every stage in the development of this article.

Philosophy of Science, 80 (December 2013) pp. 709–720. 0031-8248/2013/8005-0007$10.00
Copyright 2013 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/673717 Published online by Cambridge University Press

oratory scientists actually use "known standard" control conditions with a view to correct for confirmation biases: these controls involve "performing the experimental technique on materials where the expected result is already well known; if the expected result is obtained, the scientist can have confidence that the procedure is working" (Baker and Dunbar 2000, 345). This way, researchers do not trust their feelings about the correction of data, where confirmation biases may arise, but make a decision on the basis of an external check. These control conditions constitute an instance of what we call a debiasing procedure.

In general, debiasing procedures are experimental methods aiming at preventing the contamination of the experimental result by the preferences of the experimenter. In this article, we discuss in some detail two of them: the so-called predesignation rule and randomization. The former prescribes that tests should be designed before experimental data are obtained, so that the experimenter does not choose the most convenient test for his own interests. The latter is an allocation procedure used in experiments in which two treatments are compared, preventing any systematic connection between the preferences of the experimenter and the treatment assigned to each participant.

In this article, we discuss debiasing procedures in connection with a well-known problem in the epistemology of experiments, the so-called experimenter's regress (Collins 1981). In the case of debiasing procedures, this regress would be the impossibility of establishing in a mechanical fashion whether the procedure has actually worked, whether the experimental result is actually free from biases. Hence, we would ultimately depend on the experimenter's expert (but subjective) judgment to decide whether the data are actually unbiased. But how do we know whether this judgment is in itself bias-free? The experimenter's regress thus puts in question the epistemic justification of experimental results. Scientists would accept these results on less objective grounds than philosophers tend to think.

In section 2, we discuss the possibility of an objectivist solution to the experimenter's regress on debiasing procedures, drawing on the error-statistical analysis of the predesignation rule developed by Kent Staley and Aris Spanos. We show that there is indeed a potential experimenter's regress in our debiasing procedures because, as of today, we ignore the causal mechanisms by which biases are generated. We try to correct them, so to speak, blindly. Given this uncertainty, in section 3, we argue for a contractarian justification of our debiasing procedures: scientists would be justified in accepting a debiased experimental result not because we are certain about its efficacy in each particular case but rather because the implementation of such procedures guarantees the impartiality of the experiment. That is, they prevent any intentional manipulation of the data, so that, in the long run, as experiments are replicated, they make it more likely that the true outcome will be observed, independently of the biases of the experimenters involved.

https://doi.org/10.1086/673717 Published online by Cambridge University Press

The role of methodological norms as warrants of impartiality has been mostly emphasized by the contractarian approach, and in this article we want to present a formal argument showing how this approach overcomes the experimenter's regress in the case of debiasing procedures such as randomization (sec. 4). We claim that, from an epistemic perspective, experimenters only need to agree on which methods can impartially control their biases in order to justify any experimental result they obtain, even if the experimenters cannot establish in an objective manner in each case whether their results are actually bias free.

2. An Objectivist Solution to the Experimenter's Regress. An objectivist solution for the experimenter's regress would be one in which we would check for the biases of our data in a purely procedural fashion, without room for subjective judgment. The most promising approach in this respect today is the error-statistical philosophy of evidence developed by Deborah Mayo (1996). Instead of just drawing on a "hodgepodge" of methods aimed at correcting experimental errors, Mayo wants to formally establish the probability that a given test detects an error when it actually occurs. Within this approach, there are two alternative views as to the justification of debiasing procedures, depending on whether we take them as warrants of the objectivity of our statistical model or we require instead built-in assumptions on the effects of experimental biases.

Kent Staley (2002, 2004) provides an illustration of the more "moderate" approach in his discussion of the so-called predesignation rule in physics experiments. According to this debiasing procedure, all the relevant features of a test procedure should be predesignated in advance of any examination of the data. This is a rule of folk statistics (Mayo), generally accepted in physics without a clearly argued rationale. According to Staley, the underlying intuition is that our testing procedures can make a difference to the sampling distribution of the experiment (2002, 282), impinging on our assessment of the severity of the test. But, as the following experiment illustrates, physicists sometimes violate this rule.

In the 1990s, an experiment was conducted in the Collider Detector at Fermilab, with a view to establishing the existence of the top quark. Millions of particle collisions were measured, with a view to isolating "candidate events" revealing the sought-after top quark. After collecting the data, the threshold values of the measurements defining a candidate event (the so-called cuts) were changed. The results were nonetheless published (not without discussion among the authors) and accepted.

From an error-statistical perspective, this is clearly an objectionable procedure: when we test a hypothesis with a given significance level, we assume that our experimental data are just an instance of a series of infinite repetitions of the experiment as it has been originally designed. As Aris Spanos (2010, 567) puts it, we "embed the material experiment into a sta-

https://doi.org/10.1086/673717 Published online by Cambridge University Press

tistical model viewed as a purely probabilistic construct." The statistical distribution captured in this model $M_\theta(x)$ formally defines a legitimate event in the experiment as a typical realization of $M_\theta(x)$. A test of $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta \geq \theta_1$ in a Neyman-Pearson framework is just an analysis of the true distribution of the sample $f_*(x)$ according to the experimental data: were these data generated by $M_0(x)$ or rather by its complement $M_1(x)$ relative to $M_\theta(x)$?[1]

In the Fermilab experiment, the significance of the obtained data was not calculated taking into account that the cuts defining a candidate event had been redefined in order to increase the value of the test statistic. The assumption was that the sampling distribution $M_\theta(x)$ constructed for the original threshold value would not change with the new cut.

Staley (2002, 288) captures the statistical rationale of the predesignation rule through the concept of the *homogeneity* of a reference class: "A reference class A used in calculating the probability of an outcome E is homogenous with respect to E only if there is no factor B, under the control of the experimenter and present in that instance of the experiment that resulted in E, such that $p(E|A) \neq p(E|A\&B)$." In the Fermilab experiment, the reference class of candidates events A is defined for the original threshold through $M_\theta(x)$, and the factor B is the intervention of the experimenter to redefine the threshold, after seeing the experimental data E. The assumption made by the experimenters with the new cuts is precisely that $p(E|A) \neq p(E|A\&B)$.

The predesignation rule, so conceived, is clearly a debiasing procedure: it aims at preventing the contamination of the experimental data by the subjective preferences of the experimenter. In the Fermilab experiment, we want to make sure that the threshold is not defined in a self-serving manner according to the preferences of the experimenter regarding the existence of the top quark. If the experimenter believes in the top quark, she should not set the threshold for detecting it after seeing the data because she may recognize its existence "too easily" (if she falls prey to a confirmation bias). Staley claims that the predesignation rule can be justified, in an error-statistical perspective, because it is generally unfeasible to calculate a probability distribution for tests in which the experimenter's preferences intervene. We cannot estimate the effect of these preferences on the outcome. We can only choose whether to believe the experimenter if she claims that her intervention did not have any effect in the outcome. Hence, we justify the predesignation rule as a means to protect the objectivity of M_θ—more precisely, the homogeneity of the reference class.

1. For a given value of the parameter a, $M_0(x) = \{f(x;\theta),\ \theta \leq a\}$, $M_1(x) = \{f(x;\theta),\ \theta > a\}$.

https://doi.org/10.1086/673717 Published online by Cambridge University Press

Aris Spanos provides a more radical approach to debiasing rules, in which Staley's justification turns out to be not objective enough. According to Spanos, biases affect evidence "through the distortion of the relevant error probabilities" stemming from the data-generating mechanism captured in $M_\theta(x)$.[2] Evidence in an error-statistical perspective should be objective: we should control the generation of the data by checking the assumptions of $M_\theta(x)$. Debiasing procedures such as the predesignation rule are neither necessary nor sufficient for getting reliable data. As the Fermilab experiment could illustrate, if the existence of the top quark is confirmed through other means, we can get reliable data without applying the rule. And even if we apply the predesignation rule, the data might have been contaminated by the preferences of the experimenter in a different way we were not controlling for. Evidence in an error-statistical perspective should be objective: we should control the generation of the data through the model $M_\theta(x)$ that purportedly captures the data-generating mechanism, rather than through any debiasing procedure, whose implementation is necessarily subjective.

Were we able to identify the causal mechanisms generating biases in our data, the best conceivable solution would be indeed to incorporate such knowledge into our model in order to verify whether our results are actually bias free. The main problem here is that psychology is still far from giving a definite answer to the causal mechanisms underlying biases: even in the best studied cases, such as the confirmation bias mentioned in the introduction, we seem to have an umbrella concept covering a family of experiments dealing with selectivity in the use of evidence with many potential accounts of its sources. This is why Staley's justification of the predesignation rule as a means to protect the objectivity of our models of the experiment seems to us more cogent: given the uncertainty about the way in which the experimenter's preferences may affect the impartiality of the tests, it seems wise to sever all ties between both (which is what the predesignation rule does).

However, Staley's approach does not solve the experimenter's regress as such, and in this respect Spanos is right: we can only be certain of the homogeneity of the reference class to the extent that there is no factor under the control of the experimenter that may change the statistical distribution assumed in the model. But how do we know that no such factor has been manipulated in each particular set of results we obtain? For instance, if a scientist finds one of the results to be against his own interests (e.g., in seeing his own hypothesis succeed), he may always contest the result, claiming that the homogeneity of the reference class has been violated. In our view, the epistemic justification of any debiasing procedure should provide an answer to this last problem. Our agreement on these pro-

2. Personal communication, February 21, 2011.

https://doi.org/10.1086/673717 Published online by Cambridge University Press

cedures should be strong enough to resist any self-serving use of the experimenter's regress in order to promote the particular views of any scientist. The contractarian approach that we present in the next section provides, we believe, the best justification of such an agreement.

3. A Contractarian Approach to a Debiasing Procedure. Let us introduce the contractarian approach through an analysis of a much debated debiasing procedure: randomization, particularly as it is used in clinical trials in medicine. In these trials, once a patient is deemed eligible (according to the trial's protocol) and recruited, a treatment is assigned at random. Depending on the arrangement of the trial (number of treatments, whether it is double-blind, whether it is multicenter), randomization may be implemented in different ways. The general principle is that each patient should have an equal probability of receiving each treatment. In cases in which it is convenient to control the allocation of treatments according to patient characteristics, in order to prevent imbalances, randomization can be stratified.

From a purely statistical standpoint, randomization is justified for the contribution it makes to significance testing (e.g., Basu 1980). But, independently of this, experimenters usually adopt randomization as an allocation procedure in order to avoid selection biases: it prevents investigators from assigning (consciously or unconsciously) patients with, say, a given prognosis to any one of the treatments. For instance, an investigator might allocate the experimental treatment to the healthier patients, if she wants the trial to be positive, or to the patients with a worse prognosis, if she thinks they will benefit more. This is an argument that never fails to appear in medical textbooks.

We should distinguish here between the procedure and the outcome of randomization: procedurally, it is a warrant of experimental fairness in the sense of preventing the manipulation of the treatment allocation; however, it might happen that the allocation obtained is nonetheless biased. By sheer chance, a random allocation may yield an unbalanced distribution of the two treatments; that is, the test groups may differ substantially in their relevant prognostic factors (these are called *baseline imbalances*). This difference may bias the comparison between treatments and spoil the experiment. If one such distribution is observed, the customary solution is to randomize again seeking a more balanced allocation. Ultimately, the judge of this balance is, of course, the experimenter. This is a well-known source of Bayesian objections against randomization as grounds for significance testing.[3]

3. For example, Urbach (1985) argues that the methodology of significance testing forbids any choice between random allocations: if they are adequately generated, one allocation should be as good as any other. Hypotheses should be accepted or rejected on the basis of the experiment alone, without incorporating our personal assessment of the generated data.

https://doi.org/10.1086/673717 Published online by Cambridge University Press

It is not so often noticed that it also constitutes a clear instance of the experimenter's regress, this time about debiasing procedures: even if randomization is procedurally fair, its outcome can only be declared unbiased according to the skills of each particular experimenter, without an external standard to confirm it.

Harry Collins and Trevor Pinch (2005, 84–110) illustrated this regress (with another form of selection bias) in their analysis of the trials of vitamin C as a cure for cancer. The therapeutic use of vitamin C was originally proposed by Linus Pauling, following his conjecture on the failure of the cellular mechanisms as the ultimate cause of a number of diseases (the common cold, mental deterioration associated with age, etc.). A Scottish physician, Ewan Cameron, developed this conjecture and started experimenting with cancerous patients in the early 1970s with apparently positive results in a study with matched historical controls. After a few years of controversy, Cameron's study was replicated at a bigger scale at the Mayo Clinic in a randomized trial with negative results. Interestingly, both parties accused each other of improperly debiasing their studies: for the Mayo researchers, there has been a selection bias in Cameron's original study, by which only patients with less severe forms of cancer had been treated; for Cameron and Pauling, in turn, there was evidence that the patients at the Mayo study had violated the trial protocol, ingesting more vitamin C than the prescribed dose. Accusations of scientific fraud flew. All parties claimed to have the proper skills to test the therapy, but these claims could not be tested experimentally. For the Mayo team, the controls on patients (random urine analysis of their vitamin C levels) had been good enough because the dosage of the treatment was not as important as for the Cameron team. Conversely, the Mayo team had failed to select the type of patients who could benefit most from the therapy. The controversy never came to a formal end (Jenkins 2004, 14–35).

The experimenter's regress is one of the strongest arguments for the sociologists' claim that experiments are closed not on the basis of purely epistemic considerations but rather on a nonepistemic bargaining between self-interested agents. Experiments alone do not settle scientific controversies once and for all. Perhaps they would if there were unlimited resources for research and we could exhaust all the possible sources of controversy with further experiments. But since there are no such resources in our world, experimental evidence α will be inescapably conditioned by the sort of nonobjective considerations that most philosophers of science want to avoid: α will be evidence for an agent X, only because of the influence of some nonepistemic factors Y.

If we want to overcome the experimenter's regress, we need to justify randomization on a procedural basis, not just on the grounds of its outcome. This might seem somewhat paradoxical since we usually expect scientific

https://doi.org/10.1086/673717 Published online by Cambridge University Press

procedures to be acceptable on the basis of their results alone. If randomization did not yield unbiased results often enough, there would be no reason to allocate treatments at random. The problem that the experimenter's regress points out is that we do not have a clear criterion to distinguish in particular cases when a randomized allocation is actually free from biases. Hence, we would lack an objective reason (in the sense of Spanos) to accept its outcome.

In our view, there is an objective epistemic reason to accept a randomized allocation, despite the uncertainty about its particular outcomes. We need to explain under which circumstances randomization constitutes a better allocation rule than any other alternative procedure for assigning treatments in a comparative experiment. The problem that the experimenter's regress opens is that the uncertainty about the fairness of a randomized outcome can be exploited for purely private purposes if the decision depends on the tacit skills of the experimenter alone. It might happen that we accept as legitimate an experimental result whose methodological grounds are shaky just because the outcome favors the interest of the experimenter. In our view, the virtue of randomization is that even if the outcome is occasionally biased, this will never happen in the interest of the experimenter. Further replications might show that an experimental result was erroneous, but this will happen precisely because debiasing procedures such as randomization prevent any systematic intervention to alter the result according to anyone's particular interests.

This intuition can be substantiated from the standpoint of the *epistemic contractualism* promoted by Jesús Zamora (2002, 2006; Ferreira and Zamora 2006). Zamora's major claim is that even a community of self-interested scientists would have incentives to adopt good methodological standards. A researcher seeking to increase her professional reputation can only succeed if her peers accept her results. So even if every one of them has an incentive to promote the methodological rule most favorable to her own theory, they need to agree on a set of standards. Chances are that, if this agreement takes place before their actual research starts, they will choose standards a priori neutral regarding their particular interests. We may deem these rules impartial to the extent that they warrant "a satisfactory gain to anyone in a wide set of conceivable circumstances" (Zamora 2002, 305); that is, impartiality does not imply absolute disinterestedness: scientists are allowed to pursue their own private goals, provided that they play by the methodological rules they have agreed. Indeed, the normative force of these rules stems exclusively from their self-interested agreement: they are not universally valid principles, even if we may expect that the negotiation captures epistemic intuitions that can pass as such.

Whereas for Popper it was merely a conventional decision which statements we take as basic evidence for testing a theory, in Zamora's contrac-

https://doi.org/10.1086/673717 Published online by Cambridge University Press

tarian approach we can justify this decision in terms of agreed evidence-gathering norms. Scientists need to decide in advance what counts as legitimate evidence in order to avoid the "temptation" of contesting someone else's discovery (e.g., for lack of data or significance) in order to maximize their own chance of making it themselves. In this spirit, we can interpret a formal argument developed by Berry and Kadane (1997) in order to show, from a utilitarian perspective, why two self-interested scientists would agree on randomization as an allocation procedure, independently of their statistical creed.

We have a situation in which scientist A will test a given hypothesis about the efficacy of a drug on the data gathered in a clinical trial by scientist B. Scientist A does not know whether it is better to trust the individual judgment of B in the allocation of treatments to patients or rather to use some "mechanically objective" device, such as randomization. In the scenario of scientific competition assumed in the contractarian approach, scientist A will want to prevent that B manipulates the allocation, consciously or unconsciously, biasing the results in his own benefit. And so would B, if the situation were reversed. This is why they would agree to randomize the allocation on a purely procedural basis: even if the result is unbalanced, they cannot contest the fairness of the experiment. In order to rationalize this intuition, let us present a more analytical picture of the situation, drawing on Berry and Kadane (1997).

We have n patients: n_1 receive treatment 1 and n_2 treatment 2. Let $X_i = 1$ if patient i is a success and $X_i = 0$ if patient i is a failure. Let t_i be the treatment that patient i received.

Suppose there is a covariate h_i representing the general health of the patient, where $h_i = 1$ if the patient i is healthy and α otherwise. Let p_{jk} be the probability that a patient is a success under treatment j with covariate h_i, where $k = 0, 1$. If someone has to assign treatments to patients and perceives that p_{j0} is different from p_{j1}, she may feel tempted—for whatever reason—to do the allocation on the basis of the covariate.

Let w be the proportion of the population of patients that is healthy. The goal of the study is the estimation of the probability that a patient is a success under a given treatment. Since the health of the patient may affect the treatment's outcome, let us represent the probability that a random selected patient from the population is a success under each treatment as follows:

$$p_1^* = wp_{11} + (1 - w)p_{10},$$
$$p_2^* = wp_{21} + (1 - w)p_{20}.$$

If we ignore the existence of the covariate h_i, the sufficient statistics for estimating p_1^* and p_2^* are

https://doi.org/10.1086/673717 Published online by Cambridge University Press

$$\hat{p}_1 = \frac{\sum_{i:t_i=1} X_i}{n_1}, \quad \hat{p}_2 = \frac{\sum_{i:t_i=2} X_i}{n_2}.$$

As $n_j \to \infty$, $\hat{p}_j \to P(X = 1 \mid t = j)$ for $j = 1, 2$.

But let us now calculate this last expression, $P(X = 1 \mid t = j)$, taking into account the possibility of allocating treatments according to the covariate. Let λ_1 be the probability that a healthy patient is allocated to treatment 1 and λ_0 the probability that an unhealthy patient is allocated to treatment 1. Therefore, the probability of a patient being a success conditional on receiving treatment 1 is now as follows:

$$\begin{aligned} &P(X_i = 1 \mid t_i = 1) \\ &= P(X_i = 1 \mid t_i = 1, h_i = 1)P(h_i = 1 \mid t_i = 1) \\ &+ P(X_i = 1 \mid t_i = 1, h_i = 0)P(h_i = 0 \mid t_i = 1). \end{aligned}$$

We can now calculate the probability of being healthy, conditional on receiving treatment 1:

$$\begin{aligned} &P(h_i = 1 \mid t_1 = 1) \\ &= \frac{P(t_i = 1 \mid h_i = 1)P(h_i = 1)}{P(t_i = 1 \mid h_i = 1)P(h_i = 1) + P(t_i = 1 \mid h_i = 0)P(h_i = 0)} \\ &= \frac{w\lambda_1}{w\lambda_1 + (1 - w)\lambda_0}. \end{aligned}$$

Therefore,

$$\begin{aligned} &P(X_i = 1 \mid t_i = 1) \\ &= p_{11}\frac{w\lambda_1}{w\lambda_1 + (1 - w)\lambda_0} + p_{10}\frac{(1 - w)\lambda_0}{w\lambda_1 + (1 - w)\lambda_0}. \end{aligned}$$

We can now compare the probability of a success with treatment 1 ignoring the covariate with the probability of success with the same treatment, taking into account the covariate:

$$\begin{aligned} &p_1^* - P(X_i = 1 \mid t_i = 1) \\ &= w(1 - w)(p_{11} - p_{10})\frac{\lambda_0 - \lambda_1}{w\lambda_1 + (1 - w)\lambda_0}. \end{aligned}$$

We see that $\hat{p}_1$ will only provide a correct estimation of p_1^* in the following three cases:

a) $w = 0$ or $w = 1$: the patients are all healthy or all unhealthy; there is no covariate.

https://doi.org/10.1086/673717 Published online by Cambridge University Press

b) $p_{11} = p_{10}$: the covariate does not affect the result of the treatment.
c) $\lambda_1 = \lambda_0$: healthy and unhealthy patients have the same probability of receiving treatment 1.

Let us now return to the initial situation: scientist A has to choose an allocation rule to be used in the trial conducted by B. If, for whatever reason, A has a positive probability that there is such a covariate (denying *a*) and that this covariate affects the probability of the treatment's success (denying *b*), A should ask for the treatments to be allocated with equal probability to each patient in order to prevent any manipulation; that is, the optimal allocation rule is randomization. With this, A maximizes his chances of getting a correct estimation of p_1^*. And so would B if the situation were reversed. In this sense, randomization would be a warrant of the impartiality of the trial, on which a procedural agreement (as an evidence-gathering norm) between competing scientists is possible.

4. Randomization and the Experimenter's Regress. As we mentioned before, the main epistemic threat of the experimenter's regress is that an expert makes a biased judgment over the legitimacy of a result. Indeed, no debiasing procedure can guarantee that its outcome is actually free from the biases it aimed at preventing. The question is thus why scientists with conflicting interests would agree on a purportedly debiased result. The contractarian approach argues that even the most self-interested researchers would have an incentive to agree on impartial procedures that prevent the manipulation of a result. As the case of randomization illustrates, these procedures provide the best chance for them of seeing the truth emerge with further replications of the experiment, without giving any of the competing parties an unfair advantage in proving his or her own hypothesis.

An objectivist like Spanos may object to this understanding of randomization that it is still subjective (just as the predesignation rule), on the grounds that it depends entirely on an agreement on an experimental procedure, rather than on an actual analysis of the subsequent data. Indeed, if we knew every potential source of bias, we would be able to arrange alternative allocation procedures conditionalizing on them, just as Kadane and Seidenfeld (1990) suggest, controlling the whole procedure by purely statistical means. But the problem in clinical trials is that we are uncertain about the sources of bias that may interfere in the allocation. The virtue of randomization is that it guarantees that this uncertainty is not exploited by any competing scientist in his own interest, giving everybody an equal chance to see his own hypothesis succeed and, crucially, making it more likely that this is the true one—as the experiment is replicated.

The contractarian solution to the experimenter's regress is to implement debiasing procedures that make sure that the experiment is impartial, even

https://doi.org/10.1086/673717 Published online by Cambridge University Press

if the outcome is sometimes 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.

REFERENCES

Baker, Lisa, and Kevin Dunbar. 2000. "Experimental Design Heuristics for Scientific Discovery: The Use of Baseline and Known Controls." *International Journal of Human Computer Studies* 53:335–49.

Basu, Debabrata. 1980. "Randomization Analysis of Experimental Data: The Fisher Randomization Test." *Journal of the American Statistical Association* 75:585–81.

Berry, Scott, and Joseph Kadane. 1997. "Optimal Bayesian Randomization." *Journal of the Royal Statistical Society* B 59:813–19.

Collins, Harry. 1981. "Son of Seven Sexes: The Social Destruction of a Physical Phenomenon." *Social Studies of Science* 11:33–62.

Collins, Harry, and Trevor Pinch. 2005. *Dr. Golem: How to Think about Medicine*. Chicago: University of Chicago Press.

Ferreira, José Luis, and Jesús Zamora. 2006. "An Economic Model of Scientific Rules." *Economics and Philosophy* 22:191–212.

Franklin, Allan. 2002. *Selectivity and Discord: Two Problems of Experiment*. Pittsburgh: University of Pittsburgh Press.

Jenkins, Stephen H. 2004. *How Science Works: Evaluating Evidence in Biology and Medicine*. Oxford: Oxford University Press.

Kadane, Joseph, and Teddy Seidenfeld. 1990. "Randomization in a Bayesian Perspective." *Journal of Statistical Planning and Inference* 35:329–45.

Mayo, Deborah G. 1996. *Error and the Growth of Experimental Knowledge*. Chicago: University of Chicago Press.

Nickerson, Raymond S. 1998. "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises." *Review of General Psychology* 2:175–220.

Spanos, Aris. 2010. "Is Frequentist Testing Vulnerable to the Base-Rate Fallacy?" *Philosophy of Science* 77:565–83.

Staley, Kent W. 2002. "What Experiment Did We Just Do? Counterfactual Error Statistics and Uncertainties about the Reference Class." *Philosophy of Science* 69:279–99.

———. 2004. *The Evidence for the Top Quark: Objectivity and Bias in Collaborative Experimentation*. Cambridge: Cambridge University Press.

Urbach, Peter. 1985. "Randomization and the Design of Experiments." *Philosophy of Science* 85: 256–73.

Zamora, Jesús. 2002. "Scientific Inference and the Pursuit of Fame: A Contractarian Approach." *Philosophy of Science* 69:300–323.

———. 2006. "Science Studies and the Theory of Games." *Perspectives on Science* 14 (4): 525–59.

https://doi.org/10.1086/673717 Published online by Cambridge University Press