Dealing with the Experimenter Effect

Dick J. Bierman

Jacob J. Jolij

University of Groningen, The Netherlands
dbierman@contact.uva.nl

Abstract—Methods in experimental science assume objective facts, and those effects are generally independent of the observer or experimenter. This objectivity assumption is not warranted in the field of human studies. Results of psychological experiments tend to be dependent on among other things the expectations of the experimenter. The experimenter effect togther with the replication crisis in social psychology are major issues in experimental parapsychology. We use Houtkooper's *Hierarchical Observational Theory* to look at a model for dealing with this issue, and conclude that multiple-experimenter projects might be able to sort out experimenter effects from intrinisic effects.

Keywords: experimenter effect; replication crisis; psi; parapsychology

"Are we shamans, not scientists?"

THE PROBLEM

The above quotes was the desperate response by Rex Stanford when he realized that experimenters in psi experiments cannot avoid being a participant, too, and hence only 'subjective' data could be obtained (Stanford, 1981).

Methods in experimental science, however, have been developed under the assumption that there are objective facts, i.e. that effects are generally independent of the observer or experimenter. Rosenthal

showed that this objectivity assumption is not warranted in the field of human studies. Results of psychological experiments tend to be dependent on among other things the expectations of the experimenter (Rosenthal, 1969). These experimenter effects (E-effects) were assumed to be caused by subtle influences of the experimenter on the system under study. Through automatization of experimental procedures, these subtle influences were assumed to be reduced. However, at times, some unexplainable effects depending on the experimenter were still observed when experimenters had little interaction with the experiments.

Mainstream psychology is presently struggling with these issues: The so-called replication crisis in social psychology is attributed by several authors to the idiosyncratic effects of context (including the experimenter) on the outcomes of subtle manipulations (see for example Doyen et al., 2012). This is of course a major issue for any science that relies on careful manipulation of independent variables in experimental settings.

Experimental Parapsychology

In the field of experimental parapsychology, the role of the experimenter has been a continuous source of discussion. Some researchers (Rabeyron, 2019) have even taken the position that the contribution of the experimenter is basically uncontrollable and that what we observe is nothing more than the hopes and expectations of a few (psi-gifted) experimenters. According to this position, further experimental research is a waste of time and the focus should be on spontaneous cases.

Experimenters in this field of research sometimes have a strong worldview at stake in contrast with the unselected participants in their experiments, and hence the idea that experimenters are the main source of the anomalous effects cannot be excluded.

We Learn Nothing Intrinsic

If indeed these psi results like several differential effects (role of belief in psi, role of brain state, etc.) are just the consequence of projections of the researcher (except when there is something that

does evade explanation), then one has to ask if we really can learn from psi experiments. The effects are then not intrinsic to the process but are manifestations of the personal hopes and expectations of the experimenter. Sheep do better than Goats? Wait for a (psi-gifted) experimenter who doesn't believe this and then we will get the reverse results (Bierman, 1981).

Assumptions Needed

Can we ever improve the experimental methods so as to deal with the experimenter effect? Let us be clear: If indeed "any observer or any person in some way related to the experiment can now or in the future have an impact without any constraint," then no research is possible. However, if that were the case then our experiments would have such a large variance on this uncontrolled psi that we would get only extreme results (with an unlimited number of psi sources "participating"). In fact, we do not see this (Houtkooper, 1977). We may therefore assume that the idea that "any observer or any person in some way related to the experiment can now or in the future have an impact *without any constraint*" is false. There must be a constraint.

MODELS

But what constraint? At this point it is imperative to introduce models. Without some model, we are unable to come up with methods that would help us to deal with E-effects. As an example, I use Houtkooper's *Hierarchical Observational Theory* (1983). He assumes that:

1. any observer of the results has (retroactive) psi input into the result

2. a second observer of the same dependent variable is contributing less, even if his psi strength is as large as that of the first observer (and so on). There is a hierarchy.

It follows first that in this Hierarchical Model the 'analyzers' are probably the persons with the largest impact. They are the first to see the final results. This is in line with analyzer effects reported in the literature (Feather & Brier, 1968; Weiner & Zingrone, 1989; West & Fisk,

1953). Subjects have impact only on the trial level (hits and misses) but not on a global level (other compound measures such as run scores and of course results over all subjects).

Often, though, the Experimenter is also the Analyzer—which in some ways simplifies the problem.

Variance!

The solution to the Experimenter effect under these theoretical assumptions is basically the same as for any other source of uncontrolled *variance*: Introduce the Experimenter as a factor (hopefully resulting in some explained variance) in the design.

Interestingly, recent developments in statistical modeling have made an initial test of this idea more straightforward. Linear mixed models (lmms) have rapidly become the de facto standard in experimental psychology, in particular based on developments in psycholinguistics. In this field of study, stimuli may have an idiosyncratic effect on the dependent variable, a situation now also recognized in social psychology (Judd et al., 2012). One can control for such idiosyncrasies by using a mixed model, a model that takes into account both fixed effects (effects of factors that are under the control of the experimenter and have a known or at least predicted effect on the outcome variables) and random effects (effects that are believed to be a source of variance but that have an unknown effect on the outcome variables).

For any study in which more than one experimenter has contributed, one may compare the model fits for a model including Experimenter as random factor versus a null model in which this term is omitted. If indeed there are significant experimenter effects, regardless of the actual psi (or anti-psi) effects of any individual experimenter, the model including the random term should give a better fit than the null model.

Therefore, rather than running one experiment with one Experimenter, projects should engage say 20 experimenters. Obviously, a formal power analysis would be preferred to compute the required number of experimenters, but given that we do not know the effect size of an eventual experimenter effect, we will have to start with an initial guess.

The method practically requires a coordinator to supervise such

a project. So aren't we just transferring the problem to a next level in the hierarchy? Shouldn't we then expect a Coordinator effect? Aren't we going to just measure the expectations and hopes of this coordinator without learning anything intrinsic about the psi process?

That this coordinator has psi input is a valid argument. However, due to the assumption about decreasing effect with the order of observers, we may assume this contribution to be *limited and smaller than that of the experimenters.*

Because we assume that observational theories and in particular the Hierarchical Model are valid, the observational history of the results should be very well-controlled. For instance, no data peeking is allowed by anyone, and the first observation of the results has to be shared simultaneously by all experimenters (for instance in an online meeting). Results are then later communicated to the coordinator after this shared analysis.

Project Approach

This approach has another advantage, namely that it may empirically establish a confidence interval by analyzing the distribution of the results of the experimenters.

Thus, the requirement for replication is easier to quantify: A result is replicated if it falls within the 99.9% confidence interval obtained by the distribution of the results of the 20 experimenters. A result is not replicated if it falls outside of that confidence interval.

Another argument that favors this project approach is that with many different contributing labs, the chance for the same systematic error explaining the result is smaller. Recently, a multi-experimenter project was reported (Schlitz et al., 2019). By using an ANOVA, they were able to conclude that there was no contribution from the experimenters. Because the main psi measures did not show psi, this result in this case is not surprising.

CONCLUSION

We have shown that by using Analysis of Variance or corresponding non-parametric techniques in *multi-experimenter* projects, we may be able to separate the experimenter effect from the intrinsic effects

that shed light on the psi process itself. For this approach to work, it is mandatory to assume some theoretical framework that at least introduces a constraint so that not any observer/experimenter now and in the future can have an unlimited impact on the data. Of course, that theoretical framework may be totally wrong. And therefore, the practical implementation may be incorrect. For instance, the requirement to have the analysis be shared by all experimenters could be unnecessary or even wrong when assuming a theoretical framework other than the *Hierarchical Oservational Theory*. Frameworks like von Lucadou's Model of Pragmatic Information (MPI) may require other practical implementations (von Lucadou, 1995). But, in any case, such a project should have many experimenters in order to assess the contribution of the experimenter to the final result.

REFERENCES

Bierman, D. J. (1981). *Negative reliability, the ignored rule in psi research*. Presented at the 23rd Parapsychology Association Convention, 1980, Reykjavik, Iceland.

Doyen, S., Klein, O., Pichon, C. L., & Cleeremans, A. (2012). Behavioral priming: It's all in the mind, but whose mind? *PLOS One, 7*(1), e29081. https://doi.org/10.1371/journal.pone.0029081

Feather, S., & Brier, R. (1968). The possible effect of the checker. *Journal of Parapsychology*, 32, 167–175.

Houtkooper, J. M. (1977). A comment on Schmidt's mathematical model of psi. *European Journal of Parapsychology, 2*(1), 15–18.

Houtkooper, (1983, December). *Observational theory: A research programme for paranormal phenomena.* Ph.D. Thesis, State University of Utrecht, pp. 57–64.

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology, 103*(1), 54–69. https://doi.org/10.1037/a0028347

Lucadou, W. v. (1995). The model of pragmatic information (MPI). *European Journal of Parapsychology, 11*, 58–75.

Rabeyron, T. (2019). Presentation at the pre-conference "Workshop on Psi Theory," July 2–3, at the 62nd Annual Convention of the Parapsychological Association, July 4–9, 2019, Paris.

Rosenthal, R. (1969). Interpersonal expectancies: The effects of the experimenter's hypothesis. In R. Rosenthal & R. Rosnow (Eds.), *Artifacts in behavioral research,* Academic Press.

Schlitz, M., et al. (2019). *Experimenter effect and replication in psi research I, II, and*

III: A global initiative. Presented at the 62nd Annual Convention of the Parapsychological Association, July 4–6, 2019, Paris.

Stanford, R. G. (1981). Are we shamans or scientists? *Journal of the American Society for Psychical Research, 75,* 61–70.

Weiner, D. H., & Zingrone, N. L. (1989). In the eye of the beholder: Further research on the "checker effect." *Journal of Parapsychology, 53*(3), 203–233.

West, D. J., & Fisk, G. W. (1953). A dual ESP experiment with clock cards. *Journal of the Society for Psychical Research, 37,* 185–189.