

Insufficient Effort Responding in Experimental Philosophy, Part I: Recommendations

THOMAS PÖLZLER

Department of Philosophy, University of Graz, Austria

Experimental philosophy studies often take the form of self-report surveys on lay persons' philosophical intuitions. One thing that puzzled me when I first contributed to such a study was how little cognitive effort some subjects seem to have taken in responding to it. For example, some completed the study in less than half the time that would have been necessary to even only read all materials; and a considerable proportion failed very simple attention and comprehension checks. I found that in psychology the phenomenon that I encountered has recently mostly been referred to as “insufficient effort responding” (IER), and that those who investigate it widely believe that it can have serious psychometric effects.

So what to do about IER in experimental philosophy studies? To answer this question I dug deeper into the available literature. In this blog post I sketch the results of my research. In particular, I provide recommendations on how experimental philosophers (and other researchers who use self-report surveys) best attempt to prevent, detect and deal with IER. These recommendations are rough and preliminary; not at least because psychological research on IER is itself still at an early stage. At least, however, the following considerations may raise awareness for this topic among experimental philosophers, and may provide a helpful basis for further discussion.

Preventing IER

It is advisable to try to prevent as much IER as possible before it even occurs. Initial evidence from psychology suggests that the following strategies may be at least somewhat effective in doing so.

Increase compensation in cross-cultural research. While higher payment has not been found to decrease IER in US subjects, it does seem to have such an effect on subjects from poorer countries.

Keep surveys short. In longer surveys IER occurs more often in the middle or towards the end. Thus, another way of tackling IER is to shorten surveys (if they are long, and if doing so does not involve sacrifices in reliability or validity).

Keep surveys simple. Quite obviously, it is also recommendable to keep surveys as comprehensible as possible (again, where doing so does not decrease reliability or validity).

Use CAPTACHAs. In recent months some results and paradata have suggested the increased presence of bots on MTurk. CAPTCHAs are a simple and effective way of preventing this worst form of IER. Depending on one's survey platform, they can be included with just a few clicks.

Appeal to subjects' conscience. Subjects have been found to show less IER if they are briefly informed about the amount of work that goes into designing and conducting a study, have to type their initials next to requirements (such as an acknowledgement of the study's length), or are asked to affirm a statement such as "I promise to carefully read each item and to provide an honest response".

Promote perceptions of meaningfulness. Finally, subjects may also show less IER if the first survey page explains the study's purpose and significance, and thanks subjects for their participation.

Note that having student samples (instead of subjects from crowdsourcing platforms such as MTurk), and having a physical or virtual proctor overseeing the completion of a survey have *not* been found to decrease IER.

Detecting IER

Some insufficient effort responding is bound to occur no matter what. In addition to the above prevention strategies, researchers should thus also attempt to *identify* IER subjects. The extant psychological literature suggests the following six strategies.

Use attention checks. Attention checks can take different forms (instructional manipulation checks, instructed-response items, bogus/infrequency items, and stand-alone checks), and have been found to be effective in detecting IER. To prevent negative side-effects one's usage of these checks should be disclosed in the survey instructions.

Use comprehension checks. Failures in comprehension checks need not indicate insufficient effort, but can also mean that a subject was not *able* to grasp relevant information (despite his or her best effort). In any case, such checks are recommendable as well, and should again be mentioned in the survey instructions.

Use response time measures. If a subject completes a survey overly fast chances are that he or she did not take sufficient effort. Completion time measures have accordingly been found to be another reliable IER indicator. Such measures can be based on absolute thresholds (e.g., less than x minutes) or statistical thresholds (e.g., more than two standard deviations below the mean).

Use response pattern analyses. IER sometimes results in patterned responses. Most often, subjects give the same answer several times in a row. This manifestation of IER may be detected by two kinds of *post hoc* analyses: LongString Analysis (a subject's longest string of identical responses exceeds a certain value), and Scale Straightlining Analysis (identical responses to all items of a certain number of scales).

Use response consistency analyses. IER also serves as a potential explanation of subjects providing inconsistent responses. The only consistency indicator that is recommendable to experimental philosophers proceeds by flagging subjects who provide different responses to the same question or item.

Include open-ended questions. Open-ended questions are a simple and highly effective way of identifying IER subjects, in particular bots and responses via virtual private servers (that enable subjects to participate even if they are not qualified to do so; for example, because they do not speak English).

I recommend using several of the above indicators (as each of them can only detect specific manifestations of IER), using indicators that suit one's study (in terms of length, usage or number of items, etc.), and stipulating conservative individual thresholds, i.e., thresholds that flag only obvious IER subjects (as at the moment, there are no established thresholds, and research on IER is still at its beginning). Note also that having subjects report their level of effort themselves and the identification of atypical responses do *not* seem reliable in detecting IER.

Dealing with IER

What if a subject has been identified as an insufficient effort responder? Given current recommendations in psychology, as well as some reflections of my own, the most appropriate

way to proceed seems as follows:

Exclude IER subjects from analyses. To avoid IER's negative psychometric effects, IER subjects should be excluded from analyses, i.e., researchers should first and foremost analyze non-IER subjects.

Preregister exclusion rules. Excluding IER subjects increases researchers' degree of freedom. Hence, exclusion rules should be formulated prior to running one's study, and should be registered at dedicated websites such as aspredicted.org or osf.io.

Report exclusions, and results with IER subjects included. Any exclusion of subjects should be reported. Moreover, researchers are advised to briefly explain the effect of these exclusions on their results (even if there is no (notable) effect at all).

Do not kick IER subjects out during the study. Comparing IER and non-IER subjects requires full data sets for both groups. Hence, subjects should not be kicked out of a survey immediately after they failed IER indicators (such as attention or comprehension checks).

Start with larger samples. Excluding IER subjects reduces sample size, and hence statistical power. This means that researchers will often have to start out with somewhat larger samples than they used to.

Do not (fully) compensate IER subjects. I recommend to not (fully) compensate subjects who have been found to engage in insufficient effort responding. Withholding compensation may require special permission from one's university's Institutional Review Board, as well as dealing with the angry backlash of excluded subjects. Yet, it seems needed to maintain the integrity of subject pools such as MTurk.

Further Information

Here are three good starting points for learning more about IER in general:

Curran, P. G. (2016): Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology* 66, 4-19.

Huang, J. L.; Curran, P. G.; Keeney, J.; Poposki, E. M.; DeShon, R. P. (2012): Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology* 27 (1), 99-114.

Meade, A. W.; Craig, S. B. (2012): Identifying careless responses in survey data. *Psychological Methods* 17 (3), 437-455.

In the second part of this two-part blog post I will present the results of a survey on IER in experimental philosophy.

If you are interested in a draft of my paper on IER in experimental philosophy (which includes more detailed explanations of the above strategies as well as of the survey) please

contact me at [thomas.poelzler\(at\)uni-graz.at](mailto:thomas.poelzler@uni-graz.at).