BEHAVE 2016
4th European Conference on Behaviour and Energy Efficiency
Coimbra, 8-9 September 2016

ON THE MAGNITUDE AND PERSISTENCE OF THE HAWTHORNE EFFECT – EVIDENCE FROM FOUR FIELD STUDIES

Verena Tiefenbeck[1,2]*

1: Bits to Energy Lab, Chair of Information Management, D-MTEC
ETH Zurich
WEV G218, Weinbergstrasse 56, 8092 Zurich, Switzerland
e-mail: vtiefenbeck@ethz.ch, web: https://www.bitstoenergy.ch

2: Institute of Applied Microeconomics
University of Bonn
Lennéstraße 43, 53113 Bonn, Germany
e-mail: vtiefenbeck@uni-bonn.de, web: https://www.iame.uni-bonn.de

Keywords: Hawthorne Effect, Randomized Controlled Trial, Energy Consumption

Abstract

Participants of research studies may exhibit altered behaviour resulting from awareness of being a part of an experimental study. Although this so-called Hawthorne effect has led to increased scrutiny in social science research, little is still known about the magnitude and persistence of that phenomenon. The findings of several recent field studies on resource conservation indicate the occurrence of substantial Hawthorne effects. This raises concerns that even robust studies with treatment and control may not sufficiently disentangle Hawthorne effects from participants' actual response to the intervention, casting doubts on the treatment effects reported and on the generalizability of the findings to non-study populations. In this article, we seek to gauge the magnitude and time dynamics of the Hawthorne effect. We analyse the behaviour of the control groups in four independent randomized controlled trials on resource conservation conducted in Switzerland, the Netherlands, and Singapore. Each study comprised between 525 and 800 households and collected detailed measurement data on participants' energy and water use over two to six months. We find strikingly consistent patterns across the four studies: each control group initially increased their resource use and transitioned within the first few weeks into a relatively stable level of consumption, 5-20 percent above the initial level. Our interpretation is that control group households reduced their resource use at the beginning of the study compared to their pre-study behaviour, yet relapsed towards their pre-study behaviour as they got used to the measurement device. We can rule out seasonal trends as an alternative explanation and attribute these patterns to Hawthorne or salience effects. Our data suggest that the current practice of collecting baseline data at the beginning of studies may lead to biased reference points, as the baseline coincides with the period most affected by the Hawthorne effect.

Related work

The Hawthorne effect inherited its name from a series of studies on on workplace productivity and illumination levels at the Hawthorne plant in the 1920ies. While the results attibuted to those studies did not stand up to later more careful analyses [1], the authors of the re-examination study still qualify the concept of the Hawthorne effect to *"stand among the most influential social science research of the twentieth century"* (p. 237). The Hawthorne effect can considerably bias estimates of the actual treatment effect and represents *"one of the greatest threats to the generalizability"* of studies (Levitt and List 2011, p.225). Yet, both the magnitude [2] of the Hawthorne effect and its time dynamics have proven as difficult to capture: as Levitt and List (2011, p. 228) point out, *"it is unclear whether the Hawthorne effect is a short-run or long-run phenomenon."* The Hawthorne effect may influence the behaviour of experimental subjects through several channels [1], [3], [4]. First, increased scrutiny and processes that accompany the experiment (e.g., reading of instructions, etc.) may affect treatment and control groups alike. Second, the manipulation itself that may serve as a reminder to the participants that they may be under observation. Third, experimental subjects may attempt to act in ways that will please the experimenter or what they think the experimenter considers as appropriate behaviour ("experimenter demand effects").

Recent field studies on resource consumption have revived discussions of the Hawthorne effect: Schwartz et al. (2013) found that simple postcards reminding residents of their participation in a study on residential electricity use induced a 2.7% reduction (in the absence of any real intervention). That effect size is equivalent to the treatment effects of programs that send periodic "Home Energy Reports" to households, in which the control group receives no notification (e.g., (Allcott and Rogers 2014)). In a field study that provided near-real-time plug-level information to participants, Attari et al. (2014) measure a 12–23% decrease in electricity use for treatment apartments, yet conclude that the changes may be entirely driven by Hawthorne or salience effects. They authors explain the large magnitude of those effects with the high level of intrusiveness of their intervention. In a study with 335 airline pilots, [7] even find an increase of almost 50% in the implementation of efficient flight and taxi practices compared to the pre-experimental period, which they also attribute to Hawthorne effects. We add to that discussion by providing evidence from similar interventions in four different locations. The granularity of our data allows us to investigate the temporal dynamics of the Hawthorne effect and thus to estimate the magnitude of the effect at different stages of our four field experiments.

Methodology

In a series of four large independent, but similar field studies, participating households received smart shower meters that measured and recorded energy and water consumption in the shower over a period of two to six months. The four studies were implemented independently in different years, in different seasons, and in three different countries (two in Switzerland, one in Singapore, and one in the Netherlands). In each country, participants were recruited in collaboration with a local corporate partner who advertised the study as an energy efficiency study (in the case of Singapore: as a water conservation study). The local partner

organization who helped distribute the devices to customers or employees who were willing to participate (opt-in design). In the three European studies, participants installed the devices themselves in their showers, while that process was carried out by research assistants in the Singapore study. The accompanying user manuals stated that the smart shower meter measure and record water and energy use data and explained the group-specific content displayed on the screen of the device. All four studies were designed as a randomized controlled trial: households were randomly allocated to the control group or to one of the treatments. Treatment group devices displayed feedback on the energy and water consumption of the current shower. By contrast, control group devices displayed only water temperature (Figure 1). Thus, control group devices equally indicated the ongoing measurement activity of the device, but did not provide information on the energy or water use to the participants.

Figure 1: Control group devices displayed only water temperature.

The smart shower meters collected granular data over two to six months on a specific behaviour, recording energy and water use of every shower taken. At the end of every study, that dataset was collected by a team of researchers (for details on the data retrieval, please see [8]). The total dataset collected in the four studies consists of 429,593 datapoints. In this paper, we focus on the behaviour of the control groups of those four studies, a total or 96,279 observations recorded in 642 households. Thus, we analyse the behaviour of the control groups from four independent studies that took place in three countries. All four studies share the same measurement technology, the same target behaviour (resource conservation in the shower), and the same display content provided to the control group (water temperature in degrees Celsius).

Findings and interpretation

Figure 2 shows the mean energy use per shower of the control group (N=198) for one of the four studies. One can see a clear upward trend at the beginning of the study that transitions into a relatively stable consumption within the first month. Even if we neglect the very first datapoints, which exhibit the largest increase in consumption, our data still show an upward trend that transitions within the first month into a relatively stable level of consumption, 5-20

percent above the initial level. We observe very similar patterns in terms of magnitude and time dynamics in all four field studies, independent of the season and country in which the studies took place. Thus, we are able to rule out local or seasonal trends as alternative explanation for the patterns observed.

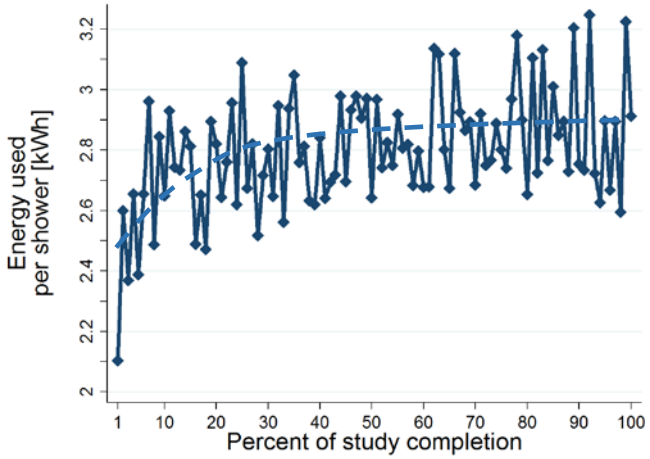

Figure 2: Average energy use per shower of the control group in one of the studies that took place in Switzerland.

Our interpretation is that the measurement data collected in the first days and weeks underestimate the amount of energy and water participants would have used in the absence of the measurement device. While we have no possibility to measure pre-study level of resource consumption for the target behaviour without the participants' knowledge (and consent), the consistency of the patterns of increase and subsequent saturation across all four studies suggest that participants reduced their resource consumption at the beginning of the studies compared to their pre-study level of consumption. We suspect that participants gradually relapsed towards their (unobserved) pre-study consumption levels, as they got used to the measurement device over time. The explanation that we deem as most plausible for these patterns is that at the beginning of every study, participants were particularly subject to Hawthorne effects. In fact, the Hawthorne effect may affect control group participants though several of the channels described by [1], [3], [4]: participants deployed the smart shower meters themselves (in three out of four studies) and the user manuals explicitly stated that the devices measured and stored data on the energy and water consumption. As a result, both the installation process and the accompanying materials may have drawn attention to the measurement process, in particular at the beginning of the study. Furthermore, participants had consented to share their shower data with the researchers at the end of the study. Thus, participants knew that their resource consumption data were recorded and that they would be analysed

by the researchers. The presence of the device may have served as a physical reminder for their participantion in the study. Over time, participants may have gotten used to that reminder. Moreover, the field trials in Switzerland and the Netherlands had been framed as energy efficiency studies and in the case of Singapore, as a water conservation study. This framing may have served as a cue to participants that resource conservation (i.e., taking shorter showers) was the appropriate behaviour expected by the researchers ("experimenter demand effect").

Discussion and limitations

The results of this series of four studies suggests that the behaviour (or measurable behavioural outcomes in terms of energy and water use) of our study participants may deviate considerably from the way the individuals would behave naturally, in particular at the beginning of a study. We observe consistent patterns across all four studies: an increase in the control group's energy (and water) consumption per shower at the beginning of every study, followed by a subsequent saturation and stabilization of the behaviour (i.e., resource use per shower).

Our primary conjecture that the control groups' behaviour at the beginning of the four studies is subject to Hawthorne effects: participants alter their behaviour because they are aware that they are part of an experimental study and being monitored. An alternative explanation might be salience: the provision of the measurement device and the user manual may have drawn participants' attention to showering as a resource-consuming behaviour. Ultimately, we will have to leave it to future research to disentangle those two mechanisms. In any case, both explanations imply that the data collected at the beginning of such studies may not be representative of individuals' natural behaviour. This is not only problematic for studies without a control group, which may produce biased results. Even for studies with a control group and difference-in-difference design, this may still generate problems: the shift of the baseline to a lower level influences the relative savings, in which the baseline appears as a denominator.

What is unique about the dataset collected is the granularity of the consumption data and the focus on a single behaviour. These characteristics make it possible to closely monitor how participants' behaviour changed over time in the course of the study. Moreover, while Hawthorne effects are extremely difficult to gauge in domains like medical studies, which are often characterized by a very small sample and few measurement points for cost reasons and ethical concerns [2], the dataset analysed consists of four larger samples. In contrast to the vast majority of other field studies on residential energy or water consumption that collect only aggregated household-level data and often only on a monthly or quarterly basis, our dataset makes it possible to better gauge the rate and magnitude of behaviour change in the control group over time. In addition, the series of four independent, but similar studies suggests that the upward trend observed is not an

accidental artefact or due to seasonal trends.

However, in the absence of measurement data, we cannot be sure about participants' behaviour prior to the installation of the measurement devices. While we consider weakening Hawthorne effects as the most plausible explanation for the initial increase in consumption, we are not able to proof that participants' resource consumption was higher prior to the study than in the first showers recorded by the measurement devices.

Future research should also determine to what extent the visibility of the measurement equipment plays an important role. Both in our studies and in [6], the devices installed in the participating households were visually and physically rather prominent. It is plausible that more discrete measurement sensors could reduce both, the potential for Hawthorne effects and undesired salience effects among control group households.

References

[1] S. D. Levitt and J. A. List, "Was There Really a Hawthorne Effect at the Hawthorne Plant? An Analysis of the Original Illumination Experiments," *Am. Econ. J. Appl. Econ.*, vol. 3, no. January, pp. 224–238, 2011.

[2] R. McCarney, J. Warner, S. Iliffe, R. van Haselen, M. Griffin, and P. Fisher, "The Hawthorne Effect: a randomised, controlled trial," *BMC Med. Res. Methodol.*, vol. 7, no. 1, p. 30, Jul. 2007.

[3] S. D. Levitt and J. A. List, "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?," *J. Econ. Perspect.*, vol. 21, no. 2, pp. 153–174, May 2007.

[4] D. J. Zizzo, "Experimenter demand effects in economic experiments," *Exp. Econ.*, vol. 13, no. 1, pp. 75–98, 2010.

[5] H. Allcott and T. Rogers, "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation," *Am. Econ. Rev.*, vol. 104, no. 10, pp. 3003–3037, 2014.

[6] S. Z. Attari, G. Gowrisankaran, T. Simpson, and S. M. Marx, "Does information feedback from in-home devices reduce electricity use? Evidence from a field experiment," National Bureau of Economic Research, 2014.

[7] G. K. Gosnell, J. A. List, and R. D. Metcalfe, "A New Approach to an Age-Old Problem: Solving Externalities by Incenting Workers Directly," 2015.

[8] V. Tiefenbeck, V. Tasic, S. Schoeb, and T. Staake, "Mechatronics to drive environmental sustainability: Measuring, visualizing and transforming consumer patterns on a large scale," in *Proceedings of the IEEE IECON 2013, special section on energy informatics*, 2013.