



**A Random Shock Is Not
Random Assignment**

Christoph Engel





A Random Shock Is Not Random Assignment

Christoph Engel

May 2016

A Random Shock Is Not Random Assignment*

Christoph Engel

May 2016

Abstract

A random shock excludes reverse causality and reduces omitted variable bias. Yet a natural experiment does not identify random exposure to treatment, but the reaction to a random change from baseline to treatment. A lab experiment comparing higher certainty with higher severity of punishment for stealing (holding the expected value of the intervention constant) shows that the difference between the effects of a random shock and random assignment can be pronounced.

Keywords: identification, random exposure, random shock, natural experiment, certainty and severity of punishment

JEL: C01, C12, C90, K14

* Helpful comments by David Jäger, Jonathan Klick, Adrian Hillenbrand and Anna Kochanova on an earlier version, excellent research assistance by Rafael Schwalb, and programming of the experiment in zTree by Anastasia Niechaieva are gratefully acknowledged. The experiment has been funded from the regular budget of the Bonn Max Planck Institute. The Hamburg Econ Lab has kindly run the experiment.

1. The issue

With the credibility revolution, empirical economics has changed fundamentally (Leamer 1983, Manski 1995, Angrist and Pischke 2010). No good publication takes correlation for causation. It seems the enthusiasm for instrumental variables has also somewhat cooled down. Too often their exogeneity is questionable, or they are weak (Bound, Jaeger et al. 1995). A way out are random shocks. They create a “natural experiment” (Meyer 1995). Researchers have warned that they are not true experiments if the untreated and the treated cases are not comparable (Sekhon and Titiunik 2012). Yet another obvious challenge seems to get short shrift. The random shock exposes individuals to a random *change*; interest tends to lie in the causal effect of the *presence* of some influence. The natural experiment identifies the *reaction* of individuals who had not been exposed to the influence in the past. Had those same individuals been exposed to that influence ever since, there might be no effect at all, or a different effect. The reaction to the sudden change may quickly fade away, or it may be conditional on experiences the individuals have made in the past.

This note uses a lab experiment to demonstrate that these concerns can be serious. As an illustration, it uses a long-standing issue in criminal policy. Which is more effective in deterring crime: higher severity, or higher certainty? The criminological literature has usually found a certainty effect (Antunes and Hunt 1973, Paternoster 1987, Klepper and Nagin 1989, Maxwell and Gray 2000, Pogarsky 2002). But severity effects have also been documented (Engel and Nagin 2015). The question is of key policy relevance. Should a country rather invest in the police force (certainty) or in building prisons (severity)? Both is costly. Which way is the taxpayer’s dollar better spent? Policymakers may be interested in the transitory effect. They may want to show their voters that they take their anxieties seriously. But from a welfare perspective, the country would want to pick an intervention that has a lasting effect. It is the long-term effect that matters for cost-benefit analysis.

Now assume an empiricist has a lucky day and finds two otherwise reasonably comparable communities that adopt diametrically opposite policies. One community stops paying extra shifts for the police (reduces certainty), and builds a new prison so that less convicts must be put on probation (increases severity). The other community closes a prison (decreases severity), and hires more policemen (increases certainty).¹ In the lab, I mimic this research design. I do indeed find that reducing certainty (while increasing severity) increases the frequency of stealing. But if severity is reduced (while certainty increases), participants steal even more. Critically in the lab I can compare the effect of changes with a difference in initial exposure. The mere exposure to a more certain (but less severe) sanction virtually has no effect. Hence what seems to be a near perfect research design only identifies the reaction to a *change*.

1 When just changing severity or certainty in isolation I would confound it with the expected value of the sanction.

2. Design

In the experiment, participants are randomly assigned to groups of two. Each participant is endowed with 20 tokens. Each decides in the active role whether she wants to take any of the passive participant's tokens. Afterwards, (using the strategy method) in every period the active role is randomly assigned. Participants know they can lose tokens if are found out to have taken tokens from their counterpart. The experiment is framed neutrally and speaks of “taking”, “checking”, tokens that are “forfeited”, and additional tokens that are “subtracted”. The expected value of this intervention is such that a participant exclusively interested in maximizing payoff is indifferent between stealing any amount and not doing so. In the *low certainty* condition, audit probability is 37.5%; upon audit the stolen tokens are forfeited and 1.66 tokens are subtracted for each token the participant had taken. In the *high certainty* condition, audit probability is 50%, severity is 1:1. The first part of the experiment lasts announced 8 periods. Participants are randomly assigned to one of the conditions. In each period, participants are rematched (in unannounced matching groups of size 6, to preserve independence (cf. Charness 2000)). Participants know that the experiment has further parts, but not what these parts are about. Actually the second phase is an exact repetition for another 8 periods. But matching groups that were in the *low certainty* condition are now shifted to the *high certainty* condition, and vice versa. After period 16, matching groups are shifted back to the original regime for another 8 periods.² All periods are paid out.

The experiment was run in the economics lab of Hamburg University. 144 (141 students of various majors) participated in the experiment, 55.56% female, mean age 25.03. The experiment was programmed in zTree (Fischbacher 2007). Participants were invited using hroot (Bock, Baetge et al. 2014). They on average earned 14.86 €(16.37 \$).

3. Results

For welfare, the sustainable deterrent effect is most relevant. In the experiment, one way to test for it is comparing choices in the first round. In this round I neither find a treatment effect on the frequency of stealing,³ nor on the stolen amount. Alternatively one may compare choices in the entire first phase (of 8 periods). This informs about the deterrent effect on would-be criminals with growing experience. In Table 1, this effect is captured by *high certainty first*. In the logit model, I find a certainty effect. Yet I do not find the effect in the categorical component of a double hurdle model (for the underlying statistical theory see Engel and Moffatt 2014). This statistical approach is more appropriate: many participants (next to) never steal (there is a hurdle). Both the frequency of stealing and the stolen amount change substantially over time (Figure 1). This suggests that the decision not to steal can result from

2 I also administer the test for risk aversion by Holt and Laury (2002), the 10 item version of the Big 5 Rammstedt and John (2007), trust and morality items from the SOEP, and ask demographic questions. I do however not need this additional data for this paper.

3 A dummy that is 1 if the participant has taken a positive amount.

type, but also from a reaction to experiences (there is a double hurdle). There is no significant certainty effect on the stolen amount, whatever the statistical model. Hence a researcher with access to the true data would conclude that the certainty effect is dubious.

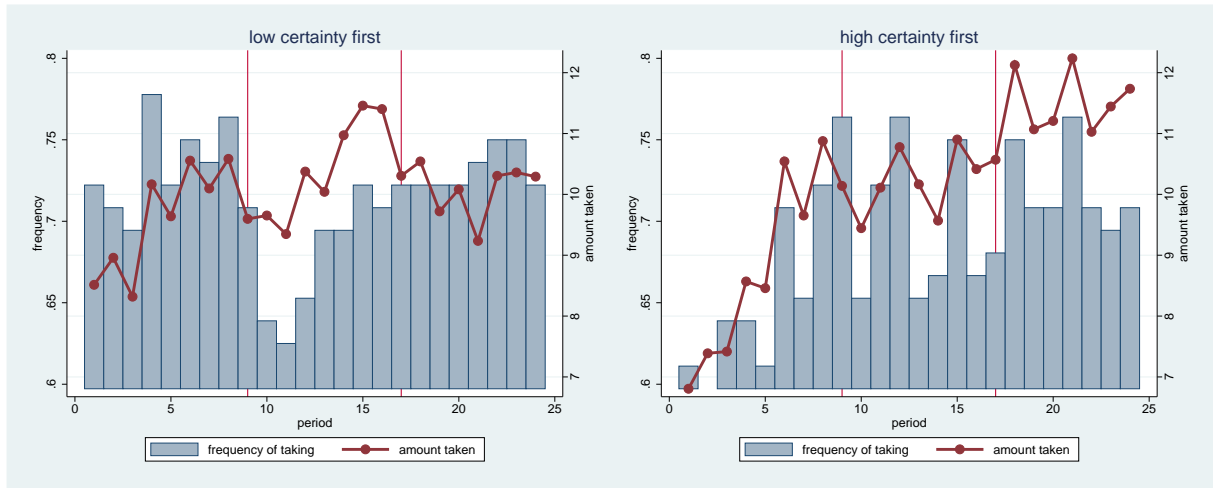


Figure 1
Descriptive Statistics

	frequency of stealing		amount stolen	
coefficients	logit	hurdle	tobit	hurdle
<i>high certainty first</i>	-1.376* (.534)	-.262 (.213)	-3.054 (3.375)	.417 (1.187)
phase 2	-.633** (.204)	-.187 (.147)	1.600+ (.911)	2.305** (.817)
phase 3	-.043 (.207)	-.024 (.094)	.934 (.899)	.821+ (.619)
<i>hcf</i> * phase 2	1.149*** (.271)	.339* (.174)	2.854* (1.288)	-1.221+ (.803)
<i>hcf</i> * phase 3	.661* (.274)	.196+ (.141)	6.539*** (1.293)	1.859** (.683)
cons	2.696*** (.400)	.670*** (.178)	8.849*** (2.385)	12.791*** (1.095)
Wald tests				
<i>lcf</i> : phase 3 – 2	.590**	.162*	-.626	-1.484*
<i>hcf</i> : phase 2 – 1	.516**	.159+	4.414***	1.084*
<i>hcf</i> : phase 3 – 2	.102	.020	3.059***	1.596***
phase 2 – 1: <i>hcf</i> – <i>lcf</i>	1.149***	.339*	2.854*	-1.221+
phase 3 – 2: <i>hcf</i> – <i>lcf</i>	-.488+	-.143	3.060**	3.080***

Table 1
Regressions

model 1: logit, model 3: tobit (lower limit 0, upper limit 20), models 2 and 4: components of double hurdle model,

models 1 and 3: mixed effects, choices nested in individuals nested in matching groups

models 2 and 4: bootstrap with cluster for matching groups

reference category: *low certainty first*

standard errors in parenthesis

*** p < .001, ** p < .01, * p < .05, + p < .1

Now turn to the evidence that might at best be available in the field. One could compare either community (within subjects) before and after the first regime change. For the *low certainty first* community, in the regressions this is the main effect of phase 2. There is a relatively clear severity effect on the stolen amount, and a more questionable certainty effect on the frequency of stealing. For the *high certainty first* community, Wald tests show a certainty effect on both dependent variables. The researcher could also exploit that she observes both communities before and after a shock and, in the spirit of a difference in difference approach, compare their sensitivity to regime change. Stealing frequency increases (more) when increasing severity while reducing certainty. But this comparison is inconclusive regarding the stolen amount.

What if, instead, the researcher did not observe the first random shock, but the second, the change back to the original regime? When going back to high severity, stealing becomes more frequent, but less is stolen (Wald test: *lcf* phase 3-2). When going back to high certainty, stealing is not more frequent, but more is stolen (Wald test: *hcf* phase 3-2). This is a severity effect. It also shows up when comparing the sensitivity to the second regime change across conditions.

4. Conclusions

Mimicking a natural experiment in the lab demonstrates that a random shock is not a panacea. One identifies reactions to a change in circumstances. This change need not be a good proxy for the presence or the absence of the circumstance in question. Essentially the random shock is just an instrument. In the example (with the first change) one would conclude that there is a certainty effect on the frequency of stealing, but a severity effect on the stolen amount. Yet there is no effect at all on unexperienced participants, and there is at best a certainty effect on the frequency of stealing, and no effect on the stolen amount, if participants have experience. Natural experiments will often be the best available evidence. Usually all individuals have been in some regime before the shock. But researchers should be cautious not to overinterpret their findings. A random shock is not random assignment to treatment. One sees the effect of an intervention *conditional* on the original state in which the observed population happened to be before the random shock.

- ANGRIST, JOSHUA AND JÖRN-STEFFEN PISCHKE (2010). "The Credibility Revolution in Empirical Economics." *Journal of Economic Perspectives* **24**: 3-30.
- ANTUNES, GEORGE AND A.LEE HUNT (1973). "The Impact of Certainty and Severity of Punishment on Levels of Crime in American States." *Journal of Criminal Law and Criminology* **64**: 486-493.
- BOCK, OLAF, INGMAR BAETGE AND ANDREAS NICKLISCH (2014). "hroot." *European Economic Review* **71**: 117-120.
- BOUND, JOHN, DAVID A JAEGER AND REGINA M BAKER (1995). "Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variable Is Weak." *Journal of the American Statistical Association* **90**(430): 443-450.
- CHARNESS, GARY (2000). "Self-Serving Cheap Talk." *Games and Economic Behavior* **33**: 177-194.
- ENGEL, CHRISTOPH AND PETER G. MOFFATT (2014). "dhreg, xtdhreg, bootdhreg. Programs to Implement Double Hurdle Regression." *Stata Journal* **14**: 778-797.
- ENGEL, CHRISTOPH AND DANIEL NAGIN (2015). "Who is Afraid of the Stick?" *Review of Behavioral Economics* *******: *******.
- FISCHBACHER, URS (2007). "z-Tree." *Experimental Economics* **10**: 171-178.
- HOLT, CHARLES A. AND SUSAN K. LAURY (2002). "Risk Aversion and Incentive Effects." *American Economic Review* **92**: 1644-1655.
- KLEPPER, STEVEN AND DANIEL NAGIN (1989). "The Deterrent Effect of Perceived Certainty and Severity of Punishment Revisited." *Criminology* **27**(4): 721-746.
- LEAMER, EDWARD E. (1983). "Let's Take the Con out of Econometrics." *American Economic Review* **23**: 31-43.
- MANSKI, CHARLES (1995). *Identification Problems in the Social Sciences*. Cambridge, Harvard University Press.
- MAXWELL, SHEILA ROYO AND M.KEVIN GRAY (2000). "Deterrence." *Sociological Inquiry* **70**(2): 117-136.
- MEYER, BREED D (1995). "Natural and Quasiexperiments in Economics." *Journal of Business & Economic Statistics* **13**(2): 151-161.
- PATERNOSTER, RAY (1987). "The Deterrent Effect of the Perceived Certainty and Severity of Punishment." *Justice Quarterly* **4**: 173-217.

POGARSKY, GREG (2002). "Identifying 'Deterrable' Offenders." *Justice Quarterly* **19**: 431-452.

RAMMSTEDT, BEATRICE AND OLIVER P. JOHN (2007). "Measuring Personality in One Minute or Less." *Journal of Research in Personality* **41**: 203-212.

SEKHON, JASJEET S AND ROCIO TITIUNIK (2012). "When Natural Experiments Are Neither Natural nor Experiments." *American Political Science Review* **106**(01): 35-57.

Online Appendix Instructions

General Instructions for Participants

Welcome to our experiment. If you read the following instructions carefully, you will be able to earn a substantial sum of money, depending on the decisions you make. It is therefore very important that you read these instructions carefully.

The instructions you have received are exclusively for your private information. **There shall be absolutely no communication during the experiment.** If you have questions, please ask us. Disobeying this rule will lead to exclusion from the experiment and any payments.

All your choices during the experiment and the payoff you receive at the end remain anonymous. Please do also not discuss them with other participants after the experiment.

During the experiment, the currency in operation is not euro, but taler. Your entire income is hence first calculated in taler. The total number of taler you and all other recipients have accumulated in the course of the experiment will then be exchanged into euro at the following rate:

1 taler = 3 euro cent.

At the end of the experiment, you will receive a **cash** payment, in euro, of whatever number of taler you have earned.

Additionally, for your participation in the experiment, you once receive a show-up fee of

25 taler

It has the purpose of compensating you for potential losses.

The experiment consists of several parts. We first explain the first part. For the remaining parts you will receive new instructions.

Instructions for the first part of the experiment

The first part of the experiment is divided into individual periods, of which there are a total of **eighth**. At the beginning of each period, every player receives an endowment of 20 taler. In each period, there are **two** roles, one active and one passive. Whether you have the active or the passive role will only be randomly determined after all players have taken their decisions. At the end of each period, the computer randomly matches you with another participant. Every period, you are randomly rematched.

In each period, you decide whether you want to take any portion of the passive player's endowment, and if so, how many taler. This choice is payoff relevant in case you will later be assigned the **active** role.

During the experiment, the computer will only sporadically check whether a participant has taken taler from the randomly matched participant. This happens with probability 50% [37.5%]. You will not know in advance whether your choice will be checked. If the computer finds out that you have taken taler, these taler are forfeited; they are not given back to the participant from which you have taken them. Additionally, you are subject to a payment that is immediately withdrawn from your account. This payment is **1 [1.66]** times the taken amount of taler. Should your previous earnings from the experiment be less than the amount subtracted, your show-up fee of 25 taler is reduced. Should also this amount be used up, you receive 0 taler. At the end of the instructions you find a table that informs you about the amount subtracted should the computer check you. Note that the computer picks participants independently of findings in previous periods. Consequently, a participant who has been found to take taler once is not more likely to be checked in future periods than all other participants.

At the end of each period, the computer executes the following random draws:

- You are randomly matched with one other player. This match is independent from matches in earlier or later periods.
- It is randomly decided whether you have the active or the passive role. You have the active role in one of two cases. There is a 50% chance of your being assigned the active role.

At the end of each period, you learn which role has been assigned to you, and your payoff.

Do you have any questions?

The following questions have the purpose of checking whether you have understood the experiment. They are only meant to quiz you. Your answers have no effect on your payoff from the experiment.

<table of payoff conditional on audit and amount taken; not reproduced>

Instructions for the second part of the experiment

The second part of the experiment lasts another eight periods. The experiment differs from the first part in two respects only: 1.) In each period, the computer will now randomly pick active participants with a probability of **37.5% [50%]** and check whether they have taken taler. If a participant is found to have taken taler, the taken amount is forfeited. Moreover this participant has to pay. 2.) Per taler taken another **1.66 [1]** taler are withdrawn from this participant's account. The following table informs you which amount is subtracted from your account, should you be singled out as the active participant, and should you be audited. Consequently all that changes in this part is the probability with which you are checked, and the number of taler that may be subtracted.

<table not reproduced>

Instructions for the third part of the experiment

The third part of the experiment is exactly the same as the first. The third part of the experiment lasts another eight periods. In each period, the computer will now randomly audit **50% [37.5%]** of the participants and check whether they have taken taler. If a participant is found to have taken taler, the taken amount is forfeited. Per taler the participant has taken **1 [1.66]** taler are withdrawn from this participant's account. The probability of being audited and the number of taler that may be withdrawn is thus exactly the same as in the first part of the experiment. The following table informs you which amount is subtracted from your account, should you be singled out as the active participant, and should you be audited.

<table not reproduced>