Sequencing Effects in the Analysis of Complex Experiments in Business Research: Mechanisms, Biases, and Recommendations

Peter Kotzian
University of Düsseldorf, Universitätsstraße 1, Düsseldorf
Peter.Kotzian@hhu.de
DOI: 10.34190/JBRM.17.3.006

Abstract: Experiments in business research became more complex over time, yielding complex sequences of stimuli and measurements. This raises the issue of sequence effects, where effects are found only in specific sequences of the experiment. One case in point is factorial surveys. Here, presenting the stimulus is followed by asking subjects to evaluate several vignettes presented in a certain sequence. The researcher is interested in the effect of the stimulus on responses to vignettes with certain features. As sequence and stimulus can be made uncorrelated by construction, holding the sequence constant or excluding the sequence from the analysis seems to be justified when researchers are only interested in effects of vignette features or the stimulus. In both cases, even if the sequence is relevant for the dependent variable, correlation between sequence and stimulus, the necessary condition for an omitted variable bias, is absent. The effect estimated for the stimulus should thus be unbiased. We show that even in the case where stimulus and sequence are uncorrelated or the sequence is held constant, an omitted variable bias occurs when the effect of the stimulus in a vignette is in its magnitude dependent on the sequence in which the vignettes were presented. Such an effect would be modeled by including a sequence-stimulus-interaction term and the omitted variable is this interaction term, which is, by construction, always correlated with each of the constitutive variables. A simulation is presented to illustrate the problem. Implications for experimental research are discussed.

Keywords: Experimental Design; Factorial Surveys; Order-effects; Omitted-Variable Bias
JEL Codes: C21; C91

1. Introduction: Simple and Complex Experiments in Business Research

Experiments became, also due to the spread of behavioral economic research issues, a standard research instrument in business research (Libby, Bloomfield and Nelson 2002). Basically, experiments are about estimating the effect of a manipulation, alternatively labeled treatment or stimulus, on a dependent variable.

An experimental setting has, in its simplest version, the following design: after a randomized assignment to control and experimental groups, participants are presented with the stimulus, which is followed by the observation of the dependent variable of interest. Measurements of the dependent variable from control and experimental groups are compared; the difference is attributed to the stimulus (Campbell and Stanley, 1963) and tested for significance. Over time, experimental designs became much more complex (Kirk 2012; Montgomery 2012). The clearest example of this development is the combination of survey elements with experimental stimuli to factorial surveys (Rossi and Anderson, 1982; Auspurg and Hinz 2015; Oll et al. 2018).

Depending on the research question, there is still an "explicit" stimulus, which is given to some but not all participants, but there are also stimuli build into the vignettes, short, stylized situations where certain situational features are manipulated, which are given to all subjects. The statistical analysis of factorial surveys allows identifying effects of the stimulus, the vignette features, and personal features on some dependent variable (Hox, Kreft and Hermkens, 1991).

As an example of a factorial survey on which the arguments presented later on are based, take the following: researchers are interested in whether a Code of Conduct, henceforth CofC, affects ethical behavior in typical situations arising in the business context. The researcher selects a random sample of participants, splits it into the control group, receiving no CofC, and the experimental group, receiving a CofC. Then each group is presented with some vignettes, i.e., decision situations involving different ethical problems from the business context. The researcher may present a set of different vignettes with constant features, for instance accepting a gift, asking for a kick-back, or misappropriating company property. Alternatively, several versions of the same vignette can be presented, where some features of the vignette vary. The basic vignette may be about accepting a gift from a business partner, but the researcher manipulates how valuable the gift is, or whether the gift is offered in private or in public; see Stöber et al. (2016) as an example of this procedure. In both

Reference this paper: Kotzian, P., 2019. Sequencing Effects in the Analysis of Complex Experiments in Business Research Mechanisms, Biases, and Recommendations. *The Electronic Journal of Business Research Methods,* 17(3), pp. 165-178, available online at www.ejbrm.com

versions, the dependent variable might be an intention to behave in a certain way, located on a continuum from unethical to ethical. From the perspective of the data generating process, the response is a result of the stimulus (Was a CofC presented?), of vignette features (Was the setting described as private?), but also of personal features (How strong is the participant's risk propensity?). Typically, participants are given several vignettes and make several decisions, yielding a data set where individual cases – the decisions of a participant – are nested at the participant level.

Factorial surveys allow for various degrees of complexity in the analysis of the data obtained. One the one hand, applying statistical models for nested data allows identifying and isolating effects of the stimulus, vignette features, and participant features, including interactions among them (Oll et al. 2018). On the other hand, a simple analysis can compare control and experimental group, which allows the researcher to estimate the effect of the stimulus, in our example the effect of the CofC on ethical behavioral intentions. Technically, this is achieved by comparing measurements of the dependent variable for control and experimental groups across all vignettes, i.e., pooling all decisions made by a participant. The effects of the vignette and its features can be isolated in an equivalent manner (Hox, Kreft and Hermkens, 1991). While the effect of a CofC might not be equal for all vignettes, this simple mode of analysis allows for a general statement of whether the CofC works or not.

Of particular relevance for the present paper is the fact that several vignettes are presented sequentially. In this, factorial surveys are similarly to other experiments, which also entail series of events, e.g. when cooperation over several rounds is analyzed (Dal Bó and Fréchette, 2011), or a series of decisions is analyzed, e.g., investment behavior in several rounds (Benartzi and Thaler, 1999; Klos, Weber and Weber, 2005).

What these experiments have in common is that one event may, potentially, affect later events. Having invested in the risky options in the first rounds may cause subjects to be more risk-averse in the latter rounds, or vice versa. In the case of more complex experiments, with several vignettes being presented sequentially, the decision in the second vignette is potentially affected by the content of the vignette presented first and also the decision the participant made in the first vignette (Johnson-Laird and Shafir, 1993). Consider the situation where one vignette concerns the acceptance of a small gift, the other the acceptance of an envelope with cash in a situation where the donor evidently offers the "gift" because s/he has a clear interest at stake.

The participant will compare the situation presented second with the situation presented first, and it is likely that this comparison will affect the choice made in the second situation. The gift may – if offered first – be regarded as acceptable, as it is, at first glance, just a gift. If a bribe is offered first, this sets the frame what an interaction is about, affecting acceptance of the gift in the second situation. As the situational frame differs, a stimulus like a CofC, may work for the gift vignette only in a particular sequence, e.g., only if the bribe vignette came first and the participant is sensitized for ethical or legal aspects of the situation.

Dealing with order effects is part of the standard repertoire of experimental psychology (Krosnick and Alwin, 1987; Hogarth and Einhorn, 1992, Zeelenberg and Pecher, 2015), survey design (Strack 1992; Tourangeau, Rips and Rasinski, 2000) and experimental design (Mitchell and Jolley, 2010). While at times, the effect of the sequence is of interest per se (Wiegmann, Okan and Nagel, 2012), the primary interest is in how order effects come about, psychologically, and how they can be reduced or precluded by choosing an appropriate experimental design. Specifically for factorial surveys, where often many, highly similar vignettes, are presented sequentially, order effects are a persisting issue (Auspurg, Hinz and Liebig, 2009). The problems involved concern, among other things, learning, rapidly decreasing attention, anchoring, framing, and assumed importance (Auspurg and Jäckle, 2017). Recommendations for conducting factorial surveys are based on arguments of plausibility, for instance that starting with extreme cases should be avoided in favor of random orders (Auspurg and Hinz, 2015, Aguinis and Bradley 2014.).

But complex experiments are not only subject to order effects in the sense that participants get more risk averse as the experiment goes on. It may also be the case that effects of stimuli are themselves conditional on the sequence of events. Consider the example of presenting a CofC before giving participants a series of vignettes covering ethical problems. If the sequence starts with a vignette, which is clearly posing an ethical problem, say, a clear attempt to bribe, participants may recognize this and use the CofC as a guidance right from the start. This effect, using the CofC as a guideline, may persist and the CofC is effective right from the start. The effect of the CofC is then stronger, if the individual vignettes are considered – say, the bribe vignette

– but also when all vignettes are pooled to obtain a summary measure of behavior. If the sequence starts with a vignette, which is less ethically charged, say the offering of a small gift, this may not be the case. The situation is not seen as posing an ethical problem and thus, the CofC is not seen as being relevant. In the analysis, the CofC will be less relevant for the first vignette, but also for a summary measure of ethical behavior. Thus, the CofCs effect depends on the sequence in which the vignettes are presented.

Given the strong evidence of order effects it is interesting to note that practical usage and dealing with order effects in real experiments and their usage in the statistical analysis are rather vague issue. Often, little can be found in terms of how sequences are used and in particular, how they are analyzed later on. As this is "footnote stuff", screening the literature is highly bothersome. To the best of our knowledge, random presentations of the vignettes or counterbalanced designs, where two sequences with inversed order are used, prevail. The issue of sequences is dealt when describing the experimental design, but rarely if ever, sequences are used in the statistical analysis, in particular not as a factor affecting the effect of a stimulus.

This paper's research questions are: under what conditions is this attribution of an observed effect to a stimulus correct? Or, is the effect of stimulus in a specific vignette partly or fully conditional on the sequence in which the vignettes were presented? What is the appropriate specification of a statistical model, when sequence effects are present and affect not only the dependent variable but also the strength of a stimulus?

The paper's contribution is the analysis of the mechanisms at work in complex experiments, which are generating the information observed. We argue that excluding the sequence is a special case of an omitted variable bias. If the effect of the stimulus is in some form dependent on the sequence, omitting either the variable indicating the sequence or the variable capturing the interaction between sequence and stimulus leads to an omitted variable bias, as by construction, the sequence-stimulus-interaction is correlated with each constitutive variable. This is true even if sequence and stimulus per se are by design of the experiment independent and uncorrelated.

2. Analysis of Simple and Complex Experiments

A simple experiment investigating our example of CofC effectiveness presents one vignette with fixed features to one group which received the CofC and another group, which did not. The only systematic difference between the randomly constructed groups is the effect of the CofC. Still, as we will show, the effect is nevertheless conditional on this constellation and any conclusion of the experiment should, appropriately, begin with: "Conditional on being presented as the only vignette or first vignette in a list....".

In complex experiments, where each participant receives several vignettes with varying features, several effects affect the participant's decisions:

- Stimulus effects – the presentation of a CofC prohibiting acceptance of gifts should reduce the acceptance of any offer, be it a small gift or a bribe.
- Vignette effects – clearly, two vignettes may differ in many regards, making the attribution of what exactly in the vignette exerts an effect difficult. For the sake of simplicity, we presume the vignettes to be identical apart from the degree to which the behavior described is immoral.
- Sequence effects – there may also be a sequence effect: compared with the open bribe, accepting a small gift seems harmless. The situation of accepting a gift after a bribe was offered differs from accepting a bribe after a gift was offered.
- Interaction effects – the effect of the CofC in a particular decision may depend on the sequence, in which the vignettes are presented. The most simple interaction effect between CofC and sequence is that the subjects forget the CofC during the experiment: the CofC matters most for the first vignette, but progressively less so for all following ones, regardless of the content of the vignette.

The researcher is interested in getting an estimate of the effect of the CofC and there are several options for the experimental design and analysis of the experimental data. As the most basic option, one may compare some kind of average "acceptance of gift/bribe" among experimental and control group across all vignettes.

We will inquire into the options to estimate the effect of the CofC unbiased, and ask how the estimated effect might be affected by taking into account sequence effects, when analyzing the data.

3. Background: Omitted Variable Bias in Linear Models

Generally, for an omitted variable bias to occur, two conditions need to be met: first, the omitted variable must be of some relevance for the dependent variable. Second, the omitted variable must be correlated with other explanatory variables in the model (Dougherty, 2011; Clarke 2005). Applied to the setting of the experiment as described above, these considerations, *prima facie*, make the omitted variable bias problem irrelevant for experimental data: the construction of an experiment can, by random variation of CofC and Sequence, guarantee there is no correlation among both. Thus, there is no reason to expect an omitted variable bias due to excluding Sequence, even if Sequence is relevant for the decision. The effect of the CofC is in this view, not subject to an omitted variable bias. One would thus conclude that whatever effect the Sequence has, it only affects the base rate, but not the effect of the stimulus. The substance of this is that the decisions may differ in their level (for instance, be more ethical) when this particular sequence is used compared to another one. *Prima facie*, holding the Sequence constant or assigning it randomly to the other experimental conditions allows for ignoring Sequence in the analysis without risking any bias.

Unfortunately, this is not true under all circumstances. If the effect of the CofC is conditional on the Sequence, i.e. if there is an effect of the interaction between Sequence and CofC on the dependent variable, excluding this interaction variable, SequenceXCofC, which is, by construction, correlated with both, Sequence and CofC, will lead to a bias. This implies that the interaction term, and the constitutive variable Sequence, must be included.

4. Omitted Variable Bias in Experiments with Sequences

We will use an experiment on CofC effectiveness as described above to illustrate the effect of omitting a relevant variable, even if this variable is seemingly uncorrelated with all other variables. The omitted variable will be the Sequence-variable, indicating that the vignettes or decisions were presented in a particular sequence. The variables of interest are the CofC, as the stimulus, and the Vignette, i.e. the content of the situation described. In our example, we use two vignettes, one, in which a small gift is offered, one, in which something is offered, which is clearly a bribe.

While sequence effects are relevant, the actual interest typically lies in getting an unbiased estimate of the coefficients of the stimulus, and we will investigate whether doing so is possible when the sequence in which the vignettes were presented is omitted from the analysis. When dealing with the problem of the sequence in which the two vignettes are presented, the researcher has several design options. We focus on the two most common:

1. hold the sequence constant for all subjects, i.e. always present one vignette as the first; in our example the gift-vignette shall come first.
2. vary the sequence. For doing this, there are, depending on the number of vignettes, several design options, e.g. invert the sequence using counterbalanced designs, or randomize the sequence. In our simplified example, there are two sequences only: gift-bribe or bribe-gift.

The first question to answer is, How does the sequence affect the dependent variable and how does dealing with the sequence variable in the analysis affect the results of the analysis? In particular the estimated coefficients for those variables the researcher is most interested in, in our case the presence or absence of a CofC?

We assume the following saturated causal model to be the true description of the mechanisms by which a decision is made by a subject:

$$(1)\ Decision_{ij} = a + b_1 CofC + \Sigma\, b_{2k} Sequence_k + \Sigma\, b_{3j} Vignette_j + \Sigma\, b_{4k} CofCXSequence_k + \Sigma\, b_{5jk} Vignette_j XSequence_k + \Sigma\, b_{6j} CofCXVignette_j + e_i$$

where the expression $\Sigma\, b_{5k} Vignette_j XSequence_k$ is shorthand for $b_{51} Vignette_j XSequence_1 + b_{52} Vignette_j XSequence_2 + (\ldots) + b_{5k} Vignette_j XSequence_k$ and equivalent for the other expressions.

And where:

$Decision_{ij}$: Decision in vignette j by subject i, in our case the acceptance of the gift or bribe.

b_1CofC : effect of the CofC (a dummy variable indicating that a CofC is present, 1, or absent, 0).

b_{2k}Sequence$_k$: there are k sequences in use and the variable Sequence$_k$ indicates, which was presented to the participant. The expression b_{2k}Sequence$_k$ indicates the effect of using this sequence rather than another one, on the decision made by the participant. In the case used later on, it shall indicate, whether the gift-vignette was the first presented to the subjects, we assume there are only two sequences, denoted 0 and 1.

b_3Vignette$_j$: the effect of the Vignette j, capturing, whether the vignette formulated strong or weak, in our case there are only two, the gift-vignette and the bribe-vignette, the former is the reference.

e_i : the error term in the usual sense, but which is also covering all individual-level features not included in the study, such as normative orientation of the subject. Random noise and individual-level features may be differentiated by using individual-level information, but this shall not be of interest here.

Of particular interest are the conditional effects, viz.

b_{4k}CofCXSequence$_k$: is the effect of the CofC has when the vignettes were presented in sequence k, compared to the alternative sequences. For instance, the effect of the CofC might be higher for the first sequence, compared to the second (if there are many more vignettes after the CofC was presented, this effect might be stronger). The interaction variable CofCXSequence$_k$ is simply the product of CofC and Sequence$_k$ (with the value of Sequence$_k$ = 1 if sequence k is in use and 0 otherwise).

b_{5jk}Vignette$_j$XSequence$_k$: is the effect of a vignette j when the vignettes were presented in Sequence$_k$. For instance, acceptance in the gift-vignette might be higher than in the bribe-vignette, but it might be lower when the gift vignette was presented first compared to when the bribe-vignette was presented first.

b_{6j}CofCXVignette$_j$: the effect of the CofC in Vignette j – indicating that the CofC as written may be effective for one Vignette, but not for the other. For instance, because the CofC contains a regulation which is relevant for Vignette 1 but no regulation which is relevant for Vignette 2. This effect may also be dependent on the sequence, but we will not further pursue this problem and thus exclude this interaction from further considerations.

While it is in practice be not possible to get estimates of all coefficients in the above equation, the saturated model is nevertheless very helpful to illustrate what components are affecting the dependent variable and what happens, if certain components of the "true" model are omitted in the statistical analysis. In particular, it serves well for illustrating which effects are mixed up into a coefficient which is estimated. The focus of the following arguments will be put on the coefficients involving the CofC, the coefficients b_{3j}Vignette$_j$ and b_{6j}CofCXVignette$_j$ will be ignored.

4.1 Omitted Variable Bias when the Sequence is held constant?

One procedure for conducting the experiment consists of holding the sequence constant. Typically, this option is chosen when the researcher is not interested in the effect of the sequence, but only in the effects of the CofC and the vignettes. As constant factors (and the sequence is a constant factor for all subjects) can be ignored in the analysis of experimental data, the researcher estimates a statistical model using the gift-vignette with no CofC as the reference group, using the following regression model to analyze the data:

(2) $\text{Decision}_{ij} = a + b_1\text{CofC} + b_{3j}\text{Vignette}_j + e_i$

The researcher then reports the intercept a as the base rate of acceptance in the reference group, the effects found for the CofC, b_1, and the bribe-vignette, b_{3j}, concluding that presenting a CofC changes decision behavior by b_1 and that, compared with the gift-vignette, the decision behavior in the bribe-vignette differs by b_{3j}.

We argue and will show that all coefficients reported would only be unbiased, when there is no effect of sequence at all, in particular in the sense that the effect of CofC not depends on the sequence used. All effects found, for CofC and the Vignettes, are strictly conditional on the particular sequence in which the vignettes were presented to the subjects. The researcher did use one and only one sequence, which was constant for all, but nevertheless, there was a particular sequence in use. We will show that according to the "true" reference

model, this sequence has some effect on the decisions made, but also on the magnitude of the effects found for the CofC, and the vignettes. The sequence has an effect, which is however not identifiable, as there is no group to compare the results as obtained for this sequence.

In our experiment with two vignettes, two sequences are possible, let's denote them as Sequence0 an Sequence1. For instance, Sequence1 shall be the sequence where the gift-vignette comes first, and the researcher has chosen this sequence for the experiment. The other sequence, Sequence0, is the one where the bribe-vignette would have come first.

The saturated model, which cannot be estimated but is highly illustrative, allows for all interaction effects involving Sequence and can, for this example, be written as follows:

(3) $$\text{Decision}_{ij} = a + b_1\text{CofC} + b_2\text{Sequence}_k + b_{3j}\text{Vignette}_j + b_{40}\text{CofCXSequence}_0 + b_{50j}\text{Vignette}_j\text{XSequence}_0 + b_{41}\text{CofCXSequence}_1 + b_{51j}\text{Vignette}_j\text{XSequence}_1 + e_i$$

For the case of two sequences, 0 and 1, the interaction-terms in the above equations become 1, i.e. are present, if the two constitutive conditions (CofC present and Sequence k in use) are met. For instance, the interaction-term CofCXSequence_1 is 1, if there is a CofC and Sequence 1 in use. The coefficient b_{41} represents the additional effect of the CofC if Sequence1 is used, while b_1 represents the unconditional effect of presenting a CofC, regardless of the sequence used. Further, if Sequence1 is in use, the variable Squence0 is zero and all terms involving Sequence0 also become zero.

Based on the true model, using the two different sequences results in two different equations:

aa) for Sequence 0

(4) $$\text{Decision}_{ij} = a + b_1\text{CofC} + b_{20}\text{Sequence}_0 + b_{3j}\text{Vignette}_j + b_{40}\text{CofCXSequence}_0 + b_{50j}\text{Vignette}_j\text{XSequence}_0 + e_i$$

which, by replacing Sequence_0 with 1, can be rearranged to

(5) $$\text{Decision}_{ij} = (a + b_{20}\text{Sequence}_0) + (b_1 + b_{40})\text{CofC} + (b_{3j} + b_{50j})\text{Vignette}_j + e_i$$

This is similar but not identical to estimating (2) from above, which was:

(2) $$\text{Decision}_{ij} = a + b_1\text{CofC} + b_3\text{Vignette}_j + e_i$$

Here, b_2, b_{40} and b_{50j} cannot be estimated independent from a, b_1 and b_3 . One gets estimates, but the intercept representing the base rate found is $(a + b_{20}\text{Sequence}_0)$ not a. The same is true for all other coefficients as well: all coefficients – not only the base rate – are conditional on this sequence.

While we will not go into detail here, what is true for the main effects is also true for other interaction effects, like CofCXVignette, which test, whether the CofC is more effective for some vignettes than for others. Part of this effect is also conditional on the sequence: a CofC matters more for this vignette ... but conditional of placing this vignette in this sequence.

4.1.1 For Sequence 1

For the alternative Sequence 1, the model would have been as follows:

(6) $$\text{Decision}_{ij} = a + b_1\text{CofC} + b_{21}\text{Sequence}_1 + b_{3j}\text{Vignette}_j + b_{41}\text{CofCXSequence}_1 + b_{51j}\text{Vignette}_j\text{XSequence}_1 + e_i$$

Which results in

(7) $$\text{Decision}_{ij} = (a + b_{21}\text{Sequence}_1) + (b_1 + b_{41})\text{CofC} + (b_{3j} + b_{51j})\text{Vignette}_j + e_i$$

Again, this is equivalent to estimating equation (2).

(2) $Decision_{ij} = a + b_1 CofC + b_3 Vignette_j + e_i$

Again, just as above, neither b_2, b_{41} nor b_{51} can be estimated in isolation, but it is easy to see that all estimates contain a component which is conditional on the sequence used, viz. $Sequence_1$.

4.1.2 Implications

The implication is that the base rate depends on the sequence chosen, but so do all other coefficients. This is not due to the fact that the explanatory variables, Sequence and CofC, are correlated, and indeed they are not: by construction of the experiment, they are independent. But the effect of one, here CofC, is in its magnitude conditional on the other, here the sequence. As in this example only one sequence in use, the sequence-variable is constant and thus cannot be included in the estimation. Even if the sequence alone were not relevant – the coefficient b_2 capturing the effect of the Sequence were zero – the interaction effect, $b_{40}CofCXSequence_0$, might be not. The crucially omitted variable is not Sequence, but CofCXSequence.

The importance of this finding should not be underestimated. The CofC's effect as estimated using the model specification which excludes any conditionality of effects on sequence is a combination of the true effect, b_1 and the effect b_{41} which is only due to the fact, that the CofC was combined with this particular sequence, viz. Sequence1.

While one is, for reasons of plausibility, tempted to assume that the CofC has an effect per se, i.e. $b_1 \neq 0$, there is no reason and no information in the model allowing us to make this statement for sure. It may well be that the only non-zero-term in the expression $(b_1 + b_{41})CofC$ is b_{41} . This problem is not mended by comparing the groups with and without CofC – the regression coefficient found for the CofC exists, surely. But still, the coefficient of the CofC is $(b_1 + b_{41})$, not b_1 . Ultimately, it is impossible to say that there exists an unconditional effect of CofC, b_1, or only a conditional effect of CofC under this particular sequence, b_{41}.

To summarize, all coefficients obtained in this setting are potentially biased, in that they may depend on a variable, which is inherent to the experiment's design but omitted from the analysis. Sequence, even when held constant, potentially shifts base rates and effects found for explanatory variables, but neither magnitude nor direction of the shifts can be known or at least guessed, unless there is at least another sequence, to compare the results with.

Technically, the bias does not depend on Sequence meeting both conditions of an omitted variable bias, correlation of the omitted variable with an explanatory variable (Sequence and CofC) and relevance of the omitted variable (Sequence) for the dependent variable. The bias arises, because the omitted variable affects the relevance of another variable, an effect modeled as the interaction effect of both (CofCXSequence). The problematic omitted variable is, in our example, not Sequence but the Sequence-CofC-interaction, CofCXSequence. For this variable there is always a correlation between each constitutive variable (here CofC) and if the interaction effect is relevant, there will always be an omitted-variable-bias.

What can be learned in terms of practical advice? When reporting the results of an experiment where a particular sequence of events is used, the sequence is potentially an important factor for how subjects decide. As such, it must be reported, together with the explicit limitation that all findings are strictly conditional on this particular sequence.

4.2 Omitted Variable Bias when Sequence is varied but is omitted from the Analysis

So how can one deal with the problem of conditionality of effects? How to get better, unbiased and *unconditional* estimates for the explanatory variables? Acknowledging that all effects found are fully conditional on the particular sequence used, the researcher may vary the sequence.

Looking at the design option to use several sequences, typical strategies in use when sequences occur in the design of an experiment are randomization and counterbalancing (Solso and MacLin, 2002). In the former design, each participant receives the steps of the experiment in a randomly chosen sequence, and the whole issue of sequence effects is often ignored in the analysis (Dülmer, 2007; Graham and Cable 2001). In the latter design, one subgroup of the experimental group receives the stimuli in one sequence, say ABC, the other in

the inverse sequence, CBA. The sequence is then sometimes taken into account, sometimes not. Interestingly, it takes a lot of effort to screen the experimental literature regarding the issue of sequences is really dealt with, and, to my knowledge, the issue of sequence is presumed to be solved once the sequences are assigned randomly, i.e. most often, sequences are not taken into account by the researcher.

Why would s/he do so?
A first reason to ignore sequences is, again, that because the experimental design is fully under control of the researcher, any correlation among the sequence and the CofC can be avoided by construction. The researcher can guarantee that there is an equal chance of a certain sequence for the experimental group (with CofC) and the control group (no CofC). As there is no correlation among sequence with any other explanatory variable, the sequence-variable, even if relevant, might be omitted, according to the standard procedure when dealing with omitted variable bias problems. However, the above argument showed that the results obtained are biased if and to the degree to which they are conditional on the sequence.

A second, more practical reason to omit the sequence in the analysis is that introducing sequence variables dramatically increases the number of explanatory variables. For N vignettes, there are N! possible sequences, resulting in N!-1 dummy variables in a regression model, which indicate the effect of this particular sequence.

Even worse, it takes either the effort to run separate regressions for all sequences, or another set of interaction effects indicating the conditional effect of, say a CofC, in this particular sequence. In ANOVA-based methods, including sequence dramatically reduces the number of subjects in the N! groups, each of which represents a certain constellation of stimuli and sequence. By ignoring the different sequences, the groups with various sequences are pooled, and one argument is that by pooling the various sequences, their effects may cancel each other out, yielding some kind of average effect for the stimulus, which is a good proxy of the true effect.

What happens if the sequence is varied in the experiment, but then omitted from the analysis? Is anything gained in terms of knowledge about the true effect of the stimulus? Is there an omitted variable bias? What do we learn from the estimated coefficients about the true effects?

Again, we start out from the simplified case of only two sequences, 0 and 1, and the true causal model of the following form:

(8) $Decision_{ij} = a + b_1 CofC + \Sigma\, b_{2k} Sequence_k + \Sigma\, b_{4k} CofCXSequence_k + (...) + e_i$

Where the expression $\Sigma\, b_{4k} CofCXSequence_k$ is again shorthand for $b_{40} CofCXSequence_0 + b_{41} CofCXSequence_1 + (\ldots) + \Sigma\, b_{4k} CofCXSequence_k$

Based on this model, we will show that varying the sequence and keeping it uncorrelated with any other explanatory variable while excluding the sequence variable in the analysis will still not lead to unbiased estimates of coefficients for, say CofC or the base rate.

4.2.1 The Base Rate

Ignoring the sequence will affect the estimated intercept. The typical interpretation of the intercept is that of a base rate expectable if nothing else is known, the "average behavior" expectable in the case for the reference group and nothing else – in our case, the average value of the dependent variable in the gift-vignette where no CofC was presented. If the sequence is excluded from the analysis, the two groups constituted by the two sequences will be merged, leading to a base rate which is the weighted average of those with $Sequence_0$, for which the base rate is $a + b_{20} Sequence_0$, and those with $Sequence_1$, for which the base rate is $a + b_{21} Sequence_1$. So, if there are is a share of p subjects with $Sequence_0$ and a share of 1-p subjects with $Sequence_1$, the estimated base rate is

(9) base rate $= p\,(a + b_{20} Sequence_0) + (1-p)\,(a + b_{21} Sequence_1)$
$= a + p\, b_{20} Sequence_0 + (1-p)\, b_{21} Sequence_1$

While the interpretation of the intercept as an "average behavior" is true, the contribution of each sequence to this average remains unknown, as a, b_{20}, and b_{21} are unknown. The base rate under $Sequence_0$ may by completely different from that under $Sequence_1$, it may also be quite similar to the latter, or there may be no difference at all.

So, unless a dummy capturing the alternative sequence is included, one cannot tell anything meaningful about the base rates, most notably nothing about whether they depend on the sequence or not. Once there are more than two sequences, the problems of assigning a meaningful interpretation to the base rate increase.

This matters for the interpretation of the base rate, because it makes a difference to say, the base rate of accepting a gift is X, or to say, the base rate of accepting a gift is X', but after being offered a clearly recognizable bribe it is only Y.

4.2.2 Coefficients of Explanatory Variables

What is the effect of omitting the Sequence variable when estimating the coefficient for CofC in a situation, where different sequences were used? The true model for k different sequences shall be:

(10) $Decision_{ij} = a + b_1CofC + \Sigma\, b_{2k}Sequence_k + \Sigma\, b_{4k}CofCXSequence_k + (...) + e_i$

which, in the case of two Sequences, 0 and 1, is reduced to

(11) $Decision_{ij} = a + b_1CofC + b_{20}Sequence_0 + b_{21}Sequence_1 + b_{40}CofCXSequence_0 + b_{41}CofCXSequence_1 + (\ldots) + e_i$

As the sequence is, per construction of the experiment, uncorrelated with the CofC, it might be excluded, if its effect were restricted to the Sequence-variable itself. Including it would yield information about the sequence effect, in the sense that there is information about the effect of using Sequence 1 instead of Sequence 0, which constitutes the reference category. But in the linear model, this effect is additive, shifting the level of the dependent variable up or down by $b_{2k}Sequence_k$ as was shown in the preceding section. For instance, if there were just two sequences, the coefficient $b_{21}Sequence_1$ indicates whether people are more likely to reject a gift, if they were e.g. offered a bribe first.

The problem for estimating the effect of CofC arises again from the interaction effect between CofC and Sequence, which captures whether the effect of a CofC is to some degree conditional on the sequence used.

As we argued above, a CofC might be much more effective in one sequence than in another one. The omitted variable is the Sequence-CofC-interaction term, not Sequence. Applied to two sequences, 0 and 1, we focus on the following part of the expression given in (11):

$$(\ldots) + b_1CofC + (\ldots)\ b_{40}CofCXSequence_0 + b_{41}CofCXSequence_1 + (\ldots)$$

Properly taking into account the sequence variable would be done by constructing interaction terms, i.e. $CofCXSequence_0$ or $CofCXSequence_1$. As both are perfectly collinear, they cannot be estimated individually, so the reference is not the effect of the unconditional effect of the CofC but the effect of the CofC conditional on the sequence serving as the reference. For instance, when Sequence0 is used as the reference, the effect of the CofC in the reference group would be $(b_1 + b_{40})CofC$. Again, little can be said about the true effect of the CofC, b_1, as it is fully conditional on this particular sequence.

In the case of two sequences, 0 and 1, excluding the information on the sequence implies that the effects for the two groups, those with $Sequence_0$ and those with sequence 1, will be merged, or rather, averaged. For the group with $Sequence_0$, the coefficient of the CofC is

(12) $b_{CofC}\ CofC = b_1CofC + b_{40}CofCXSequence_0$

which, as $Sequence_0$ has the value 1 for those for which the sequence 0 was used, reduces to

(13) $b_{CofC}\ CofC = (b_1 + b_{40})\ CofC$

As argued above, the coefficient obtained for the CofC is a composite, and it is impossible to say whether there exists a true, unconditional effect of the CofC, b_1, or only the conditional effect of b_{40} . For the group with $Sequence_1$, the complementary coefficient for the CofC is b_{41} from the expression $b_{41}CofCXSequence_1$, resulting in

(14) $b_{CofC}\ CofC = (b_1 + b_{41})\ CofC$

If the interaction effects are ignored and just the effect of the CofC, b_{CofC}, is estimated, the resulting coefficient for CofC, is the weighted average of the effect in the $Sequence_0$ group and the effect in the $Sequence_1$ group.

So, if there are p subjects with $Sequence_0$ and q=1 – p subjects with $Sequence_1$, b_{CofC} , the estimated coefficient of CofC, is the result of the following components:

(15) $b_{CofC}\ CofC = (\ p\ (\ b_1 + b_{40}\ ) + (1–p)\ (b_1 + b_{41})\ )\ CofC$

Where b_1 is the unconditional effect of CofC, b_{40} the effect of CofC conditional on using $Sequence_0$ and b_{41} the effect of CofC conditional on using $Sequence_1$.

So the coefficient estimated, b_{CofC} , is not the true coefficient of CofC, but a composite coefficient, where the constituting effects , b_1 , b_{40} and b_{41} , are all unknown in their relative magnitude. It is not possible to say whether the effect of the CofC is identical across all sequences, i.e., b_{40} =0 and b_{41} =0, or whether it is fully conditional on the sequence in use (for instance b_1 =0 and b_{40} =0).

4.2.3 Implications and an Evaluation of Standard Techniques dealing with Sequence Effects

When does using several sequences improve the results and what can we tell about the effect of a stimulus, say, a CofC?

If there are only two “events” in the experiment, randomization and counterbalancing improve the insight, albeit not because of the randomization or the counterbalancing per se, but because all possible sequences are used.

If there are more than two events, and, consequentially, many more possible sequences, counterbalancing constitutes a case of selecting two sequences from the universe of all possible sequences. In the case of randomization, one uses a random selection from the universe of all possible sequences.

The implication of the above arguments is that unless sequence is also included in the analysis, there is no insight gained from using different sequences. By using several sequences, regardless of the motivation underlying the choice of these particular sequences, one gets an average effect for CofC, but is unable to tell, whether this average effect comes about because the CofC is of equal effectiveness in all sequences or only effective in one particular sequence.

Instead, implicit assumptions stand in. For instance, in the case of counterbalancing one implicitly assumes that whatever conditionality occurs, it is lowest in one sequence and highest in the other, inverse sequence, and further, that the effect of the sequence changes somehow “linear” across all sequences, e.g., it increases linear in steps of equal magnitude when moving from the sequence ABC to the sequence CBA. When making a prediction for the effectiveness of a CofC for sequence which was not included in the design, say, BAC, one assumes an effect to exist, because one assumes that there is a true and unconditional effect, which somehow emerges from pooling the sequences.

There is however no guarantee for this to be the case. When using counterbalancing or randomization, the assumption that the coefficient obtained is “some kind of average effect” is plausible but unwarranted. True, the coefficient obtained for an experimental stimulus when the sequence and the interaction terms are excluded is a composite. But there is no basis on which to say that an effect actually occurs in all sequences. It is also possible that the observed “average effect” is due to a strong effect in just one sequence, while there is no effect in all others. Lacking information on the effect in different sequences, on simply cannot tell.

5. A Simulation

To get an impression on the practical relevance of the problem, we simulated our example experiment on Codes of Conduct as a driver of ethical behavior, but with three vignettes in use. The experimental set-up is as follows: first, participants in the experimental group receive a CofC, then all subjects receive three vignettes, A, B and C, in six different sequences, ABC to CBA. Each participant's decision is recorded, represented by three entries in the data set: the response to vignette A (the reference vignette) and the responses in vignettes B and C, which are dummy coded. The dependent variable shall be the likelihood of accepting of a gift of dubious nature. Sequence and CofC randomized, i.e. all combinations occur with identical frequency.

To create the dependent variable, we chose a "true" mechanism along the lines of our illustrative example.

(16) $\text{Decision} = 4 - \text{CofC} + \text{VignetteB} + 2 \cdot \text{VignetteC} - 2 \cdot \text{CofC} \cdot \text{SeqUneven} + e_i$

The error-term is distributed $e_i \sim N(0,1)$. The CofC is a dummy variable indicating whether the CofC was presented to the subject, and the assumption is, that presenting a CofC decreases the dependent variable, i.e. reduces the likelihood that the gift is accepted. The conditionality of the CofC effect on the sequence was modeled in that the CofC is much more relevant in sequences with an uneven number, i.e. in sequences 1, 3 and 5. For them, the dummy variable SeqUneven is 1.We ignore the substantive meaning of the mechanisms, but are simply interested in how choosing some sequences from the set of all possible sequences affects the estimates for the coefficients of CofC as the experimental manipulation. Based on this data set, we simulated what coefficients result, if only selected sequences are used. Which is to say, show which results are obtained when typical designs for sequence problems, like holding constant, randomization, or counterbalancing, are used.

Table 1 gives the results of using all sequences, as one would if the sequences were randomly assigned to the subjects, but ignoring the sequence and the potential conditionality of effects on the sequence we built into the data when analyzing the data.

Table 1: All Sequences used, but Sequence ignored or mis-specified

Variable	O0	O1	O2	O3
CofC	-1.240	-2.187	-2.187	-2.187
VignetteB	0.934		0.934	0.934
VignetteC	1.878		1.878	1.878
CofCXSeqUneven	-1.895			
SEQ2				0.507
SEQ3				-0.008
SEQ4				1.192
SEQ5				0.132
SEQ6				0.830
Intercept	4.217	5.154	4.217	3.775

Column O0 gives the correctly specified model, which results in the true equation underlying the data. O1 to O3 are other models which either ignore sequence or included it in a mis-specified way. We see that O1 to O3 would all indicate an effect of a CofC which is twice its real magnitude, and would moreover indicate that the sequence variable can be ignored, as it is uncorrelated with the CofC and the inclusion of it does not affect the coefficient obtained for the CofC.

Table 2 gives the results obtained when only one sequence was used, ABC or CBA, which would be the strategy of holding sequence constant. Or the two "extreme" cases, ABC and CBA, are used, as would be the case when counterbalancing was chosen.

We see that the different sequences yield different coefficients for the CofC, just as we would expect, there are two types of sequences, one where the CofC matters strongly, see sequence 1 in column OSq1, one, where

it does not, see sequence 6 in column OSq6. Using the counterbalanced design using sequences 1 and 6, see column OSq16, also yields coefficients, which are way off the truth.

Table 2: Holding Sequence Constant or Using Counterbalanced Design

Variable	OSq1	OSq6	OSq16
CofC	-3.280	-1.391	-2.336
VignetteB	1.577	1.557	1.567
VignetteC	2.138	2.143	2.140
Intercept	4.020	3.911	3.966

What is the range of expectable coefficients? With 6 sequences, there are 15 possible different pairs of sequences, and we ran an analysis for each, again ignoring sequence as a variable, see Table 3.

Table 3: Distribution of Estimates when using 2 out of 15 Sequences-Pairs

Variable	O12	O13	O14	O15	O16	O23	O24	O25	O26	O34	O35	O36	O45	O46	O56
CofC	-2.099	-3.390	-2.126	-3.171	-2.336	-2.209	-0.945	-1.991	-1.155	-2.236	-3.281	-2.446	-2.017	-1.181	-2.227
VignetteB	1.034	1.144	1.218	0.993	1.567	0.601	0.675	0.450	1.024	0.786	0.561	1.134	0.635	1.208	0.983
VignetteC	1.875	2.000	2.186	1.708	2.140	1.737	1.924	1.445	1.878	2.048	1.570	2.002	1.756	2.189	1.710
Intercept	3.952	4.262	4.142	4.370	3.966	4.193	4.074	4.301	3.897	4.384	4.611	4.207	4.492	4.087	4.315

If only two sequences are used, the average coefficient obtained for CofC is -2.19, which is to say, in the case of many different studies, which use the same design, the findings regarding CofC effectiveness will converge, but they will not converge to the true coefficient.

6. Conclusion and Recommendations: Dealing with Sequence Effects in Complex Experiments

Starting out from the general background of an omitted variable bias, we demonstrated that if the effectiveness of a stimulus is dependent on the sequence, there is an omitted variable bias, even if the experimental design guaranteed that sequence and stimulus are perfectly independent. The problem arises not from the sequence variable per se, but from the sequence-stimulus-interaction variable, which captures that a stimulus is more relevant in one sequence than in another one.

We discussed the typical strategies of dealing with this issue, when designing and analyzing experiments, viz. holding sequence constant, using counterbalancing designs, or randomization of sequences.

Looking at the design option of holding sequence constant by using only one sequence, it is easy to show that all findings are fully dependent on the sequence chosen. It is not possible to make a statement about the effect of a stimulus, which is unconditional, and this should be acknowledged in the limitations.

Often, different sequences are used in the experiment, using counterbalancing or randomization, but later on, sequence is neither used as an explanatory variable in the analysis nor are the coefficients tested for conditionality on sequence. In the case that more than one sequence is used, the resulting coefficients are composites, weighted averages of coefficients across all sequences occurring. Using a sequence and its inverse is one typical form of counterbalancing. It implicitly assumes that the two sequences represent the extreme cases, e.g., a very strong and a very weak coefficient of the stimulus. There is no guarantee that the strength of a coefficient under a particular sequence is comparable to the strength in other sequences. This is an assumption, nothing more. In the case of randomization, the situation is similar in that coefficients resulting from different sequences are pooled by estimating a reduced, simpler model, which excludes all effects a sequence might have.

Unless different sequences are consciously used in both, experiment and analysis, we know little about the range and occurrence of strong or weak effects in different sequences. The analysis should either include stimulus-sequence-interaction terms, or run the statistical models using subsets of cases with the same sequence. If the interactions come up as insignificant or the coefficients obtained for different subsets are stable, there is no indication of a conditionality, and sequences can be ignored. In a nutshell: using different sequences while not including the sequences in the statistical analysis does not improve our knowledge about whether the observed effects of a stimulus are real, or design-dependent.

In practice, this leads to a dilemma: On the one hand, even if several sequences are used, varying the sequences does not add any value in terms of identifying effect of the stimulus, unless the sequence is also used in the analysis. From the arguments made, it is clear, what the recommendation is: vary sequences and use all available information in the analysis, including sequence, which allows at least an insight into whether the coefficients vary or are comparable across sequences.

On the other hand, doing so may be difficult, or even impossible, because actively using sequences in both the experiment and the analysis may dramatically reduce the number of cases for a certain constellation of explanatory factors. It is oftentimes simply not possible to get enough cases to make statements based on a reasonable number of cases. Further, if there are conditionalities, interpreting them may be highly difficult and not be the actual interest of the researchers. There is no easy way out.

One solution to avoid problems with sequence-effects is, ultimately, to avoid sequences, i.e. to choose between-designs in which all participants are given only one vignette each. This would imply, for the practice of experimental research that the scope of experiments is to be reduced. Instead of using “omnibus experiments”, with many vignettes, one should use only one or two vignettes, to limit the number of possible sequences. This has the additional advantage that in real life, decision situations are rarely occurring in a rapid sequence, as they do in an experiment with many vignettes, a fact which increases external validity.

Another solution would be to use some sequences, but only to the degree, that one can also explicitly include them in the statistical analysis. Here, one is at least able to gauge the magnitude of the problem. This might apply to themes where many decisions succeed each other in a rapid sequence, for instance in performance evaluation.

By way of an outlook, it would be an interesting research effort to re-analyze data from experimental studies using several sequences in order to get a notion on the magnitude of the problem. Up to now, there is no information on how “unconditional” effects found in experiment really are. Given the replication crisis such a research effort might help to identify reasons why some findings cannot be replicated. Further, in terms of theory building, finding sequencing effects and in particular finding conditionalities of stimulus effects on sequences may open up new strains of research, which inform decision-making research.

References

Aguinis, H. and Bradley, K. J., 2014. Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational Research Method*s, 17(4) pp. 351-371.

Auspurg, K. and Hinz, T., 2015. *Factorial Survey Experiments*. Los Angeles.: Sage.

Auspurg, K., Hinz, T. and Liebig, S., 2009. *Complexity, learning effects, and plausibility of vignettes in factorial surveys*. Paper accepted for the ASA-Conference 2009. Working Paper # 4 of the DFG-Project "The Factorial Survey as a Method for Measuring Attitudes in Population Surveys".

Auspurg, K. and Jäckle, A., 2017. First equals most important? Order effects in vignette-based measurement. *Sociological Methods & Research*, 46(3), pp. 490-539.

Benartzi, S. and Thaler, R. H., 1999. Risk Aversion or myopia? Choices in repeated gambles and retirement investments. Management Science, 45(3), pp. 364-381.

Campbell, D. T. and Stanley, J. C., 1963. *Experimental and quasi-experimental designs for research*. Boston.: Houghton Mifflin.

Clarke, K. A., 2005. The Phantom Menace: omitted variable bias in econometric research. *Conflict Management and Peace Science*, 22(4), pp. 341–352.

Dal Bo, P. and Frechette, G. R., 2011. The evolution of cooperation in infinitely repeated games: experimental evidence. *American Economic Review*, 101(1), pp. 411-429.

Dougherty, C., 2011. *Introduction to Econometrics*. Oxford: Oxford University Press.

Dülmer, H., 2007. Experimental plans in factorial surveys. Random or quota design? *Sociological Methods & Research* 35(3), pp. 382-409.

Graham, M. E. and Cable, D. M., 2001. Consideration of the incomplete block design for policy-capturing research. *Organizational Research Methods*, 4(1), pp. 26-45.

Hogarth, R. M. and Einhorn, H. J., 1992. Order effects in belief updating: the belief-adjustment model. *Cognitive Psychology*, 24(1), pp. 1–55.

Hox, J., Kreft, I.G.G. and Hermkens, P.L.J., 1991. The analysis of factorial surveys. *Sociological Methods & Research*, 19(4), pp. 493-510.

Johnson-Laird, P. N. and Shafir, E., 1993. The interaction between reasoning and decision making: an introduction. *Cognition*, 49(1-2), pp. 1-9.

Kirk, R.E., 2012. *Experimental design: procedures for the behavioral sciences.* Thousand Oaks: Sage.

Klos, A., Weber, E. U. and Weber, M., 2005. Investment decisions and time horizon: risk perception and risk behavior in repeated gambles. *Management Science*, 51(12), pp. 1777-1790.

Krosnick, J. A. and Alwin, D. F., 1987. An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly*, 51(2), pp. 201-219.

Libby, R., Bloomfield, R. and Nelson, M. W., 2002. Experimental research in financial accounting. *Accounting, Organizations and Society*, 27(8), pp. 775–810.

Mitchell, M. L. and Jolley, J. M., 2010. *Research design explained*. Belmont, CA: Wadsworth.

Montgomery, D.C., 2012. *Design and analysis of experiments*, Hoboken NJ.: Wiley.

Oll, J., Hahn, R., Reimsbach, D. and Kotzian, P., 2018. Tackling complexity in business and society research: the methodological and thematic potential of factorial surveys. *Business & Society*, 57(1), pp. 26–59.

Rossi, P. H. and Anderson, A. B., 1982. The factorial survey design; an introduction. In P. H. Rossi and S. L. Nock (eds.) *Measuring social judgments*. *The factorial survey approach* (pp. 15-67). London : Sage.

Solso, R. L. and MacLin, M. K., 2002. *Experimental psychology: a case approach*. Boston : Allyn&Bacon.

Stöber, T., Kotzian, P., Weißenberger, B. E., and Hoos, F., 2016. Effective, but not all the time. Experimental evidence on the effectiveness of codes of ethics. SSRN Working Paper Series.

Strack F. 1992 "Order effects" in survey research: activation and information functions of preceding questions. In: N. Schwarz N. and S. Sudman (eds.) *Context effects in social and psychological research* (pp. 23-34). New York: Springer.

Tourangeau, R., Rips, L. J. and Rasinski, K., 2000. *The psychology of survey response*. Cambridge: Cambridge University Press.

Wiegmann, A., Okan, Y. and Nagel, J. (2012). Order effects in moral judgment. *Philosophical Psychology,* 25(6), pp. 813-836.

Zeelenberg, R. and Pecher, D., 2015. A method for simultaneously counterbalancing condition order and assignment of stimulus materials to conditions. *Behavior Research Methods*, 47(1), pp. 127–133.