



# Applying Linear Mixed Effects Models (LMMs) in Within-Participant Designs With Subjective Trial-Based Assessments of Awareness—a Caveat

Guido Hesselmann\*

Visual Perception Laboratory, Department of Psychiatry and Psychotherapy, Charité—Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Berlin Institute of Health, Berlin, Germany

**Keywords:** *post hoc* trial sorting, anova, linear mixed models, unbalanced data, visibility, consciousness, awareness

## OPEN ACCESS

### Edited by:

Samuel Evans,  
University of Westminster,  
United Kingdom

### Reviewed by:

David Amooti Magezi,  
Universität Hamburg, Germany  
Miguel A. Vadillo,  
Universidad Autonoma de Madrid,  
Spain

### \*Correspondence:

Guido Hesselmann  
g.hesselmann@gmail.com

### Specialty section:

This article was submitted to  
Consciousness Research,  
a section of the journal  
Frontiers in Psychology

**Received:** 03 January 2018

**Accepted:** 03 May 2018

**Published:** 25 May 2018

### Citation:

Hesselmann G (2018) Applying Linear Mixed Effects Models (LMMs) in Within-Participant Designs With Subjective Trial-Based Assessments of Awareness—a Caveat. *Front. Psychol.* 9:788. doi: 10.3389/fpsyg.2018.00788

## INTRODUCTION

In experimental psychology, including consciousness research, within-participant designs are typically more powerful than between-participant designs. During the last decades or so, the most widely used statistical method to analyze data from within-participant designs has been the repeated-measures (rm-) ANOVA. In recent years, however, empirical studies have increasingly turned toward using linear mixed effects models (LMM) to analyze data from within-participant designs (Baayen et al., 2008; Magezi, 2015).

LMMs are sometimes preferred over rm-ANOVA for a single practical reason, namely their ability to deal with unbalanced and incomplete data sets. For the sake of illustration, I will briefly describe a simple psychophysical experiment. Participants are instructed to respond as quickly and as accurately as possible to some predefined feature of a visual target (e.g., its semantic category). Response times (RTs) are recorded as the dependent variable. Immediately before the presentation of the target, another visual stimulus is presented. In the following, I will refer to this stimulus as the cue stimulus. The cue stimulus is presented at the threshold (liminally) or near the threshold of perception (peri-liminally), so that participants sometimes see the stimulus, and sometimes miss it. Liminal presentation can be achieved by low visual contrast, brief presentation times, and various psychophysical suppression methods (Breitmeyer, 2015). On each trial, participants first provide the speeded response to the target stimulus, and then rate the subjective visibility of the liminal cue stimulus using a binary scale (“seen,” “not seen”). Alternatively, a continuous subjective visibility scale such as the perceptual awareness scale (PAS) can be binarized by selecting and/or pooling across ratings.

A classical research question would be whether some features of the cue facilitate the response to the target (i.e., priming). Based on the subjective trial-based assessment of awareness, or *post hoc* trial sorting, priming can be compared when the cue is not consciously perceived relative to when it is (Van den Bussche et al., 2013; Avneon and Lamy, 2018). Another research question could be whether the visibility of the cue stimulus itself affects the RTs to the target stimulus. Due to cognitive costs associated with seeing the cue, responses to targets preceded by “seen” cue stimuli may be slower or less accurate than the responses to targets preceded by “unseen” cue stimuli.

In these examples, one commonly used statistical approach is to first sort trials into “seen” and “not seen” trials, and then average and submit the data to rm-ANOVA<sup>1</sup>.

In the absence of individual adjustment of stimulation parameters, it is a realistic scenario that the distribution of visibility ratings will be different for individual participants. In the extreme case, some participants might only rarely, or even never provide “seen” ratings, while other participants mostly provide “seen” ratings and therefore hardly any or no “not seen” ratings. Since the rm-ANOVA is based on the mean RTs per condition, participants with no “seen” trials or no “not seen” trials will be removed from the statistical analysis. Some researchers might also be reluctant to use mean RTs based on <10 trials or so. In this situation, and primarily for the pragmatic reason to use most of the obtained experimental data, the researcher might decide to calculate LMMs, which can accommodate missing data. Before introducing the data simulation, the next paragraph provides a brief overview of the assumptions for rm-ANOVA and LMMs.

## ASSUMPTIONS OF RM-ANOVA AND LMMs

Rm-ANOVA and LMMs are extensions of linear regression, and a number of assumptions are therefore common to both methods (Field, 2012). Among these, the absence of correlation with external variables is most important for the current data report. Beyond the common assumptions, rm-ANOVA additionally requires compound symmetry and complete data (Magezi, 2015). Compound symmetry means that the variances as well as the covariances of the repeated measures are similar (or, homogeneous). If the stringent assumption of compound symmetry is violated, then the sphericity assumption is still a necessary and sufficient condition for the F-tests to be valid. Sphericity means that the variances of the difference scores (between the levels of the repeated factor) are similar. The assumption of sphericity is often violated in experimental psychology, which may increase Type I errors, but this violation can be accounted for by correcting the degrees of freedom (e.g., using the Greenhouse-Geisser correction). Please note that in the current data simulation, the sphericity assumption is met because the repeated factor has only two levels (“seen” and “not seen”; see data simulation below). Follow-up simulations employing graded levels of visibility will need to consider sphericity when calculating rm-ANOVA. Complete data, the second additional assumption of rm-ANOVA, means that, for each participant, measurements must be available for all levels of the repeated factor. In contrast to rm-ANOVA, LMMs do not depend on assumptions about the variance-covariance matrix, and LMMs can accommodate missing data (Magezi, 2015).

## SIMULATED DATA SETS (METHODS)

In this data report, unbalanced within-participant data sets were generated in a numerical simulation, in order to compare the

<sup>1</sup>Please note that statistical test including a factor that was not under experimental control are generally considered exploratory, not confirmatory (de Groot, 2014).

statistical outcomes of rm-ANOVA and LMMs. The R code for the data simulation is available at the Center for Open Science (OSF; <https://osf.io/d7y8h/>).

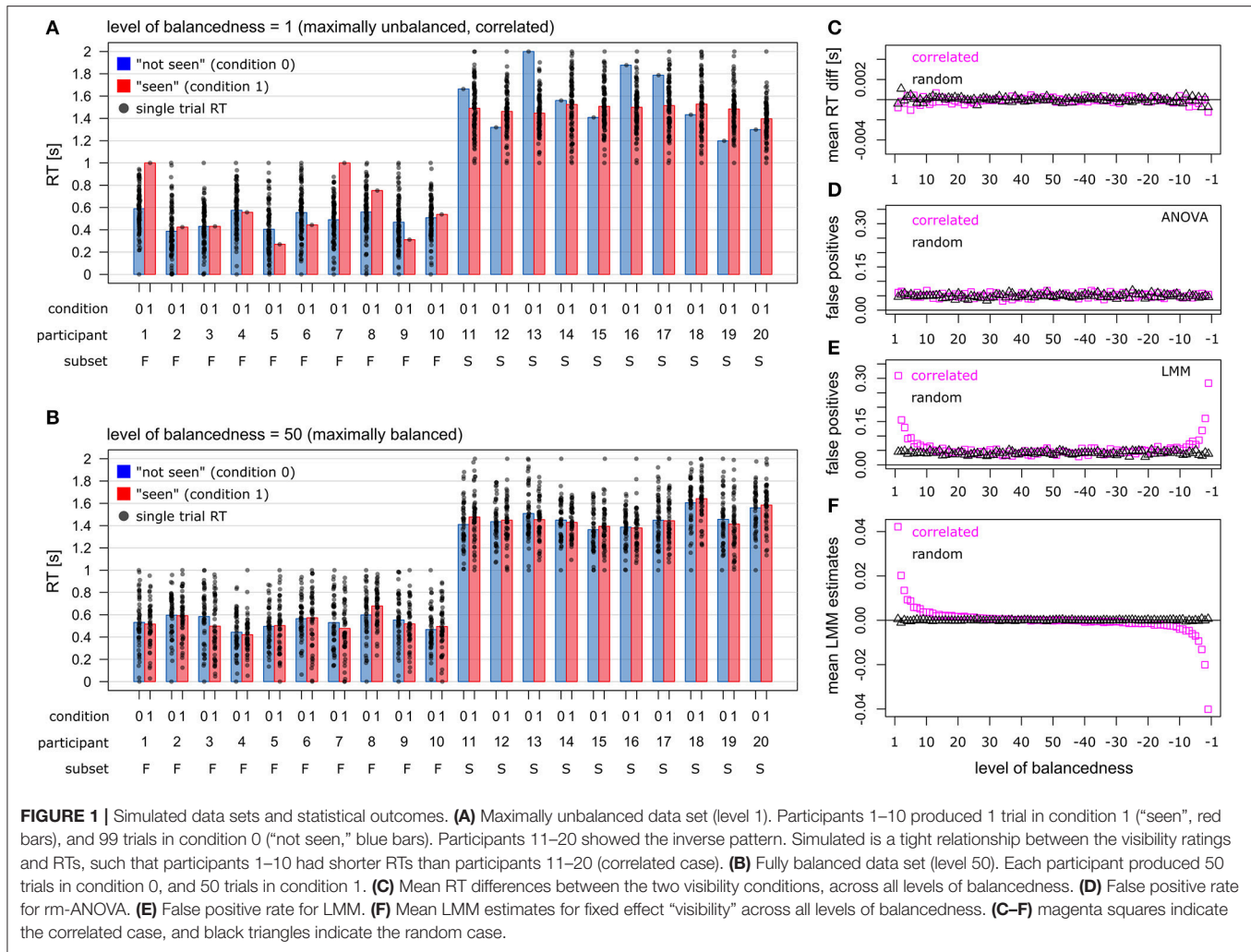
In total, 100 trials per participant ( $N = 50$ ) were generated. The level of balancedness ranged from 1 to 50. At level 1, one trial was rated as “seen,” while 99 trials were rated as “not seen” (Figure 1A). At level 50, 50 trials were rated as “seen,” and 50 trials were rated as “not seen,” thus indicating maximal balancedness (Figure 1B). Negative levels of balancedness indicate the inverse pattern, such that at level  $-1$  one trial was rated as “not seen,” while 99 trials were rated as “seen.” Accordingly, levels 1 and  $-1$  indicate the lowest level of balancedness.

Figures 1A,B illustrates that two subsets of participants were simulated: participants with short RTs (subset “Fast”, F), and participants with long RTs (subset “Slow”, S). Subset “Fast” ( $N = 25$ ) had RTs ranging from 0s to 1s (mean 0.5s), while subset “Slow” ( $N = 25$ ) had RTs ranging from 1s to 2s (mean 1.5s). Both RT distributions were based on the *norm* function in R (mean = 1, standard deviation = 1), and rescaled to the range 0–1s. For subset “Slow”, 1s was added to each RT data point. Note that for the sake of illustration, Figures 1A,B plots the data from only 20 participants (subset “Fast”: 1–10, subset “Slow”: 11–20).

Two different relationships between the participants’ subset membership and the pattern of subjective visibility ratings were simulated. In the “correlated” case, a tight relationship was simulated (Figure 1A). For example, at balancedness level 1 in subset “Fast,” one trial was rated as “seen,” and 99 trials were rated as “not seen.” Subset “Slow” showed the inverse rating pattern so that at level 1 one trial was rated as “not seen,” while 99 trials were rated as “seen.” This scenario is not implausible, as it might be the case that participants with shorter RTs (subset “Fast”) consciously perceive the liminal cue stimulus less often than participants with longer RTs (subset “Slow”). For the sake of consistency, we simulated the inverse pattern for each level of balancedness: at levels 1–49, subset “Fast” was the subset with short RTs, while at levels  $-1$  to  $-49$  the same subset showed long RTs. In the “random” case, participants were randomly assigned to the subsets so that there was no relationship between the participants’ subset membership and the pattern of subjective visibility ratings (data not shown). In both cases, 1,000 data sets were generated for each level of balancedness. Crucially, no effect of factor “visibility” was simulated, neither in the correlated case nor in the random case.

## STATISTICAL OUTCOMES (RESULTS)

Data analysis focused on the relative frequency of significant ( $p < 0.05$ ) tests of the two-level factor “visibility” (i.e., false positive rate), as resulting from the 1,000 rm-ANOVAs and 1,000 LMMs at each level of balancedness. The rm-ANOVA was calculated using the *aov\_car()* function of the *afex* package in R. Type III sums of squares were used, as these are default in many commercially available statistical packages (e.g., SPSS). The LMM was calculated using the *lmerTest* package in R, with Satterthwaite’s approximation of degrees of freedom. The LMM



included the two-level factor "visibility" (cond) as fixed factor as well as random slope [ $RT \sim (1+cond|subj) + cond$ ]. Note that "subset" was not part of the LMMs, as the relationship between the participants' subset membership and the subjective visibility ratings was assumed to be unknown to the researcher prior to the experiment.

**Figure 1C** plots the mean RT difference between the two visibility conditions across all levels of balancedness, separately for the "correlated" case and the random case. "Mean RT" refers here to the global mean across the 1,000 mean RT differences from the simulated data sets. As expected, the mean RTs vary around zero, since no effect of factor "visibility" was simulated.

**Figure 1D** plots the false positive rate for the rm-ANOVA. For each level of balancedness, the false positive rate is close to 0.05. **Figure 1E** plots the false positive rate for LMMs. When the level of balancedness is approx. below 10, the false positive rate begins to exceed 0.05. For maximally unbalanced data, the false positive rate is about 30%. However, this increase of false positives is only observed in the correlated case.

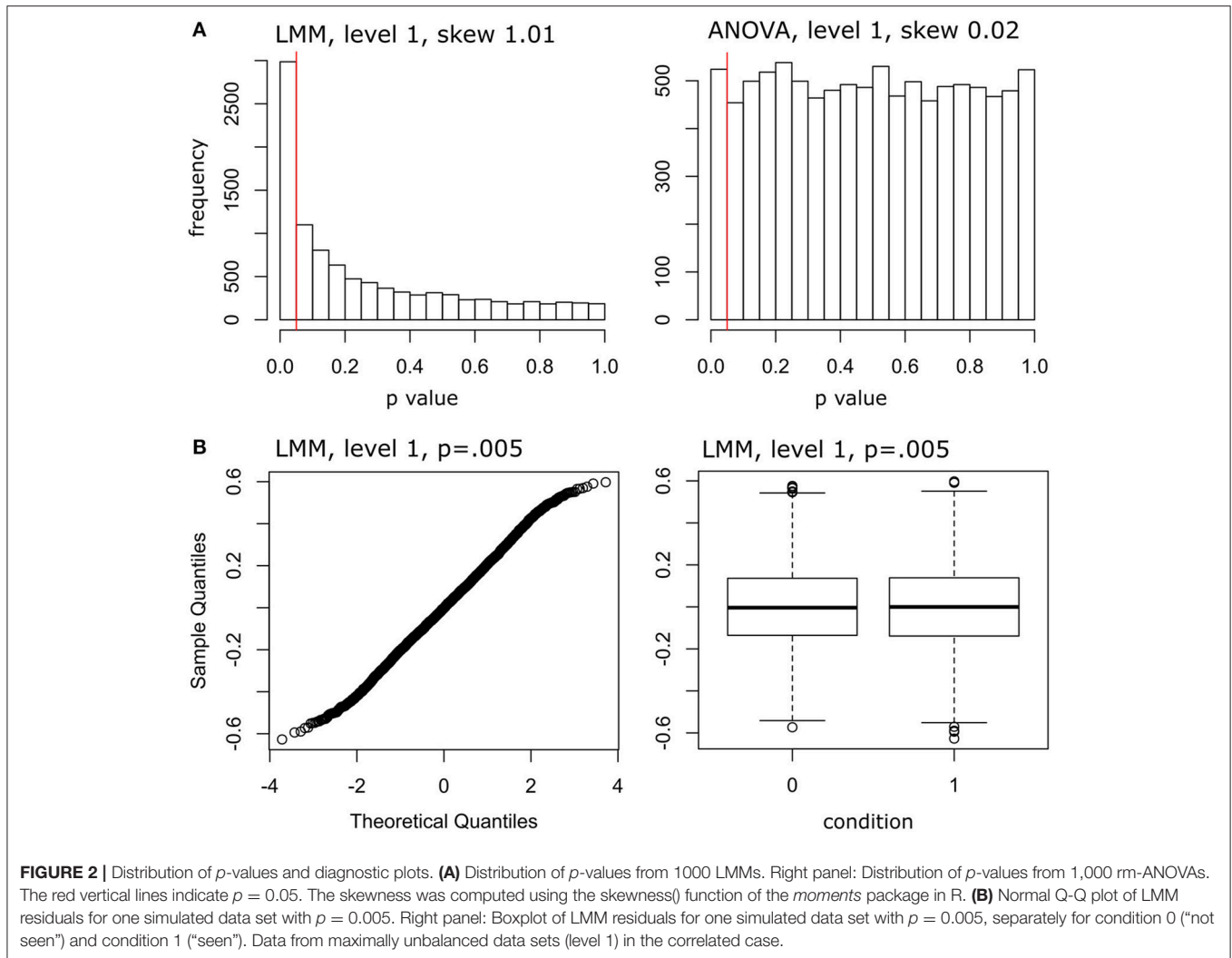
**Figure 1F** plots the mean LMM estimates for the fixed effect "visibility" across all levels of balancedness. In

the random case, the estimates vary around zero. In the correlated case, the estimates deviate from zero for massively unbalanced data sets (i.e., toward the left/right limits of the x-axis), in agreement with the increased false positive rate.

**Figure 2** provides a number of diagnostic plots. **Figure 2A** plots the p-value distributions of rm-ANOVA and LMM, when the data are maximally unbalanced (balancedness level 1, correlated case). As can be seen, the distribution is heavily skewed in the case of LMM. **Figure 2B** plots the residuals for one simulated data set with  $p = 0.005$  (balancedness level 1, correlated case). The figure suggests that a violation of the normality assumption cannot explain the observed increase in false positives for LMMs.

## CONCLUSION

LMMs yielded strikingly more false positives than rm-ANOVA in the case of massively unbalanced within-participant data



sets, when a previously unknown grouping factor remained unaccounted for. In the specific example used in this data report, the LMMs frequently indicated an effect of factor “visibility,” while only a RT difference between two subsets of participants was present in the simulated data. One solution to this problem would be to include the rating behavior of each participant in the linear model (e.g., the ratio between “seen” and “not seen” trials). The behavior of LMMs in more realistic and complex scenarios (e.g., including participants with missing data in one visibility condition) awaits further investigation. Follow-up data simulations should also investigate the rate of false negatives in similar scenarios.

While LMMs are more flexible than rm-ANOVA, and therefore are becoming increasingly popular in experimental psychology, researchers using LMMs for unbalanced data sets should be aware of the caveat described in this data report. It is generally advisable to carefully visualize the data prior to analysis, as well as to consult diagnostic plots when using LMMs. Finally, if

a data set turns out to be massively unbalanced such that subsets of participants have only few trials at a specific factor level, or that participants show very different patterns of visibility ratings, the experimental design per se might have to be reconsidered.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

## FUNDING

GH was supported by the German Research Foundation (grant HE 6244/1-2).

## ACKNOWLEDGMENTS

I would like to thank Olga Shurygina for her invaluable support in preparing the R script.

## REFERENCES

- Avneon, M., and Lamy, D. (2018). Reexamining unconscious response priming: A liminal-prime paradigm. *Conscious. Cogn.* 59, 87–103. doi: 10.1016/j.concog.2017.12.006.
- Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005
- Breitmeyer, B. G. (2015). Psychophysical “blinding” methods reveal a functional hierarchy of unconscious visual processing. *Conscious. Cogn.* 35, 234–250. doi: 10.1016/j.concog.2015.01.012
- de Groot, A. D. (2014). The meaning of “significance” for different types of research [translated and annotated by Eric-Jan Wagenmakers, Denny Borsboom, Josine Verhagen, Rogier Kievit, Marjan Bakker, Angélique Cramer, Dora Matzke, Don Mellenbergh, and Han L. J. van der Maas]. *Acta Psychol.* 148, 188–194. doi: 10.1016/j.actpsy.2014.02.001
- Field, A. (2012). *Discovering Statistics Using R*. London: SAGE.
- Magezi, D. A. (2015). Linear mixed-effects models for within-participant psychology experiments: an introductory tutorial and free, graphical user interface (LMMgui). *Front. Psychol.* 6:2. doi: 10.3389/fpsyg.2015.00002
- Van den Bussche, E., Vermeiren, A., Desender, K., Gevers, W., Hughes, G., Verguts, T., et al. (2013). Disentangling conscious and unconscious processing: a subjective trial-based assessment approach. *Front. Hum. Neurosci.* 7:769. doi: 10.3389/fnhum.2013.00769

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Hesselmann. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.