

Easy methods for extracting individual regression slopes: Comparing SPSS, R, and Excel

Roland Pfister,
*Julius-Maximilians
University of Würzburg*

Katharina Schwarz,
*University Medical Center
Hamburg-Eppendorf*

Robyn Carson,
*University of
Ottawa*

Markus Janczyk
*Julius-Maximilians
University of Würzburg*

Three different methods for extracting coefficients of linear regression analyses are presented. The focus is on automatic and easy-to-use approaches for common statistical packages: SPSS, R, and MS Excel / Libre Office Calc. Hands-on examples are included for each analysis, followed by a brief description of how a subsequent regression coefficient analysis is performed.

An increasingly popular analysis of within-subjects designs revolves around regression coefficients that are estimated individually for each participant. More precisely, a dependent variable (criterion) is regressed on an independent variable (predictor) individually for each participant. The extracted values for slopes and intercept are then compared between conditions or tested against a population value of 0 via standard significance tests such as paired-samples *t*-tests or repeated-measures analyses of variance (ANOVA). This procedure is commonly known as *regression coefficient analysis* (RCA; Lorch & Myers, 1990, Method 3).

RCA circumvents methodological problems of standard regression analysis which assumes different observations to be independent from each other. This assumption is routinely violated by data from within-subjects designs, but it does not apply to the coefficients that were extracted from individual data sets (cf. Lorch & Myers, 1990). In contrast, RCA only assumes a linear relationship between predictor and criterion for each individual participant and can be used for both, continuous and dichotomous predictors (Ahn, Jung, & Kang, 2002; Lorch & Myers, 1990; Myers & Broyles,

2000).^{*} RCA thus offers a flexible alternative to more common ANOVA approaches, and it has been applied to numerous different topics, covering as diverse areas as reading, emotion, cognitive control, and numerical cognition (see Table 1 for a more detailed overview).

In light of RCA's widespread use, the present paper attempts to give a brief overview of how to extract individual regression slopes efficiently with commonly used computer programs. To this end, we compare available methods for extracting regression slopes for three widespread statistical packages: SPSS 19, R 2.15 and MS Excel 2010 / LibreOffice 3.6 Calc. As a hands-on example, we demonstrate how regression slopes can be extracted to probe for spatial-numerical associations in a parity judgment task (Dehaene, Bossini, & Giraux, 1993; cf. Fias, Brysbaert, Geypens, & d'Ydewalle, 1996, Pfister, Schröder, & Kunde, in press).

* A major drawback of RCA is that this analysis does not yield any measures of variance-accounted-for (such as R^2). Such measures can be obtained from alternative approaches such as linear trend analyses for repeated-measures ANOVA on the one hand (Pinhas, Tzelgov, & Ganor-Stern, 2012) as well as hierarchical linear modeling or mixed modeling on the other hand (Baayen, Davidson, & Bates, 2008; Hoffman, & Rovine, 2007; Quené & van den Bergh, 2008; Van den Noortgate & Onghena, 2006).

Adress correspondence to Dipl.-Psych. Roland Pfister, Julius-Maximilians University of Würzburg, Department of Psychology III, Röntgenring 11, 97070 Würzburg, Germany
Tel +49-931-31-81363, Email: roland.pfister@psychologie.uni-wuerzburg.de.

Table 1. Recent applications of regression coefficient analysis (RCA) to diverse fields across psychology. The studies listed are necessarily a selection and do not attempt to give a complete overview of RCA in psychological research.

Area	Representative studies
Cognitive Control	Braem, Verguts, Roggeman, & Notebaert (2012); Kunde, Augst, & Kleinsorge (2012); Notebaert & Verguts (2007)
Emotion	Baumgartner, Willi, & Jäncke (2007); Petrova & Wentura (2012)
Grammar Learning	Lotz & Kinder (2006); Scott & Dienes (2008); Tunney (2010)
Numerical Cognition	Andres, Michaux, & Pesenti (2012); Cappelletti, Butterworth, & Kopelman (2012); Cohen Kadosh, Muggleton, Silvanto, & Walsh (2010); Duyck, Lagrou, Gevers, & Fias (2008); Ganor-Stern, Karasik-Rivkin, & Tzelgov (2011); Hartmann, Grabherr, & Mast (2012); Imbo, De Brauwer, Fias, & Gevers (2012); Lindemann & Tira (2011); Semenza, et al. (2012); Shaki, Petrusic, & Leth-Steensen (2012); Vierck & Kiesel (2010)
Reading and word recognition	Gao, Levinthal, Stine-Morrow (2012); Kliegl, Nuthmann, & Engbert (2006); Lewis & Mensink (2012); Wang, Pomplun, Chen, Ko, & Rayner (2010); Yap, Balota, Sibley, & Ratcliff (2012)
Risk perception	Pachur, Hertwig, & Steinmann (2012); Price, Smith, & Lench (2006)

Extracting regression slopes: Available methods

In a simulation of the mentioned task, we assume four participants to have judged the parity of the numbers 0-9 in a series of trials. Furthermore, participants are assumed to have used two mappings in the experiment: “even” responses were mapped to a left response key during one half of the experiment and to a right response key during the other half. Response times (RTs) can thus be analyzed as a function of target number and response side.

Similar experiments have consistently shown smaller numbers to facilitate left responses and larger numbers to facilitate right responses: the SNARC effect (spatial-numerical association of response codes; Dehaene et al., 1993; for an overview see Wood, Willmes, Nuerk, & Fischer, 2008). This finding is typically quantified as a negative slope when the SNARC effect ($RT_{\text{right}} - RT_{\text{left}}$) is regressed on the target number (Fias et al., 1996). The corresponding relations of target number and SNARC effect for the four simulated participants are plotted in Figure 1. These data were generated by adding normally distributed noise ($\mu = 0$; $\sigma = 25$) to the mean SNARC effects reported by Dehaene et al. (1993, Exp. 1). Data and scripts for the following slope extractions are available as supplementary material.

SPSS: Slope extraction via OMS

An efficient way to extract regression slopes with SPSS involves two separate steps (Figure 2). Individual regression analyses are first run for each participant and each condition of interest. The resulting coefficient tables are then automatically read from the output via the Output

Management System (OMS).[†] The two steps are described in detail below.

For the following example of the SNARC effect, we assume the data to consist of three variables (see the supplementary material for the corresponding data set). The variable `Subject` contains the participant number (1-4), the variable `Number` codes the target digit (0-9), and the variable `SNARC` codes the corresponding SNARC effect as the mean difference $RT_{\text{right}} - RT_{\text{left}}$ in milliseconds. The data thus contains ten cases (rows) per participant, each listing the SNARC effect for one of the target digits.

The two steps described above can then be defined in the syntax editor (see the supplementary material for the entire syntax file). To prepare the individual regression analyses, the data is first split according to the variable `Subject` using the menu *Data > Split File...* and the corresponding option *Compare groups*. Furthermore, a new data set is declared to prepare the call to OMS.

* Analyze each participant separately and initialize

* a new data set for regression output.

SORT CASES BY Subject.

SPLIT FILE LAYERED BY Subject.

[†] Alternative ways to obtain regression coefficients are described by Thompson (2008) and by Weaver and Dubois (2012). However, we believe the present solution to offer a more intuitive way because it does not require the user to have advanced background knowledge about SPSS macro facilities and/or training in matrix algebra as previous methods did.

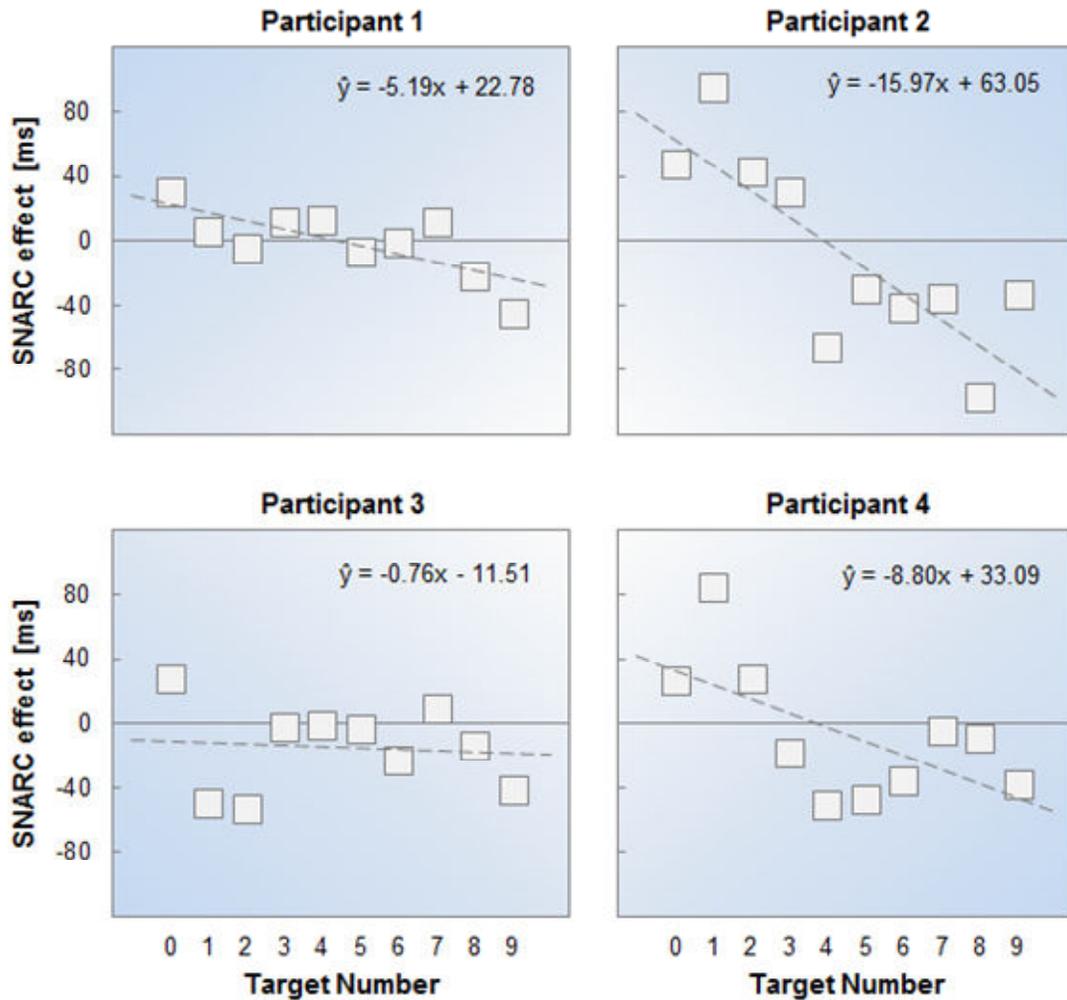


Figure 1. Example data of four participants who are assumed to have completed parity judgment task. SNARC effects ($RT_{\text{right}} - RT_{\text{left}}$) as a function of the corresponding target number clearly show a negative slope.

DATASET DECLARE SNARC_Slopes.

Before the actual regression analysis is performed (via *Analyze > Regression > Linear...*), we need to set up a call to OMS. This can be done either via the menu *Utilities > OMS Control Panel...* or by writing the syntax directly (which is preferred in this case). Basically, OMS accesses the output of upcoming function calls and can distill any information from the output (such as figures and tables). For the present purposes, we are looking for a specific table that is produced by the analysis command `REGRESSION`. Furthermore, we want to write the content of this table to the newly defined data set `SNARC_Slopes` (see above).

- * Define the request to the Output Management
- * System (OMS).

OMS

```

/SELECT TABLES
/IF COMMANDS=['Regression'] SUBTYPES=['Coefficients']
/DESTINATION FORMAT=SAV
OUTFILE = SNARC_Slopes.

```

Once this call is initialized, OMS starts collecting information from all upcoming regression commands until the call is stopped again by the user. Thus, we can now specify the regression analysis and stop OMS afterward.

- * The SNARC effect is regressed on the Number with
- * regression (as in *Analyze > Regression > Linear...*).

REGRESSION

```

/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)

```

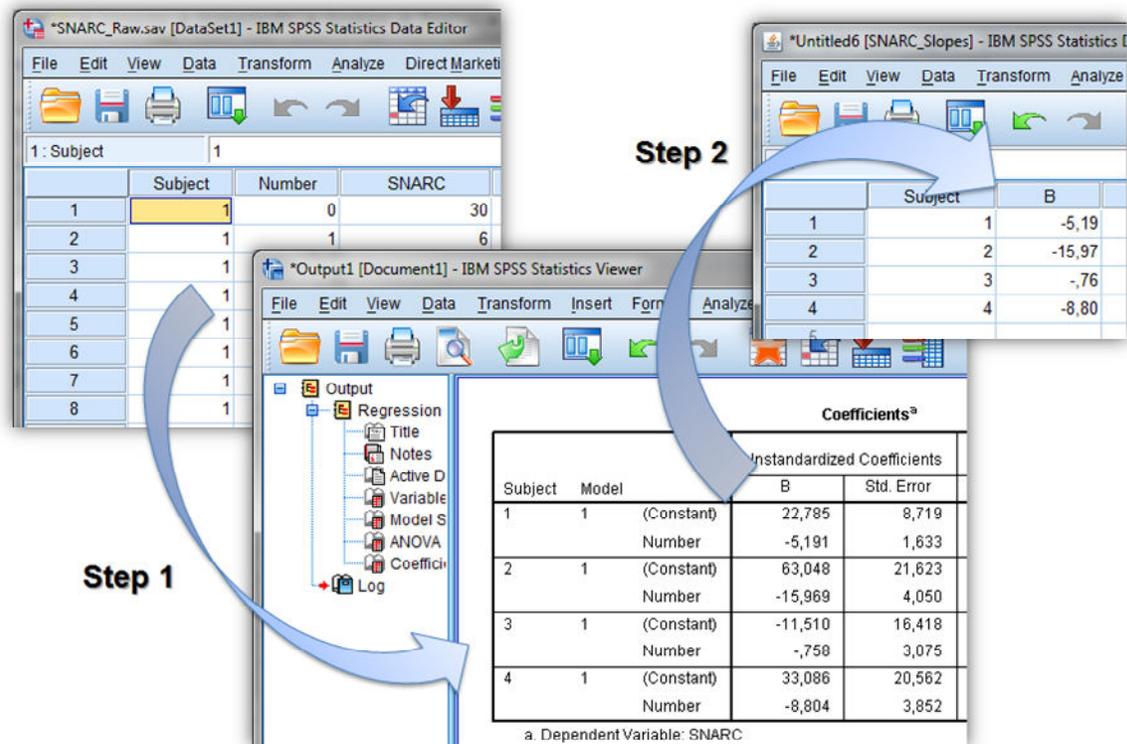


Figure 2. Slope extraction with SPSS 19. Individual regression analyses are first run for each participant and each condition of interest. The resulting coefficient tables are then automatically read from the output via the Output Management System (OMS).

```

/NOORIGIN
/DEPENDENT SNARC
/METHOD=ENTER Number.

```

* The OMS request is stopped; the SPLIT command
* is disabled.

OMSEND.

SPLIT FILE OFF.

Running the code from OMS to OMSSEND will now populate the newly defined data set `SNARC_Slopes` with the data from the coefficient tables of the regression analysis. Among others, the data includes the intercepts and slope coefficients for each participant. The new data set can be cleaned up easily to store only subject numbers and corresponding regression slopes (see the supplementary material). Other efficient ways to constrain the OMS output are explained in the SPSS Command Syntax Reference guide (IBM, 2010).

R: Slope extraction using linear models

Individual regression slopes can be extracted with only a few lines of R code and the most straightforward solution

uses the command `lm`. This command allows fitting a linear model to empirical data. The following demonstration is based on a data frame called `snarc_raw` (see the supplementary material). This data frame consists of three variables: The variable `Subject` codes the participant number (1-4), the variable `Number` codes the target number (0-9), and the variable `SNARC` codes the corresponding SNARC effect as the mean difference $RT_{\text{right}} - RT_{\text{left}}$ in milliseconds.

As a basic procedure, we use the `for` command to loop through the data frame and to compute separate analyses for each participant. For the example data set, the loop index `i` can thus be defined as a counter from 1 to 4 (a more flexible definition of the loop index is described in the supplementary material). Furthermore, we define the empty variable `snarc_coefs` to store the to-be-extracted regression slopes:

```

snarc_coefs = c(NA, NA, NA, NA)
for (i in c(1:4)) {
  # [...] run participant-wise analyses
}

```

Inside the loop, we perform four separate steps. First, the

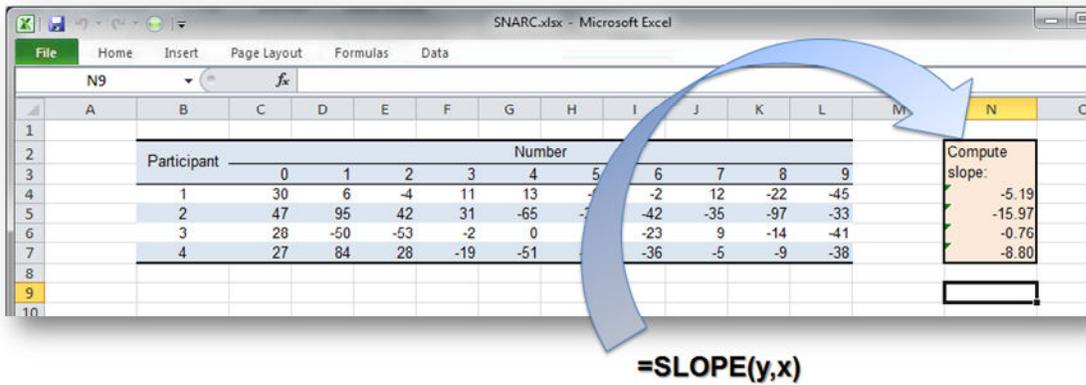


Figure 3. Slope extraction with Excel 2010. Individual regression slopes are readily available via the SLOPE function. Similarly, intercepts can be accessed via INTERCEPT. Non-English versions of Excel and Calc are likely to use translated function names and they might also use different ways to delimit the input arguments (e.g., a semicolon instead of the displayed comma).

relevant observations of the data frame `snarc_raw` is copied to a temporary data frame `snarc_tmp`. Then, the command `lm` is used to perform the linear regression of the SNARC effect on the corresponding number; the results of this analysis are saved as the linear model `reg_result`. The function `coef` is used subsequently to access the linear model. For a simple linear regression, the `coef` function returns a vector of two elements: the first element corresponds to the intercept, the second element is the slope coefficient. Finally, this coefficient is saved inside the summary vector `snarc_coefs`:

```
# Create temporary data frame:
snarc_tmp <-
  snarc_raw[snarc_raw$Subject==i,]
# Perform regression:
reg_result <- lm(snarc_tmp$SNARC ~
  snarc_tmp$Number)
# Get coefficient:
tmp_coef <- coef(reg_result)
# Store coefficient:
snarc_coefs[i] <- tmp_coef[2]
```

If the above series of steps is implemented inside a suitable loop, extracted coefficients will be readily available via the vectorarray `snarc_coefs`.

Excel / Calc: The SLOPE function

Extracting regression slopes is remarkably convenient with Excel 2010 and its open source counterpart Calc. For the simulated data, we can simply use the function `=SLOPE` to request the slope coefficient (similarly, `=INTERCEPT` returns the intercept). This function takes the criterion data (SNARC effect) as first argument and the predictor data

(numbers) as second argument (Figure 3).

This function can then be applied to each individual participant by dragging down the formula using the Fill Handle tool. To fix the predictor data for this operation, we need to define an absolute range of `x` values using the `$` operator. In the example of Figure 3, the correct specification of the `x` range would be `C$3:L$3`.

Concluding remarks

The preceding description of different methods for slope extraction can be summarized as follows. In SPSS, extracting regression coefficients involves two separate steps during which separate regression analyses are carried out and the corresponding output is fed back to a new data set using the OMS facilities. In R, we follow a similar procedure and loop through the data set while storing the output of a linear regression in each iteration. In Excel / Calc, regression slopes and intercepts for simple linear regressions are readily available via built-in functions.

Thus, Excel and Calc do seem to offer a very efficient way to perform participant-wise regression analyses (Lorch & Myers, 1990, Method 3) and we do indeed suggest that these programs offer an interesting alternative to standard statistical packages such as R or SPSS. This conclusion, however, only holds true for simple linear regressions using one predictor variable. More advanced setups (starting with multiple linear regression) are more difficult to handle with Excel or Calc, whereas the described methods for SPSS and R generalize easily and intuitively to these settings. In fact, the demonstrated algorithms for SPSS and R will work just as well for multiple regression and will simply enhance the output by data relating to the additional predictors.

Independent of the extraction method used, the obtained coefficients are then ready for subsequent RCA (Lorch &

Myers, 1990; for limitations of this approach and possible alternatives, see Van den Noortgate & Onghena, 2006). Such analyses may include rather simple comparisons such as testing whether the obtained coefficients differ significantly from zero via a one-sample *t*-test. This simple comparison would indeed be appropriate for the data presented in the above example of the SNARC effect; however, RCA can of course also involve any statistical test depending on the research question and the data obtained.

References

- Ahn, C., Jung, S.-H., & Kang, S.-H. (2002). Modified regression coefficient analysis for repeated binary measurements. *Journal of Applied Statistics*, 29(5), 703-710. doi: 10.1080/02664760120098766
- Andres, M., Michaux, N., & Pesenti, M. (2012). Common substrate for mental arithmetic and finger representation in the parietal cortex. *NeuroImage*, 62(3), 1520-1528. doi: 10.1016/j.neuroimage.2012.05.047
- Baumgartner, T., Willi, M., & Jäncke, L. (2007). Modulation of corticospinal activity by strong emotions evoked by pictures and classical music: A transcranial magnetic stimulation study. *NeuroReport*, 18(3), 261-265. doi: 10.1097/WNR.0b013e328012272e
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 387-556. doi:10.1016/j.jml.2007.12.005
- Braem, S., Verguts, T., Roggeman, C., & Notebaert, W. (2012). Reward modulates adaptations to conflict. *Cognition*, 125(2), 324-332. doi: 10.1016/j.cognition.2012.07.015
- Cappelletti, M., Butterworth, B., & Kopelman, M. (2012). Numeracy skills in patients with degenerative disorders and focal brain lesions: A neuropsychological investigation. *Neuropsychology*, 26(1), 1-19. doi: 10.1037/a0026328
- Cohen Kadosh, R., Muggleton, N., Silvanto, J., & Walsh, V. (2010). Double dissociation of format-dependent and number-specific neurons in human parietal cortex. *Cerebral Cortex*, 20(9), 2166-2171. doi: 10.1093/cercor/bhp273
- Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, 122(3), 371-396. doi: 10.1037/0096-3445.122.3.371
- Duyck, W., Lagrou, E., Gevers, W., & Fias, W. (2008). Roman digit naming: Evidence for a semantic route. *Experimental Psychology*, 55(2), 73-81. doi: 10.1027/1618-3169.55.2.73
- Fias, W., Brysbaert, M., Geypens, F., & d'Ydewalle, G. (1996). The importance of magnitude information in numerical processing: Evidence from the SNARC effect. *Mathematical Cognition*, 2(1), 95-110. doi: 10.1080/135467996387552
- Ganor-Stern, D., Karasik-Rivkin, I., & Tzelgov, J. (2011). Holistic representation of unit fractions. *Experimental Psychology*, 58(3), 201-206. doi: 10.1027/1618-3169/a000086
- Gao, X., Levinthal, B. R., & Stine-Morrow, E. A. L. (2012). The effects of ageing and visual noise on conceptual integration during sentence reading. *The Quarterly Journal of Experimental Psychology*, 65(9), 1833-1847. doi: 10.1080/17470218.2012.674146
- Hartman, M., Grabherr, L., & Mast, F. W. (2012). Moving along the mental number line: Interactions between whole-body motion and numerical cognition. *Journal of Experimental Psychology: Human Perception and Performance*, 38(6), 1416-1427. doi: 10.1037/a0026706
- Hoffman, L., & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods*, 39(1), 101-117. doi: 10.3758/BF03192848
- IBM. (2010). *IBM SPSS Statistics 19 Command Syntax Reference*. Retrieved on February 12, 2013, from the IBM website: <http://www-01.ibm.com/support/docview.wss?uid=swg27022445>
- Imbo, I., De Brauwer, J., Fias, W., & Gevers, W. (2012). The development of the SNARC effect: Evidence for early verbal coding. *Journal of Experimental Child Psychology*, 111(4), 671-680. doi: 10.1016/j.jecp.2011.09.002
- Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, 135(1), 12-35. doi: 10.1037/0096-3445.135.1.12
- Kunde, W., Augst, S., & Kleinsorge, T. (2012). Adaptation to (non)valent task disturbance. *Cognitive, Affective, & Behavioral Neuroscience*, 12(4), 644-660. doi: 10.3758/s13415-012-0116-8
- Lewis, M. R., & Mensink, M. C. (2012). Prereading questions and online text comprehension. *Discourse Processes*, 49(5), 367-390. doi: 10.1080/0163853X.2012.662801
- Lindemann, O., & Tira, M. D. (2011). Operational momentum in numerosity production judgments of multi-digit number problems. *Journal of Psychology*, 219(1), 50-57. doi: 10.1027/2151-2604/a00004
- Lorch, R. F., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(1), 149-157. doi: 10.1037/0278-7393.16.1.149
- Lotz, A., & Kinder, A. (2006). Transfer in artificial grammar learning: The role of repetition information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*,

- 32(4), 707-715. doi: 10.1037/0278-7393.32.4.707
- Myers, L., & Broyles, S. T. (2000). Regression coefficient analysis for correlated binomial outcomes. *Journal of Applied Statistics*, 27(2), 217-234. doi: 10.1080/02664760021754
- Notebaert, W., & Verguts, T. (2007). Dissociating conflict adaptation from feature integration: A multiple regression approach. *Journal of Experimental Psychology: Human Perception and Performance*, 33(5), 1256-1260. doi: 10.1037/0096-1523.33.5.1256
- Pachur, T., Hertwig, R., & Steinmann, F. (2012). How do people judge risks: Availability heuristic, affect heuristic, or both? *Journal of Experimental Psychology: Applied*, 18(3), 314-330. doi: 10.1037/a0028279
- Petrova, K., & Wentura, D. (2012). Upper-lower visual field asymmetries in oculomotor inhibition of emotional distractors. *Vision Research*, 62, 209-219. doi: 10.1016/j.visres.2012.04.010
- Pfister, R., Schroeder, P. A., & Kunde, W. (in press). SNARC struggles: Instant control over spatial-numerical associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi: 10.1037/a0032991
- Pinhas, M., Tzelgov, J., & Ganor-Stern, D. (2012). Estimating linear effects in ANOVA designs: The easy way. *Behavior Research Methods*, 44(3), 788-794. doi: 10.3758/s13428-011-0172-y
- Price, P. C., Smith, A. R., & Lench, H. C. (2006). The effect of target group size on risk judgments and comparative optimism: The more, the riskier. *Journal of Personality & Social Psychology*, 90(3), 382-398. doi: 10.1037/0022-3514.90.3.382
- Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, 59, 413-425. doi:10.1016/j.jml.2008.02.002
- Scott, R. B., & Dienes, Z. (2008). The conscious, the unconscious, and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(5), 1264-1288. doi: 10.1037/a0012943
- Semenza, C., Bonollo, S., Polli, R., Busana, C., Pignatti, R., Iuculano, T., ... Murgia, A. (2012). Genetics and mathematics: FMR1 premutation female carriers. *Neuropsychologia*, 50(14), 3757-3763. doi: 10.1016/j.neuropsychologia.2012.10.021
- Shaki, S., Petrusic, W. M., & Leth-Steensen, C. (2012). SNARC effects with numerical and non-numerical symbolic comparative judgments: Instructional and cultural dependencies. *Journal of Experimental Psychology: Human Perception and Performance*, 38(2), 515-530. doi: 10.1037/a0026729
- Thompson, G. L. (2008). Eliminating aggregation bias in experimental research: Random Coefficient Analysis as an alternative to performing a 'by-subjects' and/or 'by-items' ANOVA. *Tutorials in Quantitative Methods for Psychology*, 4(1), 21-34.
- Tunney, R. J. (2010). Similarity and confidence in artificial grammar learning. *Experimental Psychology*, 57(2), 160-168. doi: 10.1027/1618-3169/a000020
- Van den Noortgate, & Onghena, P. (2006). Analysing repeated measures data in cognitive research: A comment on regression coefficient analyses. *European Journal of Cognitive Psychology*, 18(6), 937-952. doi: 10.1080/09541440500451526
- Vierck, E., & Kiesel, A. (2010). Congruency effects between number magnitude and response force. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(1), 204-209. doi: 10.1037/a0018105
- Wang, H.-C., Pomplun, M., Chen, M., Ko, H., & Rayner, K. (2010). Estimating the effect of word predictability on eye movements in Chinese reading using latent semantic analysis and transitional probability. *The Quarterly Journal of Experimental Psychology*, 63(7), 1374-1386. doi: 10.1080/17470210903380814
- Weaver, B., & Dubois, S. (2012). SPSS macros to compare any two fitted values from a regression model. *Behavior Research Methods*. doi: 10.3758/s13428-012-0204-2
- Wood, G., Willmes, K., Nuerk, H.-C., & Fischer, M. H. (2008). On the cognitive link between space and number: A meta-analysis of the SNARC effect. *Psychology Science*, 50(4), 489-525.
- Yap, M. J., Balota, D. A., Sibley, D. E., & Ratcliff, R. (2012). Individual differences in visual word recognition: Insights from the English lexicon project. *Journal of Experimental Psychology: Human Perception and Performance*, 38(1), 53-79. doi: 10.1037/a0024177

Manuscript received 19 December 2012

Manuscript accepted 26 February 2013