Pre-Registration of Replication Study (Levy & Keller, 2013)

Daniela Mertzen

University of Potsdam, Germany

Lena A. Jäger

University of Potsdam, Germany

Shravan Vasishth

University of Potsdam, Germany

May 11, 2017

Author Note

Please send correspondence to vasishth@uni-potsdam.de.

Pre-Registration of Replication Study (Levy & Keller, 2013)

Background to the current study

We attempt to replicate a psycholinguistic study by Levy and Keller (2013) (henceforth, LK13) who investigated locality and anti-locality effects in German. The main claim of this study is that both locality and anti-locality effects play a role in the processing of verb-final clauses of German, however, locality effects dominate over anti-locality effects when memory load is high.

The study comprises of two eyetracking-while-reading experiments. *Experiment 1* tested predictions of surprisal theory: Intervening material between a verb and its subject should *facilitate* the parser's prediction of the verb and, thus, its processing (anti-locality effect). *Experiment 2* tested predictions of memory-based theories: There should be more *processing difficulty* at the verb due to intervening material between the verb and its subject (locality effect). This was investigated using complex syntactic structures where the subject and the critical matrix verb had four configurations in which the position of a *dative noun phrase* (Dat NP) and of a *prepositional adjunct* (PP Adj) were manipulated (see Table 1 for a simplified example of an experimental item of Experiment 1). Experiment 1 had the target construction in a main clause; in Experiment 2 the same target construction as in Exp. 1 was more deeply embedded in a relative clause, making the sentence syntactically more complex (see Table 2 for a simplified experimental item of Exp. 2). The critical region in Exp. 1 was the matrix clause verb; in Exp. 2 the critical region was the head verb of the relative clause and the auxiliary (*versteckt* in Exp. 1 vs. *versteckt hat* in Exp. 2 (Tables 1 and 2, respectively).

Table 1

LK13 Exp. 1: Example of experimental item (simplified).

a. PP Adjunct in subordinate clause, Dative NP in subordinate clause

Nachdem	der	Lehrer	[**PP** zur	Ahnd.]	[**NP** dem	Sohn]	…,	hat	Hans Gerstner …		den	Fußball	**versteckt**, …
After	*the*	*teacher*	[**PP** *as*	*payback*]	[**NP** *the*	*son*]	…,	*has*	*Hans Gerstner …*		*the*	*football*	*hid, …*

b. PP Adjunct in main clause, Dative NP in subordinate clause

Nachdem	der	Lehrer			[**NP** dem	Sohn]	…,	hat	Hans Gerstner …	[**PP** zur	Ahnd.]	den	Fußball	**versteckt**, …
After	*the*	*teacher*			[**NP** *the*	*son*]	…,	*has*	*Hans Gerstner …*	[**PP** *as*	*payback*]	*the*	*football*	*hid, …*

c. PP Adjunct in subordinate clause, Dative NP in main clause

Nachdem	der	Lehrer	[**PP** zur	Ahnd.]			…,	hat	Hans Gerstner …		[**NP** dem	Sohn]	den	Fußball	**versteckt**, …
After	*the*	*teacher*	[**PP** *as*	*payback*]			…,	*has*	*Hans Gerstner …*		[**NP** *the*	*son*]	*the*	*football*	*hid, …*

d. PP Adjunct in main clause, Dative NP in main clause

Nachdem	der	Lehrer					…,	hat	Hans Gerstner …	[**PP** zur	Ahnd.]	[**NP** dem	Sohn]	den	Fußball	**versteckt**, …
After	*the*	*teacher*					…,	*has*	*Hans Gerstner …*	[**PP** *as*	*payback*]	[**NP** *the*	*son*]	*the*	*football*	*hid, …*

'*After the teacher imposed detention classes, Hans Gerstner hid the football from the naughty son of the industrious janitor as additional payback for the multiple wrongdoings, and thus corrected the affair.*'

Table 2

LK13 Exp. 2: Example of experimental item (simplified)

a. PP Adjunct in subordinate clause, Dative NP in subordinate clause

| Nachdem | der | Lehrer | [**PP** zur | Ahnd.] | [**NP** dem | Sohn] | ..., | hat | der Mitschüler, | der | | | | den | Fußball | **versteckt hat**, ... |
| *After* | *the* | *teacher* | [**PP** *as* | *payback*] | [**NP** *the* | *son*] | ..., | *has* | *the classmate*, | *who* | | | | *the* | *football* | *hid*, ... |

b. PP Adjunct in relative clause, Dative NP in subordinate clause

| Nachdem | der | Lehrer | | | [**NP** dem | Sohn] | ..., | hat | der Mitschüler, | der | [**PP** zur | Ahnd.] | | den | Fußball | **versteckt hat**, ... |
| *After* | *the* | *teacher* | | | [**NP** *the* | *son*] | ..., | *has* | *the classmate*, | *who* | [**PP** *as* | *payback*] | | *the* | *football* | *hid*, ... |

c. PP Adjunct in subordinate clause, Dative NP in relative clause

| Nachdem | der | Lehrer | [**PP** zur | Ahnd.] | | | ..., | hat | der Mitschüler, | der | | | [**NP** dem | Sohn] | den | Fußball | **versteckt hat**, ... |
| *After* | *the* | *teacher* | [**PP** *as* | *payback*] | | | ..., | *has* | *the classmate*, | *who* | | | [**NP** *the* | *son*] | *the* | *football* | *hid*, ... |

d. PP Adjunct in relative clause, Dative NP in relative clause

| Nachdem | der | Lehrer | | | | | ..., | hat | der Mitschüler, | der | [**PP** zur | Ahnd.] | [**NP** dem | Sohn] | den | Fußball | **versteckt hat**, ... |
| *After* | *the* | *teacher* | | | | | ..., | *has* | *the classmate*, | *who* | [**PP** *as* | *payback*] | [**NP** *the* | *son*] | *the* | *football* | *hid*, ... |

'After the teacher imposed detention classes, the classmate who hid the football from the naughty son of the industrious janitor as additional payback for the multiple wrongdoings corrected the affair.'

Predictions of the original study

Predictions for the original LK13 Experiments 1 and 2 were as follows: According to surprisal, the verb *versteckt* in condition (a) should be read slower than in (b), and (c) slower than (d); hence, (d) should show the most facilitation (*anti-locality effect*) as the upcoming verb is more expected (see left panel of Fig. 1).

Memory-based theories, on the other hand, predict a *locality effect*: (a) should be read faster than (b), and (c) faster than (d); i.e., (d) should yield the greatest slowdown at the critical matrix clause verb (see right panel of Fig. 1).
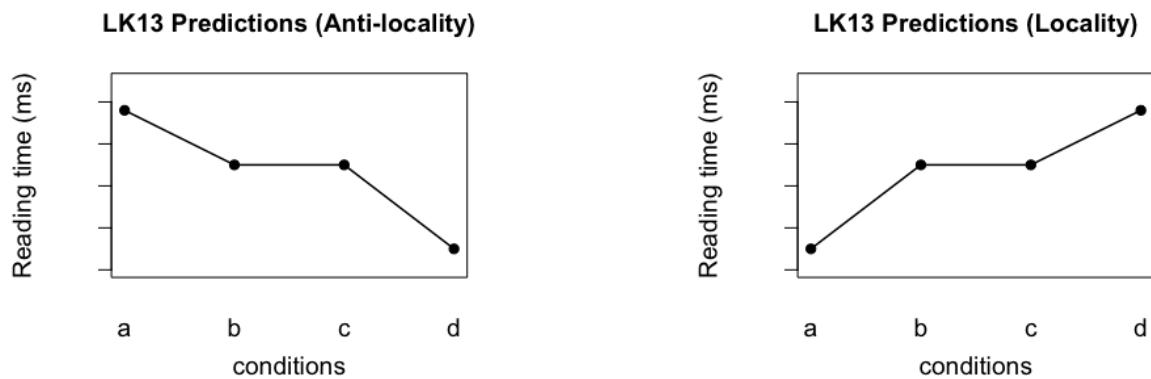
Figure 1. LK13 predictions for Exp. 1 and 2: Left panel: Anti-locality effect predicted by surprisal theory. Right panel: Locality effect predicted by memory-based theories.

Results of the original study

The original analysis of reading measures was carried out on the raw scale. To display the original results, we show model output of the raw reading time. However, we will re-run the analysis using a log-transformation of the reading measures. Analyzing the reading measures (here, total reading time) on the raw scale yields residuals which are non-normally distributed violating the normality assumption of the model whereas using a log-transformation yields normal residuals (see Fig. 2), satisfying model assumptions.

Results for Experiment 1 show a speed-up for (d) over (c) consistent with surprisal (i.e., an anti-locality effect, see left panel of Fig. 3).
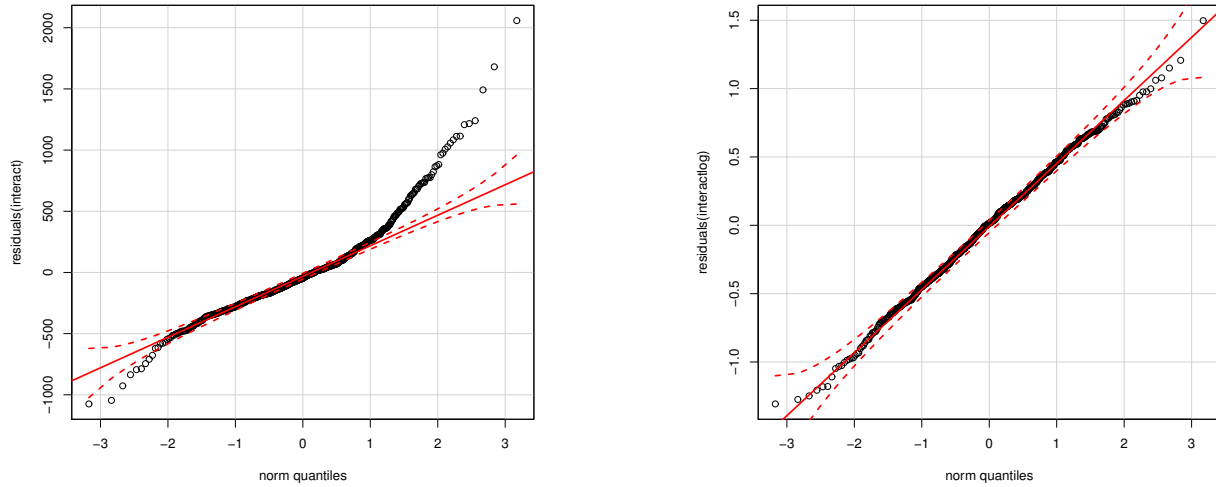
Figure 2. Left panel: Residuals (plotted against quantiles of the normal distribution) of linear mixed effects model analyzing raw reading time. Right panel: Residuals of linear mixed effects model analyzing log-transformed reading time.
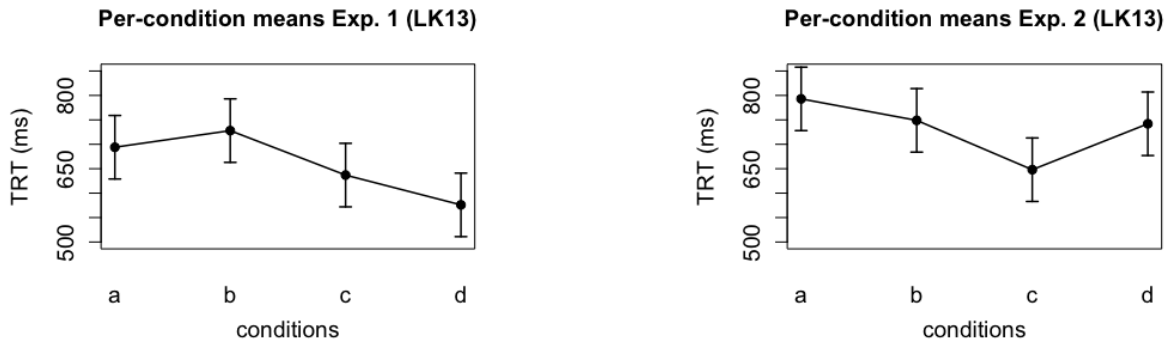

Figure 3. Left panel: LK13 Exp. 1: Per-condition means (total reading time) and 95% confidence intervals at the critical region. Right panel: LK13 Exp. 2: Per-condition means and 95% confidence intervals at the critical region

In Experiment 1, it was "found that the presence of a dative noun phrase led to decreased reading time at the corresponding verb, compared to a condition in which there is no preceding dative noun phrase. This can be explained by assuming that the presence [of] the additional preverbal material allows the processor to predict the upcoming verb, which leads to a facilitation effect" (Levy & Keller, 2013, p.214). See Fig. 4 for coefficient estimates of main effects and interaction.

However, in Experiment 2, the opposite pattern was observed: (d) was read slower than (c)
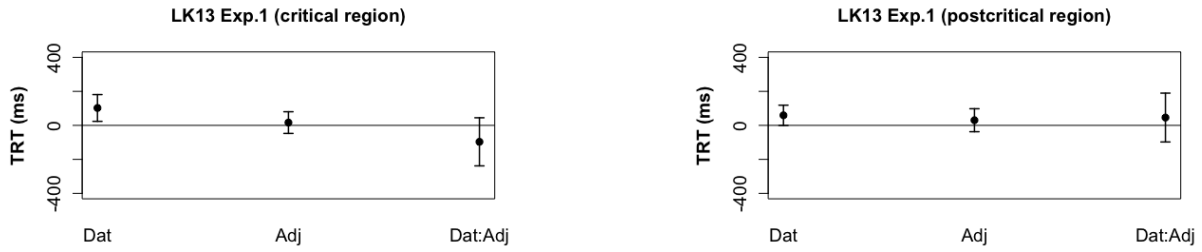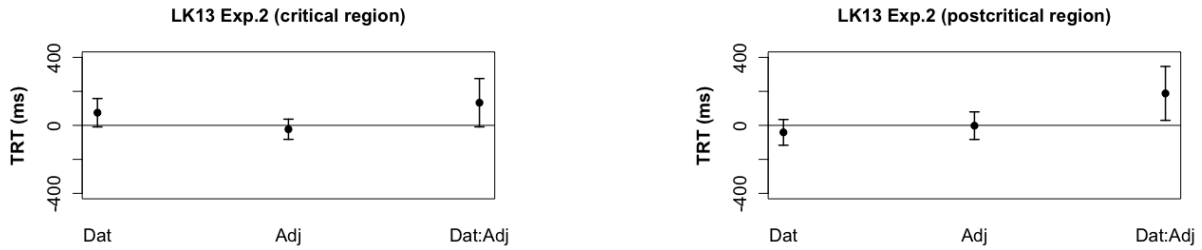
Figure 4. LK13 Exp. 1: Coefficient estimates (total reading time in ms) and 95% confidence intervals at the critical and postcritical region for main effect of 'Dat', main effect of 'Adj' and interaction of 'Dat:Adj'

(i.e., a locality effect as predicted by memory-based theories, see right panel of Fig. 3).

Figure 5. LK13 Exp. 2: Coefficient estimates (total reading time in ms) and 95% confidence intervals at the critical and postcritical region for main effect of 'Dat', main effect of 'Adj' and interaction of 'Dat:Adj'

Experiment 2 "showed an interaction of adjunct position and dative position, with the verb more difficult to process when both the adjunct and the dative phrase were present than when only one was present. This suggests the presence of a locality effect, i.e., the additional material that needs to be integrated at the verb, leading to a distance-based cost" (Levy & Keller, 2013, p.214). See Fig. 5 for coefficient estimates of main effects and interaction. The "interaction of dative and adjunct position [is] significant in second-pass times for the critical and spillover regions and in total times for the spillover region." The "interaction in total time at the critical region is not significant in Experiment 2, the qualitative pattern is the same as second-pass time at the region, for which the interaction is significant" (Levy & Keller, 2013, p.213).

To summarize, Levy & Keller found evidence of both locality and anti-locality effects for

the processing of complex syntactic structures in German. Anti-locality effects were found in Experiment 1 where a *facilitation* of processing was observed at the critical matrix verb when a dative noun phrase intervened between the verb and its subject. A locality effect was present in Experiment 2 in which the target construction was embedded in a relative clause as opposed to a main clause leading the authors to conclude that locality effects dominate over anti-locality effects when syntactic complexity is increased, as in Experiment 2.

Aim of our replication attempts

Statistical theory (Gelman & Carlin, 2014) states that, if an effect is indeed present in reality, repeatedly conducting an experiment with low power will lead to a high proportion of (i) null results, and (ii) exaggerated effect sizes which could be in the wrong direction.

Our aim was to investigate whether the relatively small sample sizes in psycholinguistic studies lead to statistically significant results that subsequently cannot be reproduced. We chose the study by Levy and Keller (2013) as a test case, because it has a sample size typical for psycholinguistic studies (28 subjects in each of their two studies) and its results are highly plausible. Findings are supported by surprisal- (Hale, 2001; Levy, 2008) and memory-based theories (Gibson, 2000; Lewis & Vasishth, 2005) and there is independent empirical evidence in favor of locality and anti-locality effects (Locality: *English*: Bartek, Lewis, Vasishth, and Smith (2017); Grodner and Gibson (2005); *Hindi*: Husain, Vasishth, and Srinivasan (2014); *Persian*: Safavi, Husain, and Vasishth (2016); Anti-locality: *English*: Boston, Hale, Patil, Kliegl, and Vasishth (2008); Linzen and Jaeger (2016); *German*: Boston, Hale, Vasishth, and Kliegl (2011); Frank, Trompenaars, and Vasishth (2015); Konieczny (2000).

Existing replication data

We conducted a total of six attempted replications of LK13 Experiment 1 and 2 using the same 24 experimental items and identical participant sample size of 28 for each of the six replication attempts. More specifically, we tested:

- LK13 Exp. 1: Attempted replication 1, *self-paced reading (SPR)*, N = 28

- LK13 Exp. 1: Attempted replication 2, *eye-tracking (ET)*; N = 28

- LK13 Exp. 2: Attempted replication 3, *SPR*, N = 28

- LK13 Exp. 2: Attempted replication 4, *ET*, N = 28

As LK13 found statistically significant effects in only conditions (c) and (d) of their Experiments 1 and 2, we combined conditions (c) and (d) from each of their two experiments and ran this as one experiment, using both self-paced reading and eye-tracking:

- LK13 Exp. 1,2 (c, d): Attempted replication 5, *SPR*, N = 28

- LK13 Exp. 1,2 (c, d): Attempted replication 6, *ET*, N = 28

Replication results (Experiment 1)

Figures 6 and 7 show a summary of the results of the attempted replication of Exp. 1, for both the SPR and the ET study. Dependent measures are raw reading time (RT) and total reading time (TRT) for SPR and ET, repectively. Note that we originally ran the analyses of our previous six replication attempts using a log-transformation of reading measures. However, here, we show coefficient estimates based on models on the raw scale (as was done by LK13). No dependent measure showed any statistically significant effect.

Replication results (Experiment 2)

Figures 8 and 9 summarize results for the replication attempt of LK13 Experiment 2.

Results ('combined' Exp. 1 & 2, conditions c, d)

A summary of the results for combined replication attempts 5 and 6 are shown in Figures 10 and 11. Recall that in the 'combined' studies, conditions (a) and (b) had the target construction in a main clause whereas in (c) and (d) the target construction was embedded
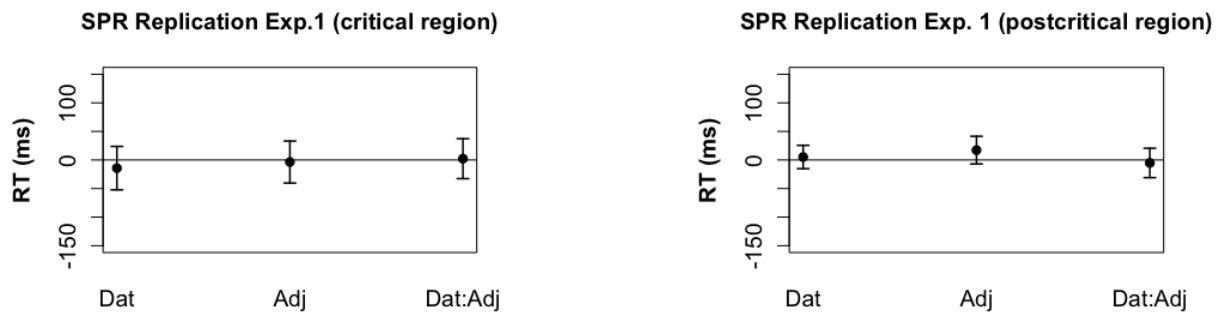
Figure 6. SPR Replication of LK13 Exp. 1: Coefficient estimates (RT in ms) and 95% confidence intervals at the critical and postcritical region for main effect of 'Dat', main effect of 'Adj' and interaction of 'Dat:Adj'
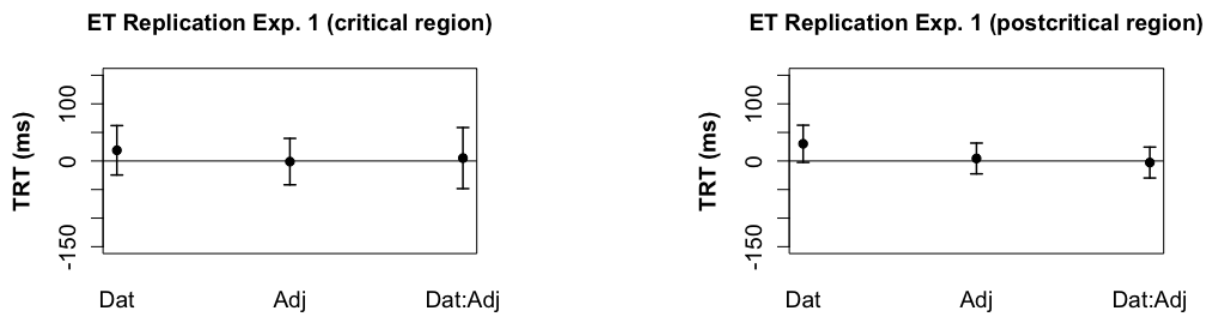

Figure 7. ET Replication of LK13 Exp. 1: Coefficient estimates (reading time in ms) and 95% confidence intervals at the critical and postcritical region for main effect of 'Dat', main effect of 'Adj' and interaction of 'Dat:Adj'
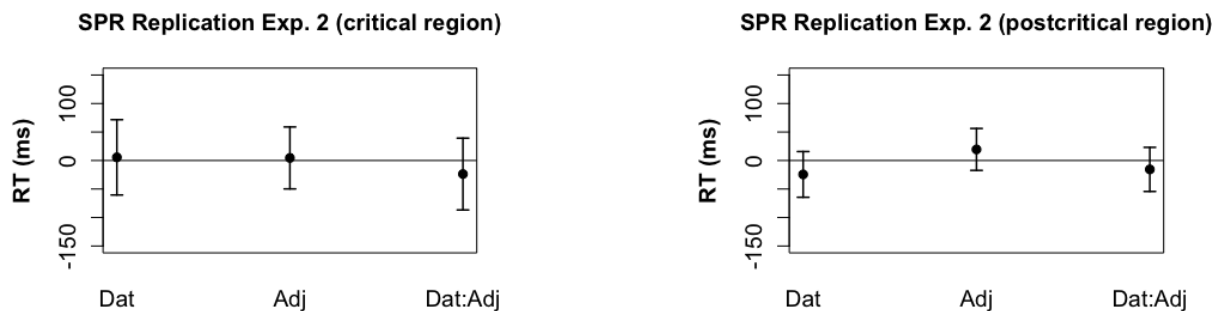

Figure 8. SPR Replication of LK13 Exp. 2: Coefficient estimates (reading time in ms) and 95% confidence intervals at the critical and postcritical region for main effect of 'Dat', main effect of 'Adj' and interaction of 'Dat:Adj'

in a relative clause (i.e., 'load' refers to the target construction in main clause vs relative
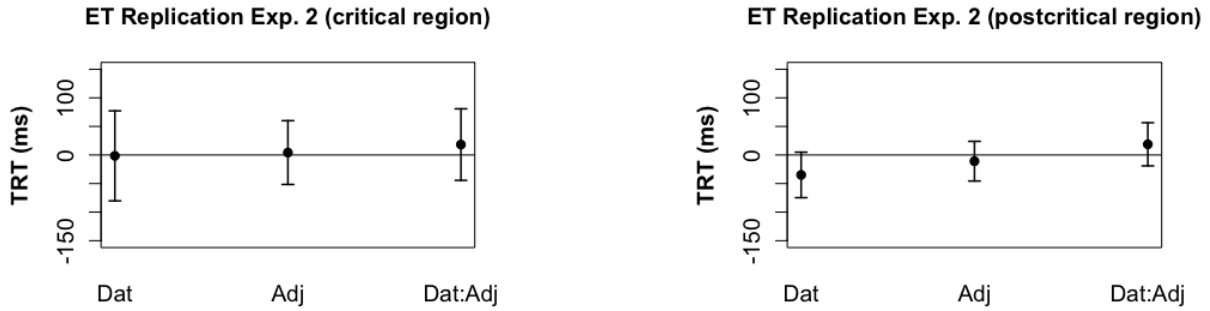
Figure 9. Coefficient estimates (total reading time in ms) and 95% confidence intervals at the critical and postcritical region for main effect of 'Dat', main effect of 'Adj' and interaction of 'Dat:Adj'

clause[1]). Conditions (a) and (c) had an intervening dative NP between the critical verb and its subject; in (b) and (d) both a dative NP and a PP adjunct served as the intervening material between the critical verb and the subject (this is referred to as the 'distance' manipulation here). Pairwise comparisons of (a) vs (b), and (c) vs (d) showed no statistical significance.
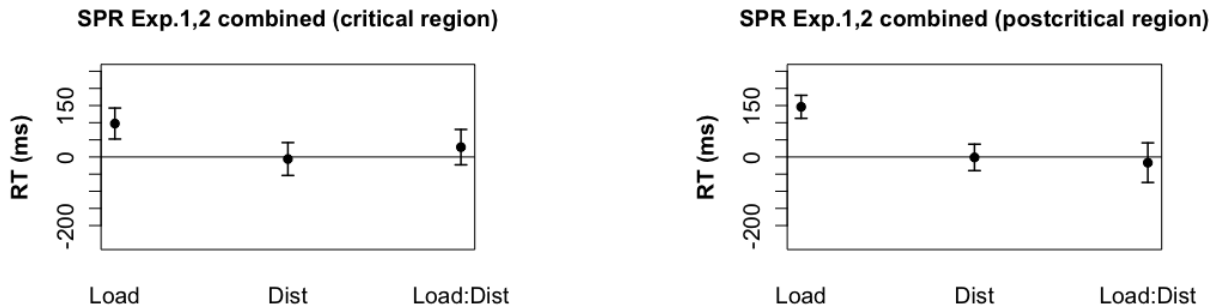

Figure 10. Combined SPR Replication Exp. 1, 2: Coefficient estimates (reading time in ms) and 95% confidence intervals at the critical and postcritical region for main effect of 'load', main effect of 'distance' and interaction of load:dist

All six attempts to reproduce the originally significant results were unsuccessful; no dependent measure showed any effect.

_____

[1]Note that the main effect of 'load' is statistically significant; however, this is not a contrast of interest.
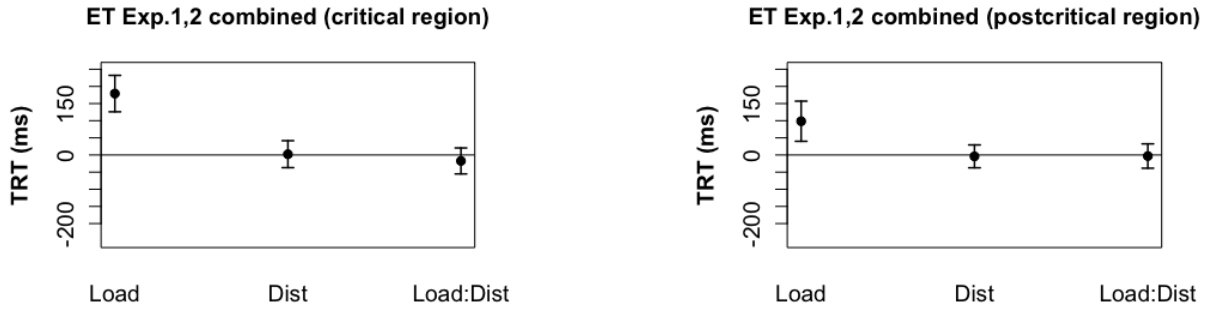
Figure 11. Combined ET Replication Exp. 1,2: Coefficient estimates (total reading time in ms) and 95% confidence intervals at the critical and postcritical region for main effect of 'load', main effect of 'distance' and interaction of load:dist

The planned replication study

To investigate whether the originally statistically significant effects are present, we re-run our attempted replication 6 (see details above), aiming for a high degree of precision in our estimate.

The reason for choosing to replicate the 'combined' study rather than attempting to replicate the original experiments (Exp. 1 and 2) separately, is that statistically significant effects were found in only conditions (c) and (d) of Levy & Keller's original studies. Furthermore, we choose to run the replication study using the eye-tracking rather than the self-paced reading method, as the original study was conducted using the eye-tracking method and the originally significant effects were observed in eye-tracking reading measures.

Predictions

Recall that in Experiment 1 it was "found that the presence of a dative noun phrase led to decreased reading time at the corresponding verb, compared to a condition in which there is no preceding dative noun phrase" (i.e., an anti-locality effect as predicted by surprisal) (Levy & Keller, 2013, p.214) whereas results of Experiment 2 "showed an interaction of adjunct position and dative position, with the verb more difficult to process when both the adjunct

and the dative phrase were present than when only one was present" (i.e., a locality effect as predicted by memory-based accounts) (Levy & Keller, 2013, p.214).
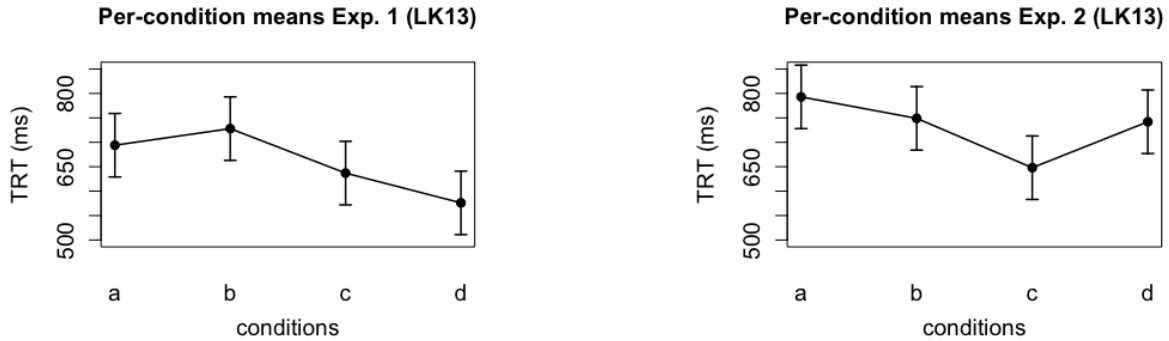

Figure 12. Our predictions based on original LK13 results. Left panel: LK13 Exp. 1: Per-condition means (total reading time) and 95% confidence intervals at the critical region. Right panel: LK13 Exp. 2: Per-condition means and 95% confidence intervals at the critical region. We will attempt to replicate the patterns for only *conditions (c) and (d)* of Exp. 1 and 2.

On the basis of the originally statistically significant effects (Exp. 1: speed-up in (d) vs (c); Exp. 2: slowdown in (d) vs (c)), we will attempt to replicate the patterns for (c) and (d) (Figure 12) of the original Exp. 1 and 2. Pairwise comparisons in our planned replication study should therefore show a speed-up for (b) vs (a) (these were conditons (d) and (c) of the original Experiment 1) and a slowdown for (d) compared to (c) (these were conditions (d) and (c) of the original Experiment 2).

Method

An eyetracking-while-reading study will be conducted.

Participants. Participants will be native German (mostly) undergraduate students from the University of Potsdam, Germany.

Participants will be recruited via the undergraduate participant pool of the language processing laboratory (Prof. Shravan Vasishth, Department of Linguistics, University of Potsdam, Germany).

All participants will receive financial compensation, or, if preferred, will receive course credit for their participation.

Design.   2 x 2 fully-crossed factorial design

- Factor 1: memory load (*target construction in main clause* vs. *target construction embedded in relative clause*)

- Factor 2: distance between the critical matrix clause verb and its subject (*dative NP only* vs. *dative NP and PP adjunct*)

- Both factors are within-items, within-subjects manipulations

Experimental items.   Experimental items were made available to us by the authors of the original study (Levy & Keller, 2013). In our planned 'combined' replication study, conditions (a) and (b) were orginally conditions (c) and (d) of LK13 Exp. 1 and have the target construction in a main clause whereas conditions (c) and (d) of our planned study were originally conditions (c) and (d) of LK13 Exp. 2 and have the target construction in a relative clause (see Table 3 for a simplified example item of the 'combined' study).

Table 3

Example of experimental item (simplified) 'combined' replication study

a. PP Adjunct in subordinate clause, Dative NP in main clause

Nachdem	der	Lehrer	[**PP** zur	Ahnd.]	...,	hat	Hans Gerstner,			[**NP** dem	Sohn]	den	Fußball	**versteckt hat**, ...
After	*the*	*teacher*	[**PP** *as*	*payback*]	*...,*	*has*	*Hans Gerstner,*			[**NP** *the*	*son*]	*the*	*football*	*hid, ...*

b. PP Adjunct in main clause, Dative NP in main clause

Nachdem	der	Lehrer			...,	hat	Hans Gerstner,	[**PP** zur	Ahnd.]	[**NP** dem	Sohn]	den	Fußball	**versteckt hat**, ...
After	*the*	*teacher*			*...,*	*has*	*Hans Gerstner,*	[**PP** *as*	*payback*]	[**NP** *the*	*son*]	*the*	*football*	*hid, ...*

c. PP Adjunct in subordinate clause, Dative NP in relative clause

Nachdem	der	Lehrer	[**PP** zur	Ahnd.]	...,	hat	der Mitschüler,	der		[**NP** dem	Sohn]	den	Fußball	**versteckt hat**, ...
After	*the*	*teacher*	[**PP** *as*	*payback*]	*...,*	*has*	*the classmate,*	*who*		[**NP** *the*	*son*]	*the*	*football*	*hid, ...*

d. PP Adjunct in relative clause, Dative NP in relative clause

Nachdem	der	Lehrer			...,	hat	der Mitschüler,	der	[**PP** zur	Ahnd.]	[**NP** dem	Sohn]	den	Fußball	**versteckt hat**, ...
After	*the*	*teacher*			*...,*	*has*	*the classmate,*	*who*	[**PP** *as*	*payback*]	[**NP** *the*	*son*]	*the*	*football*	*hid, ...*

'After the teacher imposed detention classes, Hans Gerstner/the classmate (who) hid the football from the naughty son of the industrious janitor as additional payback for the multiple wrongdoings corrected the affair.'

Filler items from the original experiments were used for our previous replication attempts as well as our planned replication study. Items will be presented in a Latin-square design and will be pseudo-randomised.

The stopping rule.   Data will be collected until either

(a) we reach our goal of achieving a high degree of precision. We aim for a small standard error of ≤ 10 for total reading time (TRT). Thus, when our estimate reaches $\hat{\beta} \pm 2 \times 10$ — fitting a Bayesian linear mixed model in Stan (Stan Development Team (2016) assuming a log-normal distribution for reading times) — testing will stop. The reason for aiming for high precision in total reading time, as opposed to first-pass reading time, is that FPRTs are commonly lower (with a smaller standard error) than TFTs, and, thus, a desired precision would be reached sooner in FPRTs compared to TRTs. Aiming for a certain degree of precision in total reading time is therefore more conservative. Or,

(b) the study will run until we reach a maximum participant sample size of 150 due to logistical limitations.

Procedure.   The procedure of the upcoming replication study will be identical to the previous small sample (N = 28) study, i.e., our previous replication attempt 6. The planned replication will be conducted in the same laboratory under the same conditions. Participants will be recruited via the Vasishth laboratory participant pool. Before the start of the experiment, participants will be briefed and must give their consent. Instructions regarding the experimental procedure will be given in writing and will be identical to the previous replication study. The setup of the eye-tracker, the eye-tracking procedure as well as the room conditions will be identical to those of the previous replication study. The experiment will take approximately 45 minutes.

Eye-tracker settings.   The experiment was created with the Experiment Builder software (SR Research). We use an EyeLink 1000 eye-tracker (SR Research) with a desktop-mounted camera setup for data collection. All settings will be identical to those used for our replication

attempt 6. (Further information on the experimental procedure and set-up of the eye-tracker available upon request).

Data exclusion/missing data. No data cleaning will be carried out, i.e., fixations which fall outside of the pre-defined interest areas will be excluded from analysis.

Deviations from the original study. In our 'combined' experiment, filler items from both the original Exp. 1 and 2 were used.

Different eye-tracking systems were used: The original experiments were carried out with an Eyelink II (500 Hz sampling rate) head-mounted system by SR Research. The 'Eyetrack' software (which was developed at the University of Massachusetts at Amherst) was used to implement the experiment. Our previous replication attempts as well as the planned study will be conducted using an Eyelink 1000 (SR Research) with a sampling rate of 1000 Hz. The Experiment Builder software by SR Research was used to create the experiment.

Presentation of the experimental items on the presentation monitor varied from the original study. In the original Exp. 1 the critical verb appeared in the middle of either the third or fourth line of the presented text. In Exp. 2 the critical verb was always the fourth word of the fourth line on the presentation screen. In our replication attempts, the critical verb always appears in the same position, i.e., as the fourth word of the fourth line of the presented text. The original LK13 study was conducted with native German undergraduate students (attending a summer school) at Edinburgh University, UK.

Analysis of reading measures

As in the original LK13 study as well as our six previous replication attempts, reading measures will be analyzed at the critical region (matrix clause verb) and the postcritical region (the word following the matrix clause verb). Linear mixed-effects models (with varying intercepts and varying slopes for subjects and items — given model convergence) will be used to analyze the following eye-tracking reading measures: First-pass reading time (FPRT), total reading time (TRT) and re-reading time (RRT). In the analysis of the original study,

zero values were not removed from the RRT analysis. This implies an invalid underlying generative model for the data: In the data approximately 50% of the trials had zero re-reading times. It is arguable whether a linear mixed model assuming a normal distribution as a generative distribution should be fit in this case. Therefore, we will run two models for RRT — including zero values in one of the models and excluding zero values in the other. We will further analyze re-reading probability (RRP) at the critical and post-critical region of the original study, as well as our replication attempts. The reason for looking at RRP is that it allows us to investigate whether interposing a dative NP or PP adjunct increases or decreases the probability of re-reading. The Box Cox function will determine the appropriate transformation of the reading measures.

In the analysis of our six previous replication attempts, reading measures were log-transformed as mentioned above, so we will analyze on the log scale in addition to whatever the Box Cox procedure suggests.

Additionally, we will conduct two separate analyses of the reading measures at the critical and postcritical region. In the first analysis outliers will not be removed. In a second analysis, outliers will be removed as is commonly done in psycholinguistic reading studies. Here, fixations > 800 ms will be excluded before analysis of the reading measures, and, for each reading measure, outliers that are below or above 3.5 standard deviations of participants' mean RT will be removed.

We will subsequently carry out an analysis of the data within a Bayesian framework for all replication attempts of the original study, as well as the original studies (raw data of the original study was obtained from the authors[2]). We will run a correlated varying intercepts, varying slopes mixed-effects model in Stan (Stan Development Team, 2016) to derive a posterior distribution of our parameters of interest.

---

[2]Our thanks to Roger Levy and Frank Keller for sharing experimental items and code of the original analyses.

References

Bartek, B., Lewis, R. L., Vasishth, S., & Smith, M. (2017). In search of on-line locality effects in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *5*(37), 1178–1198.

Boston, M. F., Hale, J. T., Patil, U., Kliegl, R., & Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the potsdam sentence corpus. *Journal of Eye Movement Research*, *1*(2), 1–12.

Boston, M. F., Hale, J. T., Vasishth, S., & Kliegl, R. (2011). Parallel processing and sentence comprehension difficulty. *Language and Cognitive Processes*, *26*(3), 301–349.

Frank, S. L., Trompenaars, T., & Vasishth, S. (2015). Cross-linguistic differences in processing double-embedded relative clauses: Working-memory constraints or language statistics? *Cognitive Science*, 1–25.

Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing type s (sign) and type m (magnitude) errors. *Perspectives on Psychological Science*, *6*(9), 641–651.

Gibson, E. (2000). Dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain* (pp. 95–126). Cambridge, MA: MIT Press.

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science*, *29*, 261–291.

Hale, J. (2001). A probabilistic early parser as a psycholinguistic model. In *Naacl '01: Second meeting of the north american chapter of the association for computational linguistics on language technologies 2001* (pp. 1–8). Morristown, NJ, USA: Association for Computational Linguistics.

Husain, S., Vasishth, S., & Srinivasan, N. (2014). Strong expectations cancel locality effects: Evidence from Hindi. *PLoS ONE*, *9*(7), e100986.

Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research*, *29*(6), 627–645.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*, 1126–1177.

Levy, R., & Keller, F. (2013). Expectation and locality effects in German verb-final structures. *Journal of Memory and Language*, *68*, 199–222.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*(3), 375–419.

Linzen, T., & Jaeger, T. F. (2016). Uncertainty and expectation in sentence processing: Evidence from subcategorization distributions. *Cognitive Science*, *40*(6), 1382–1411. Retrieved from http://dx.doi.org/10.1111/cogs.12274 doi: 10.1111/cogs.12274

Safavi, M. S., Husain, S., & Vasishth, S. (2016). Dependency resolution difficulty increases with distance in persian separable complex predicates: Implications for expectation and memory-based accounts. *Frontiers in Psychology*(7).

Stan Development Team. (2016). *RStan: the R interface to Stan.* Retrieved from http://mc-stan.org/ (R package version 2.14.1)