Replication attempt 1 of Levy and Keller 2013

Shravan Vasishth

2 Dec 2015, Tokyo

Introduction

Levy and Keller 2013 Expt 1 design:

1. Nsubj Nacc Verb
2. Nsubj Adj Nacc Verb
3. Nsubj Ndat Nacc Verb
4. Nsubj Adj Ndat Nacc Verb

Main result of interest: d faster than c.

Introduction

Levy and Keller 2013 Expt 2 design:

1. Nsubj relpron Nacc Verb
2. Nsubj relpron Adj Nacc Verb
3. Nsubj relpron Ndat Nacc Verb
4. Nsubj relpron Adj Ndat Nacc Verb

Result: d slower than c.

Introduction

- In Levy and Keller 2013, in Expt 1, in conditions c,d, we see a facilitation at the verb when the sentences are main clauses, but in Expt 2 we see a slowdown at the verb when the sentences are embedded clauses
- We tried to replicate this effect, focusing only on conditions c and d of the two experiments.
- The prediction is a cross-over interaction.

Replication attempt of Levy and Keller 2013

- 28 subjects as in Levy and Keller
- Their items and fillers. However, some of their items had mistakes/problems, these were fixed.

item 5c "...der Vorstand, der dem senilen Vorstand ..."

item 6c "...hat der Verdächtige (...) den Fall abgeschlossen". A suspect usually doesn't solve a case.

- A filler experiment had obviously different items, removed.

Design

2×2: Main vs embedded clause; intervener Dat vs Dat+PP

1. Nsubj Ndat Nacc Verb
2. Nsubj Adj Ndat Nacc Verb
3. Nsubj relpron Ndat Nacc Verb
4. Nsubj relpron Adj Ndat Nacc Verb

Prediction: at verb, b faster than a, but d slower than c (interaction)

Results: question-response accuracies

Accuracy on target sentences is remarkably low:

```
round(100*with(qdat,tapply(correct,expt,mean)),digits=0)
```

```
##    filler      Kevy practice
##        91        63       96
```

Results: question-response accuracies
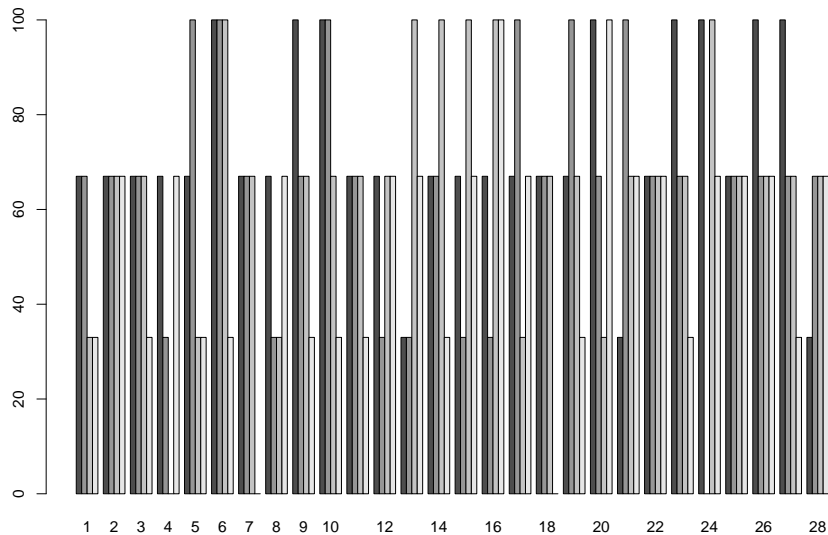
Accuracy for the target items by condition:

```
(qmeans<-round(100*with(qdatkevy,
                        tapply(correct,cond,mean)),
     digits=0))
```

```
##  a  b  c  d
## 73 64 65 51
```

They seem to be at chance with condition d. (But note we have only 3 data points per subject per condition.)

Results: question-response accuracies by subject

Results: question-response accuracies

```
## Loading required package: Matrix
```

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.70	0.32	2.16	0.03
load	-0.27	0.13	-2.10	0.04
dist	-0.30	0.13	-2.35	0.02
load:dist	-0.07	0.13	-0.52	0.60

Discussion of question-response accuracies

- Increasing distance causes a reduction in accuracy
- Increasing load causes a reduction in accuracy
- Our mean accuracies by condition are much lower than Levy and Keller's.

Discussion of question-response accuracies

Levy and Keller Expt 1: "Per-participant accuracy ranged from 67.6% to 89.2%, with a mean of 79.4%. Accuracy did not differ significantly across lists."

Levy and Keller Expt 2: "Per-participant accuracy ranged from 62.5% to 96.0%, with a mean of 80.0%. Accuracy did not differ significantly across lists."

I think they are referring to all items, including fillers.

Discussion of question-response accuracies

A comparable summary to theirs from our data:

```
meansallsubj<-round(100*with(qdat,tapply(correct,subj,mean)
## average accuracy:
mean(meansallsubj)
```

```
## [1] 82.5
```

```
## min and max:
min(meansallsubj); max(meansallsubj)
```

```
## [1] 74
```

```
## [1] 92
```

Look at the by condition accuracies again.

Discussion of question-response accuracies

```
## by condition
qmeans
```

```
##  a  b  c  d
## 73 64 65 51
```

```
## min and max among subjects:
min(meanssubj); max(meanssubj)
```

```
## [1] 0
```

```
## [1] 100
```

Discussion of question-response accuracies

- Probably their subjects' accuracies by condition were also very low?
- Implications: how much are these subjects understanding condition d? What can we conclude from reading times when accuracy in condition d is 51%?

Results: reading times

We will analyze data at the region before the verb, the verb, and the region following the verb.

Results: reading times
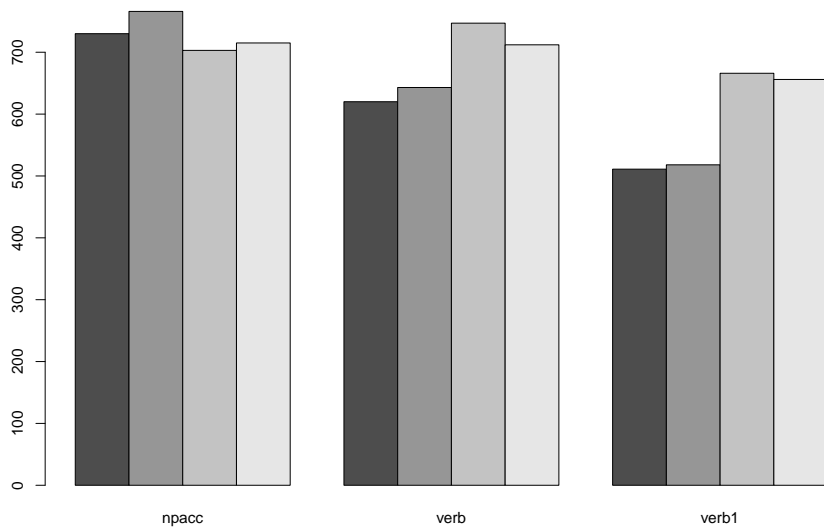
Results: reading times at pre-critical region

	Estimate	Std. Error	t value
(Intercept)	6.53	0.05	142.88
load	-0.02	0.01	-2.58
dist	0.02	0.01	1.18
load:dist	-0.01	0.01	-0.49

Results: reading times at verb

|	Estimate	Std. Error	t value
(Intercept)	6.44	0.05	132.08
load	0.06	0.01	5.39
dist	0.00	0.01	0.20
load:dist	-0.00	0.01	-0.30

Results: reading times at region after verb

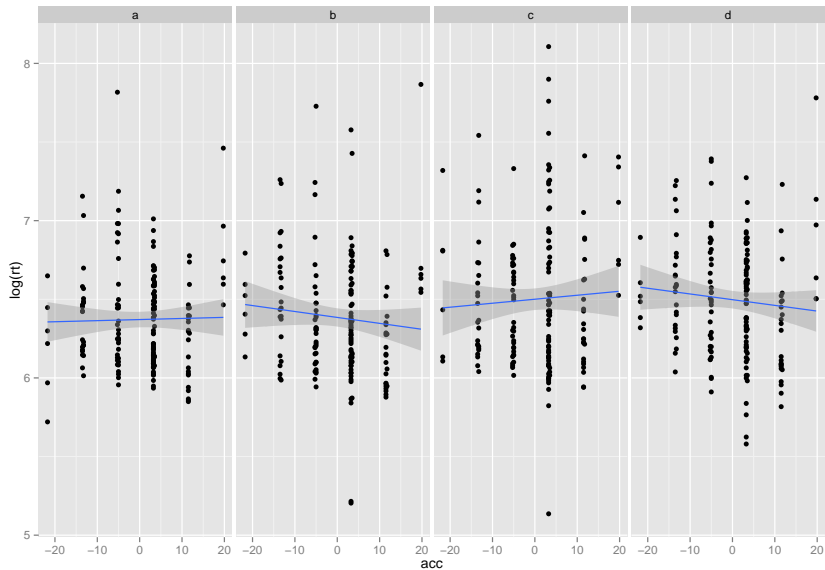|	Estimate	Std. Error	t value
(Intercept)	6.32	0.03	193.36
load	0.11	0.01	7.99
dist	0.00	0.01	0.39
load:dist	-0.00	0.01	-0.35

Results: secondary analysis

Now we can explore our data. Maybe we can take accuracy of subjects by condition into account?

Results: secondary analysis

red is the long distance condition
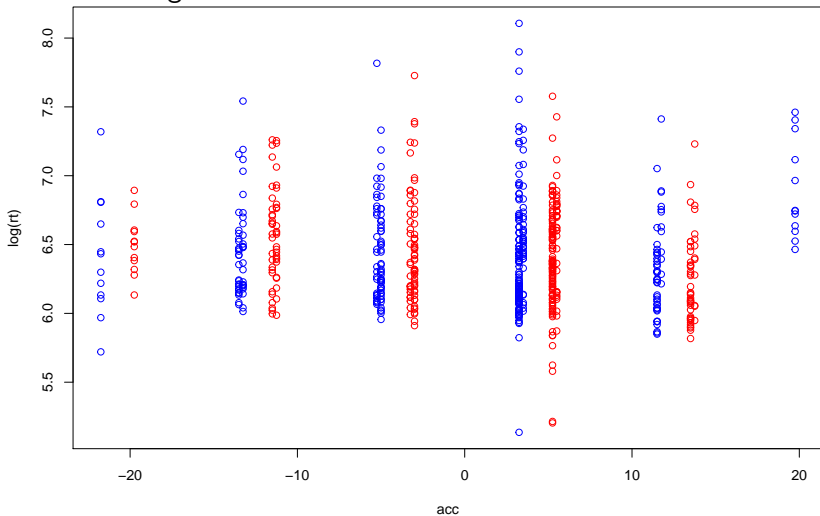
Results: secondary analysis

Results: secondary analysis

We see at verb a distance x accuracy interaction: if distance is short, for higher accuracy we get slower rt. If distance is long, for higher accuracy we get faster rt (possibly spurious, see plot).

	Estimate	Std. Error	t value
(Intercept)	6.44	0.05	129.75
load	0.06	0.01	5.31
dist	0.00	0.01	0.23
acc	-0.00	0.01	-0.21
load:dist	-0.00	0.01	-0.36
load:acc	0.00	0.00	0.28
dist:acc	-0.00	0.00	-2.11
load:dist:acc	0.00	0.00	0.04

Discussion

- The main effect of load is uninteresting (length of high load verb is longer, as it includes the auxiliary)
- We failed to find any effect of distance
- We failed to find any interaction between load and distance
- If anything, the tendency is towards facilitation in d vs c, and slowdown in b vs a—the opposite pattern to that expected from Levy and Keller's argument.
- Discussion for today: What shall we conclude from this?

Discussion

- It may be worth redoing the study with comprehension questions for all items and fillers.
- It may be worth doing the Levy and Keller experiment exactly as they did it (no changes at all).