Understanding language comprehension: the challenge of measuring individual differences within experimental designs

Lena Blott[1], Anna Gowenlock[1], Kate Nation[2], & Jennifer Rodd[1]

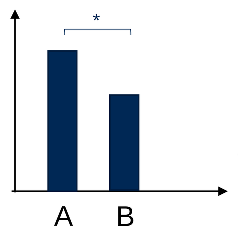
[1] Department of Experimental Psychology, University College London; [2] Department of Experimental Psychology, University of Oxford

Background

- Individual differences in language comprehension are important for theory development, and can help answer questions about classic debates in psycholinguistics, e.g. relative contributions of domain-general vs domain-specific factors to processing (see e.g. Kidd et al., 2018)
- Findings at the group-level suggest that people sometimes find it hard to access relevant meanings of ambiguous words, especially when they are *subordinate* (Duffy et al., 1988; Foss, Bever, & Silver, 1968; Rodd, Johnsrude, & Davis, 2010; and many more…)

- We know less about individual differences in using context to resolve lexical-semantic ambiguity (but see e.g. Norbury, 2005; Nation & Snowling, 1998; Khanna & Boland, 2010)

The experimental research tradition

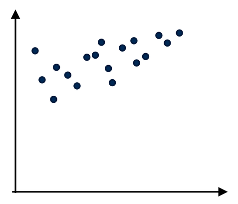


Focus on within-participant variance

Minimise variability between participants
(all participants showing a similar effect is a good thing!)

Reliability = the task consistently shows similar effect size at the group level

The individual differences research tradition



Focus on between-participant variance

Maximise variability between participants
(different participants should show different effects/scores)

Reliability = the measure consistently rank-orders individuals

Can we develop a measure that reliably captures individual differences in the ability to use context to resolve lexical-semantic ambiguity?

Participants

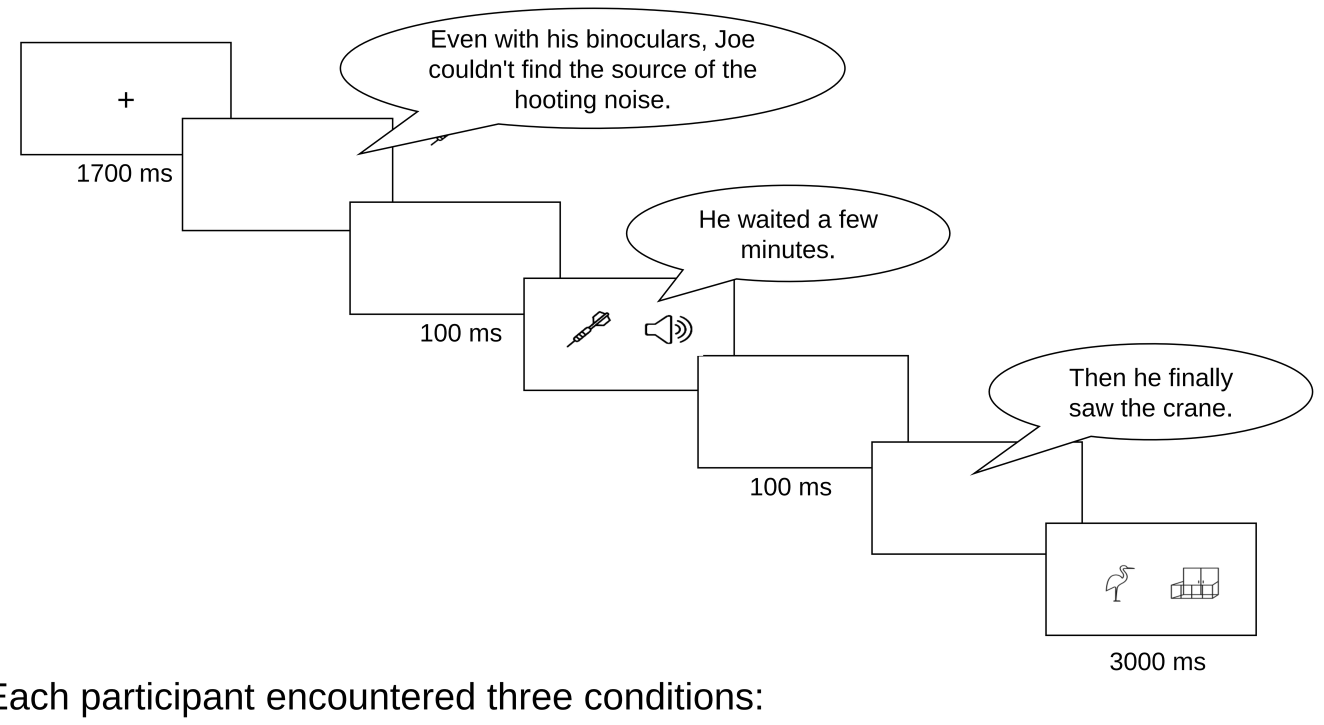
	List A (n=26)	List B (n=24)	Total (n=50)
Male (n (%))	15 (57.7%)	14 (58.3%)	29 (58%)
Female (n (%))	11 (42.3%)	10 (41.7%)	21 (42%)
First language			
British English (n (%))	25 (96.2%)	24 (100%)	49 (98%)
Any other English (n (%))	1 (3.85%)	n/a	1 (2%)
Dominant language			
British English (n (%))	25 (96.2%)	24 (100%)	49 (98%)
Any other English (n (%))	1 (3.85%)	n/a	1 (2%)
Monolingual (n (%))	22 (84.6%)	22 (91.7%)	44 (88%)
Bi- (or multi-)lingual (n (%))	4 (15.4%)	2 (8.3%)	6 (12%)
Age			
Mean (SD)	29.5 (6.95)	30.4 (7)	29.96 (6.92)
Median	31	30.5	30.5
Range	18-38	18-41	18-41

Prolific

For this pilot study, we recruited 50 adult participants, who completed a web-based disambiguation task.

Methods

Test of disambiguation ability: listen to a short narrative and choose appropriate picture



GORILLA

Each participant encountered three conditions:

Ambiguous (N=66)
Even with his binoculars, Joe couldn't find the source of the hooting noise.
He waited a few minutes.
Then he finally saw the **crane**.

Matched control (N=66)
Even with his binoculars, Joe couldn't find the source of the hooting noise.
He waited a few minutes.
Then he finally saw the **heron**.

Unmatched control (N=66)
The baby didn't want to play with any of her own toys.
Everything else was more interesting.
She managed to steal her mum's **phone**.

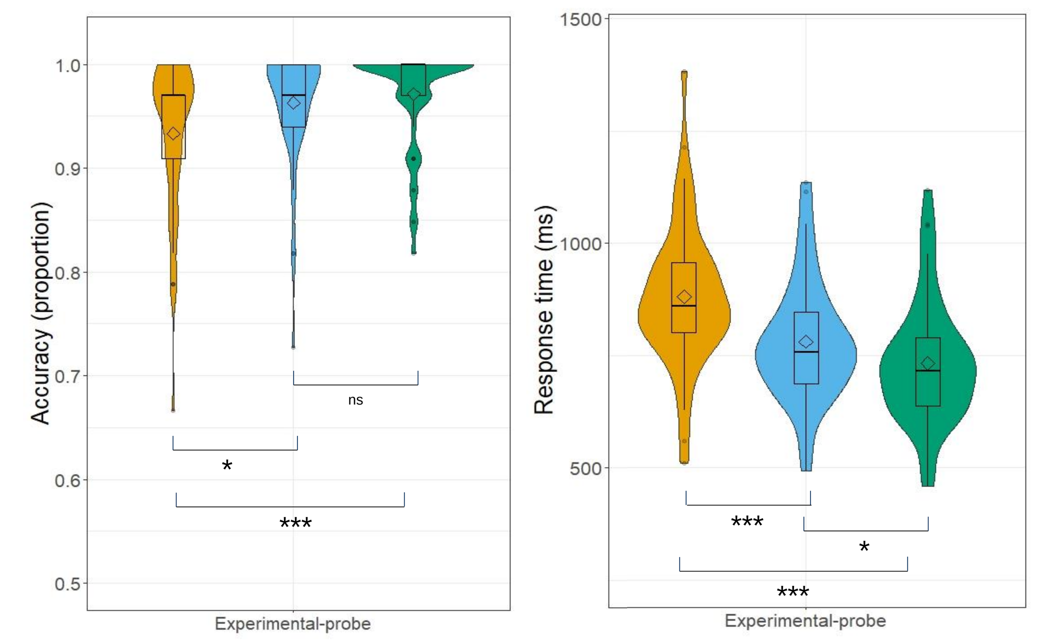
- Comparison of Ambiguous vs matched condition allows for statistical inferences about the effect of ambiguity, whilst controlling for narrative- or picture-related factors, at the *group-level*
- In a within-participant individual differences design, we need to compare Ambiguous vs a condition that is unmatched

	Ambiguous	Matched control	Unmatched control
Narrative context Story naturalness ratings on a 1-7 Likert scale, t(130) = 0.02, p=0.99, N = 60	5.27 (1.66)	5.27 (1.66)	5.27 (1.67)
Final word Fit between final word and remainder of the narrative, based on LSA, F(2,191)=0.153, p=0.86	0.07 (0.11) crane	0.07 (0.09) heron	0.08 (0.10) phone (matched on word-form frequency, AoA, number of syllables)
Picture representativeness Ratings of how well the picture represents the word's meaning on a 1-7 Likert scale, F(2, 195) = 15.05, p<.001, N = 90	4.93 (1.78)	5.16 (1.78)	5.74 (1.48)

Work is ongoing to untangle effects of ambiguity from nuisance differences between conditions.

Results

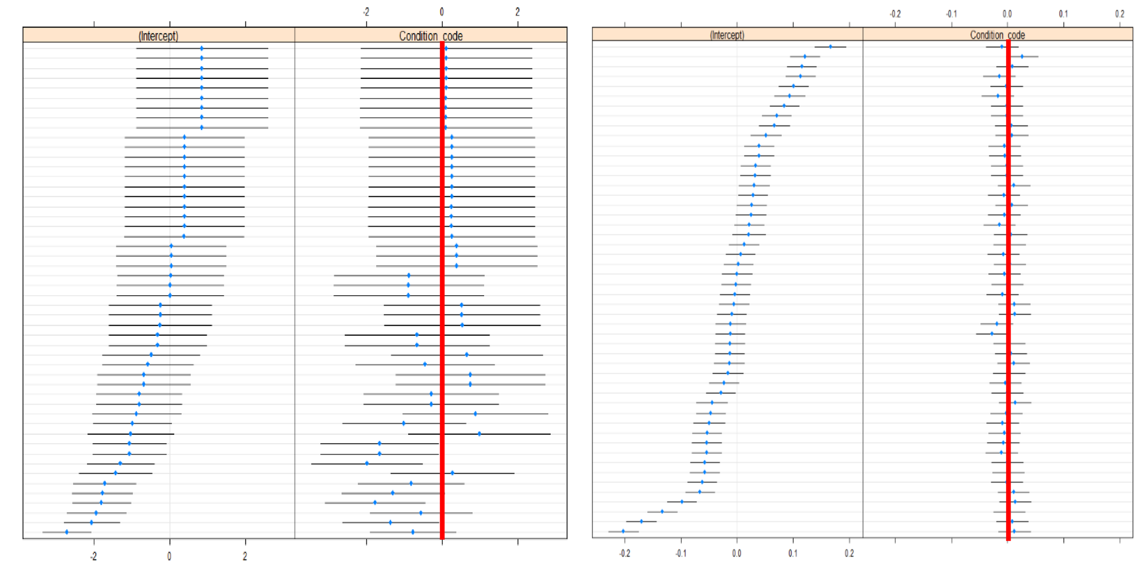
Group-level†



Compared to either control condition, Ambiguous condition had lower accuracy and slower RTs on correct trials, replicating earlier findings of ambiguity-related difficulty in language processing

† Inference based on mixed effects models with fixed effects of Ambiguity and mean picture representativeness as covariate, maximal random effects structure for participants and items

Individual differences††



Condition slope for accuracy, p = .011 Condition slope for response times, p = .044

Effects of ambiguity on accuracy (and, to a lesser extent, response times) were characterised by individual differences

†† Figures show conditional modes for each participant, extracted from mixed effects models. Inference based on comparisons of mixed effects models with varying random effects structures.

Reliability

Difference scores between Ambiguous and Unmatched control

Using 5000 random splits, Spearman-Brown corrected reliability estimates are relatively low:

	Accuracy	RT (log-transformed)
List 1	0.43, 95% CI [-0.08, 0.75]	0.27, 95% CI [-0.2, 0.63]
List 2	0.44, 95% CI [-0.08, 0.77]	-0.09, 95% CI [-0.54, 0.44]

Instead of traditional, aggregated scores, rely on **trial-level** data and a mixed modelling approach

Recommendations for measuring individual differences within experimental designs (see also Hedge et al., 2018; Goodhew & Edwards, 2018; Parsons et al., 2019; Rouder & Haaf, 2018)

- Isolate process of interest, e.g. with a comparison between well-matched conditions
- Design items to tap into the same construct and cover a range of difficulty
- Avoid adding noise that is unrelated to the process of interest, e.g. by presenting the same list of items in the same order to all participants
- Aim for the highest N of trials you can get away with
- Pilot the paradigm and desired outcome measures carefully, and check that your measure is sensitive to individual differences
- Report reliability of your measure(s) in each sample you use them on
- Make use of trial-level data
- If possible, use a latent variable approach to minimise measurement error

@LenaMBlott

lena.blott.12@ucl.ac.uk

References

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language, 27*(4), 429–446. thttps://doi.org/10.1016/0749-596X(88)90066

Foss, Donald J., Bever, T. G., & Silver, M. (1968). The comprehension and verification of ambiguous sentences. *Perception & Psychophysics, 4*(5), 304–306. https://doi.org/10.3758/BF03210520

Goodhew, S. C., & Edwards, M. (2019). Translating experimental paradigms into individual-differences research: Contributions, challenges, and practical recommendations. *Consciousness and Cognition*, *69*, 14-25. https://www.sciencedirect.com/science/article/pii/S1053810018304963

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior research methods*, *50*(3), 1166-1186. https://doi.org/10.3758/s13428-017-0935-1

Khanna, M. M., & Boland, J. E. (2010). Children's use of language context in lexical ambiguity resolution. *Quarterly Journal of Experimental Psychology*, *63*(1), 160–193. https://doi.org/10.1080/17470210902866664

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual Differences in Language Acquisition and Processing. Trends in Cognitive Sciences, 22(2), 154–169. https://doi.org/10.1016/j.tics.2017.11.006

Nation, K., & Snowling, M. J. (1998). Individual Differences in Contextual Facilitation: Evidence from Dyslexia and Poor Reading Comprehension. Child Development, 69(4), 996–1011. https://doi.org/10.1111/j.1467-8624.1998.tb06157.x

Norbury, C. F. (2005). Barking up the wrong tree? Lexical ambiguity resolution in children with language impairments and autistic spectrum disorders. *Journal of Experimental Child Psychology*, *90*(2), 142–171. https://doi.org/10.1016/j.jecp.2004.11.003

Parsons, S., Kruijt, A. W., & Fox, E. (2019). Psychological science needs a standard practice of reporting the reliability of cognitive-behavioral measurements. Advances in Methods and Practices in Psychological Science, 2(4), 378-395. https://doi.org/10.1177/2515245919879695

Rodd, J. M., Johnsrude, I. S., & Davis, M. H. (2010). The role of domain-general frontal systems in language comprehension: Evidence from dual-task interference and semantic ambiguity. *Brain and Language, 115*(3), 182–188. https://doi.org/10.1016/j.bandl.2010.07.005

Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic bulletin & review*, *26*(2), 452-467. https://link.springer.com/article/10.3758/s13423-018-1558-y