A large-scale comparison of cognitive task measures of self-regulation: raw measures vs. model parameters for individual difference analyses

A. Zeynep Enkavi, Ian W. Eisenberg, Patrick G. Bissett, Russell A. Poldrack
Department of Psychology, Stanford University

Introduction

- Psychology is rich with behavioral tasks measuring of impulsivity, self-control, inhibition, delay discounting
- These measures are assumed to capture trait-like individual differences without evaluating their stability over time
- We compare two common measure types from a large battery of behavioral tasks to determine best trait measures and their features

Methods

- 14 tasks from larger battery on reliability of self-regulation measures[1,2] (N=150): N-back, ANT, choice RT, directed forgetting, DPX, local global, recent, probes, shape matching, simon, stop signal (x3), stroop, cued task switching
- Raw measures: RT and accuracy
- 2 types of DDM: EZ and HDDM
- Non-contrast measures = use all trials; contrast variables = subtraction of two conditions; condition variables = subset of trials

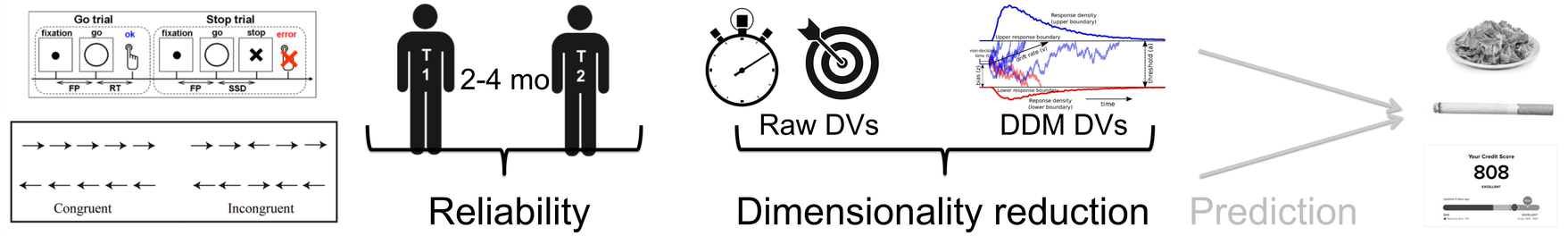
Conclusions

- DDM parameters show similar reliability to RT and accuracy
- Reliability estimates stabilize n>15
- Hierarchical estimates do not change parameter value or reliability
- Applying the same model across tasks yields 1. interpretable measures that 2. reduce to lower and more reliable trait measures

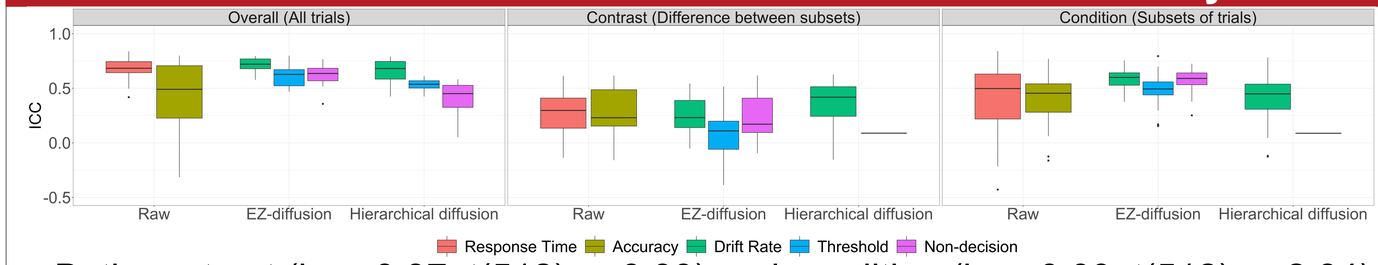
References

Eisenberg, I., Bissett, P., Enkavi, A. Z., Li, J., MacKinnon, D., Marsch, L., & Poldrack, R. (2018). Uncovering mental structure through data-driven ontology discovery
Enkavi, A. Z., Eisenberg, I., Bissett, P., Mazza, G. L., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. (2018). A large-scale analysis of test-retest reliabilities of self-regulation measures.

Overview of Procedure

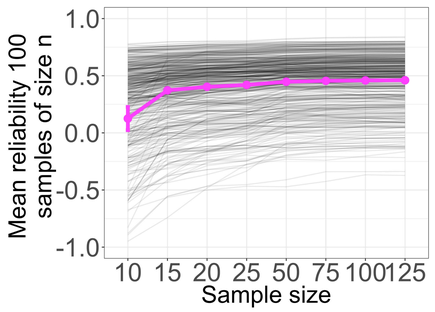


DDM vs Raw measure reliability



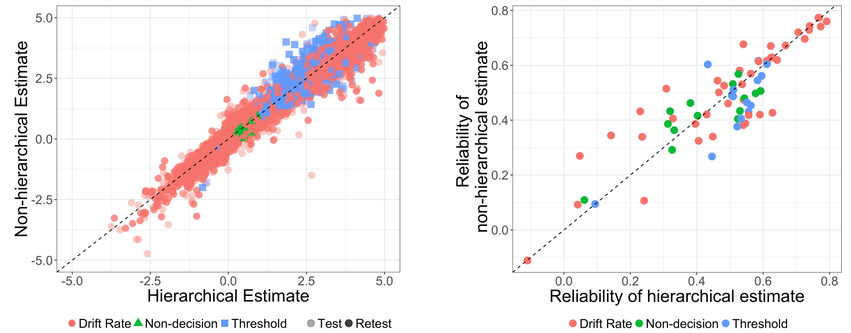
- Raw measures are comparable in reliability to DDM parameters (b = -0.03, t(512) = -0.83)
- Both contrast (b = -0.37, t(512) = -9.99) and condition (b = -0.09, t(512) = -2.84) measures are less reliable than measures that use all trials

Sample size effects on reliability



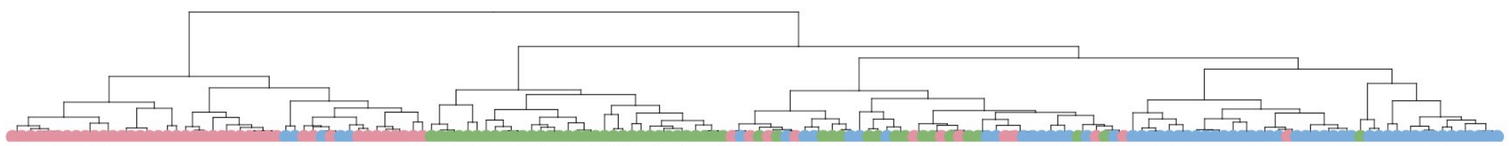
- Are pilot studies helpful in choosing trait variables?
- Yes, BUT samples <15 yield too variable and lower reliability estimates (b=0.001, t(505)=4.92)
- Conclusion does not change depending on measure type (raw vs. ddm, all trials vs. contrasts)

Analysis of hierarchical estimates



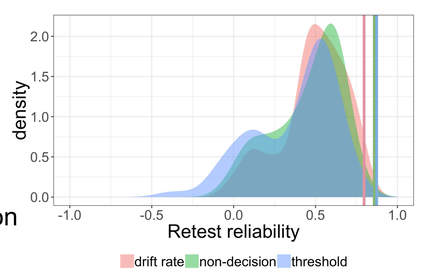
- No systematic difference in parameter estimate or reliability using hierarchical estimates

Measures vs. lower dimensional projections as trait measures



EZ measures from T1 (164 measures, n=552) → PCA → Hierarchical clustering of factor *loadings*

Predict factor *scores* from T1 PCA for T2 → ICC for the 3 PCs vs ICC of 164 EZ measures measures



- PCs are more reliable than individual measures that compose them

	Drift rate	Threshold	Non-decision
DVs (median)	0.51	0.48	0.51
PC	0.80	0.87	0.86

Contact: A. Zeynep Enkavi <zenkavi@stanford.edu>