A large-scale analysis of test-retest reliabilities of self-regulation measures

A. Z. Enkavi[1], I. W. Eisenberg[1], P. G. Bissett[1], G. L. Mazza[2], D. P. Mackinnon[2], L. A. Marsch[3], R.A. Poldrack[1]

[1]Department of Psychology, Stanford University, [2]Department of Psychology, Arizona State University, [3]Department of Psychiatry, Dartmouth College

Introduction

- Psychological literature is rich with measures of impulsivity, self-control, inhibition, delay discounting etc.
- These measures are often assumed to capture trait-like individual differences without evaluating their stability over time.
- We present a comprehensive literature review as well as novel analyses on a new large dataset containing both types of measures

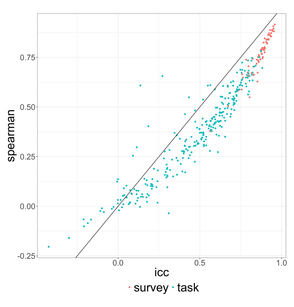
Methods

- Battery consisted of 37 cognitive tasks and 23 questionnaires putatively related to self-regulation https://expfactory.github.io/table.html
- N=150 passed QC
- Average retest delay = 115 days (range = 60 - 228 days)
- ICC's were used as the main retest reliability metric (no changes with Spearman or Pearson correlations)

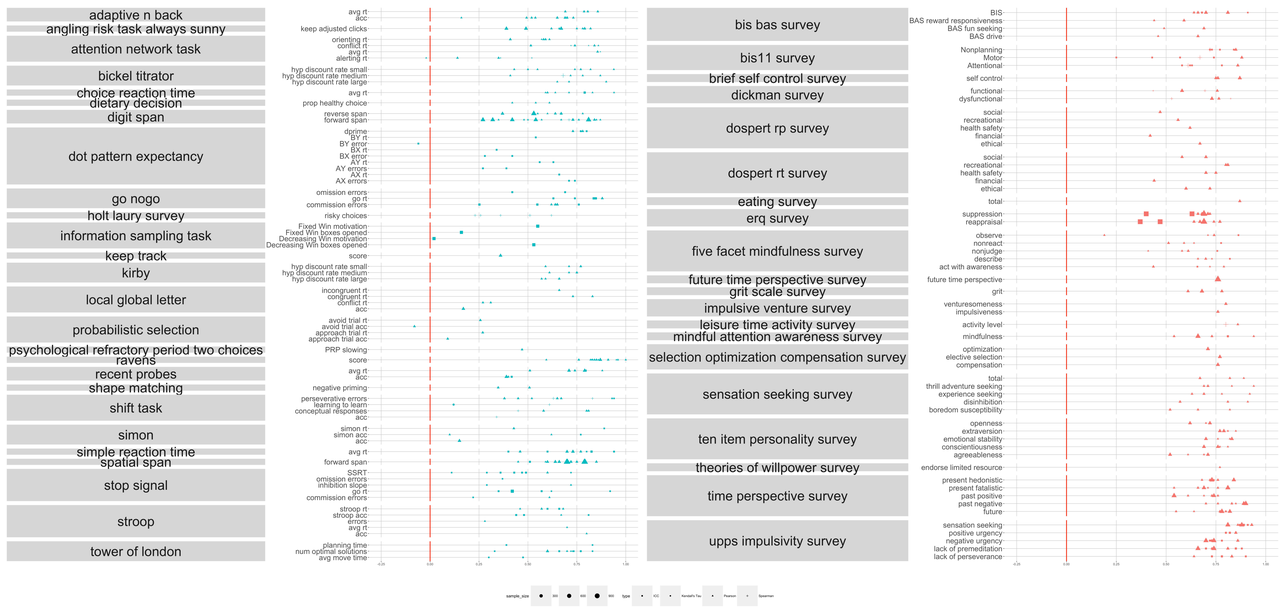
Task	Min ICC	Mean ICC	Max ICC
ravens	0.81	0.87	0.91
kirby	0.66	0.81	0.92
discount titrate	0.68	0.80	0.89
cognitive reflection survey	0.63	0.78	0.89
psychological refractory period	0.38	0.73	0.87
simple reaction time	0.55	0.73	0.86
hierarchical rule	0.41	0.70	0.85
adaptive n back	0.11	0.68	0.93
digit span	0.43	0.67	0.85
tower of london	0.25	0.66	0.96
choice reaction time	0.07	0.65	0.88
spatial span	0.28	0.63	0.83
keep track	0.41	0.63	0.75
columbia card task hot	0.14	0.62	0.91
stroop	-0.13	0.62	0.88
go nogo	0.27	0.59	0.77
information sampling task	-0.17	0.58	0.92
shift task	0.03	0.57	0.85
holt laury survey	0.29	0.56	0.72
simon	-0.28	0.56	0.84
shape matching	-0.52	0.54	0.87
stop signal	-0.49	0.52	0.85
stim selective stop signal	0.01	0.51	0.79
columbia card task cold	0.04	0.51	0.76
motor selective stop signal	-0.45	0.47	0.86
directed forgetting	-0.59	0.42	0.85
local global letter	-0.77	0.42	0.84
dietary decision	-0.08	0.40	0.74
writing task	-0.29	0.39	0.61
angling risk task always sunny	-0.54	0.38	0.73
recent probes	-0.95	0.36	0.88
dot pattern expectancy	-0.84	0.35	0.88
attention network task	-1.19	0.33	0.86
bickel titrator	-0.05	0.31	0.95
threebytwo	-0.72	0.28	0.88
probabilistic selection	-0.61	0.20	0.85
two stage decision	-0.98	0.14	0.82

Survey	Min ICC	Mean ICC	Max ICC
self regulation survey	0.93	0.95	0.97
brief self control survey	0.91	0.94	0.97
grit scale survey	0.91	0.94	0.97
ten item personality survey	0.79	0.93	0.97
time perspective survey	0.83	0.92	0.97
bis bas survey	0.81	0.90	0.97
dospert rt survey	0.79	0.90	0.97
five facet mindfulness survey	0.77	0.90	0.97
mindful attention awareness survey	0.83	0.90	0.95
erq survey	0.77	0.89	0.94
impulsive venture survey	0.75	0.89	0.97
sensation seeking survey	0.79	0.89	0.96
mpq control survey	0.81	0.88	0.93
upps impulsivity survey	0.64	0.87	0.97
eating survey	0.78	0.87	0.94
future time perspective survey	0.74	0.85	0.93
bis11 survey	0.61	0.84	0.96
dickman survey	0.60	0.84	0.93
dospert rp survey	0.69	0.84	0.93
dospert eb survey	0.58	0.80	0.90
theories of willpower survey	0.69	0.80	0.89
leisure time activity survey	0.67	0.80	0.88
selection optimization compensation survey	0.50	0.78	0.91

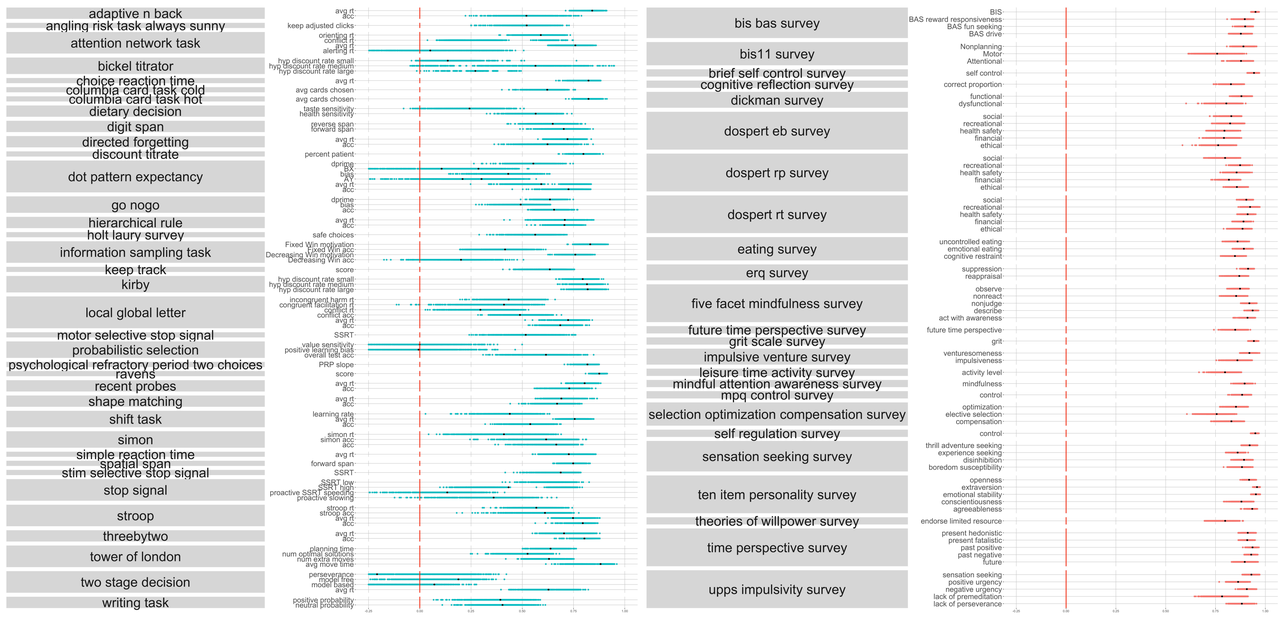
- Bootstrapped reliabilities (n=1000)



Literature review



Our results



Discussion

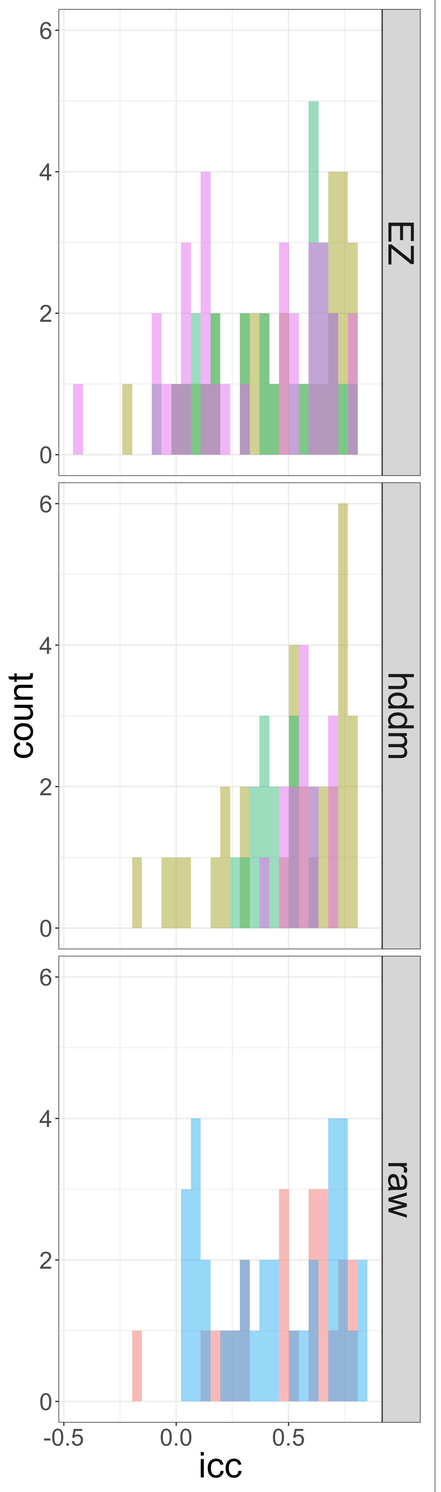
- Dependent variables from cognitive tasks show larger variability and lower reliability compared to measures from surveys
- Drift diffusion parameters show similar reliability to RT and accuracy

References

An interactive version of the literature review as well as a list of all the references can be found at goo.gl/gM7Pgr

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

Exploratory analyses



- Raw ~ HDDM > EZ (b = 0.098, t(143) = 2.17, p = 0.032; b = -0.045, t(11) = -1, p = 0.32)