Iowa Gambling Task: Comparison of the Classical Scoring and Cognitive Modeling Approach and its Convergent Validity with Other Clinical Tasks

Hynek Cígler, Martin Šmíra, Vojtěch Viktorin

FACULTY OF SOCIAL STUDIES
Masaryk University

Institute for Research on Children, Youth and Family | Masaryk University, Brno, Czech Republic

The Iowa Gambling Task (IGT) is a popular test for assessing executive functions and decision-making deficits in several clinical populations, especially if the symptomatology includes excessive risk taking. The IGT is a complex task which requires participants to correctly integrate and remember information, evaluate consequences and make a decision. However there are many papers studying criterial and convergent validaty for many purposes (for overview see Buelow & Suhr, 2009), less papers concern reliability or validity of the test scores construction. Moreover, the majority of studies uses "traditional IGT scores" mainly based on the differences between "good" and "bad" blocks or on the total net scores, although there are other modern and up-to-date methods for IGT scoring.

One of the alternative approaches is Bayesian Cognitive Modeling, for example PVL-delta model (Steingroever et al., 2013; 2014). However, these models are not widely used. High computational demands can be one of the possible reasons, other barrier is a highly theoretical nature of cognitive modeling combined with low knowledge about the parameters validity. Although there is a few studies on differences between clinical and normal population (e.g. Steingroever et al., in press), convergent validity to classical test scores or to other tests of executive functions is unknown.

The question is if the Bayesian parameters can be used in similar way as the traditional IGT scores for diagnostic purposes. The goal of this explorative study is to present results of convergent validity analysis of parameters from PVL-delta model to the classical IGT test scores and to other tasks of executive functions.

Participants and data collection

Data were collected in PC lab during approx. 3 hours between other questionnaires in two waves as a part of bigger project focused on a Need for Cognitive Closure. Unfortunately, each wave received different test battery (1st wave: DDT, GNG, SST; 2st wave: the others) – only IGT were administered to all the respondents. The total $N = 171$ respondents were aged 20–30 ($Me = 24$) with majority of women ($n = 128$; 75 %); numbers of students ($n = 82$; 48 %) and workers ($n = 89$; 52 %) were similar.

Reliability estimation

For IGT (classical scores), GNG, SST, Stroop (subscores used for total Stroop score) and NL, reliability was estimated using McDonald's omega based on the subscores/reaction times from different items/blocks of the test. For N-back, where Horn's parallel analysis suggested two factors (2back vs. 3back), McDonald's omega total was used. For DDT, we used correlation of both conditions with Spearman-Brown correction. Reliability of difference scores (e.g. total Stroop score) was estimated using reliabilities of subtests and their Pearson's correlation (e.g. Thomas & Zumbo, 2012). All the estimates can be biased due to non-normal score distributions.

Test Materials

Iowa Gambling Task (IGT): Bayesian Cognitive Modeling Approach

We used classical IGT design (e.g. Bechara et al., 1994) twice longer (200 trials). The PVL-delta model use four parameters to predict block choices:

- **Shape parameter (A)**, which determines the shape of the "utility function". Greater A-parameter means higher respondent's sensitivity to feedback outcomes.
- **Loss aversion parameter (w)**. People with low w-parameter are more sensitive to gains than to the losses, people with high w-parameter are more sensitive to losses than to gains.
- **Updating parameter (a)**, which quantifies memory and learning. Its high value indicates that the recent card outcomes have bigger influence on the expected utility of blocks and forgetting is quick.
- **Response consistecy (c)**, which captures the exploitation vs. exploration process. Low value indicates more random choices (more explorative behavior), while the high value indicates strong choices dependency on expected utility of block (more exploitative behaviour).

The PVL-delta model was estimated using Markov chain Monte Carlo (MCMC) sampling algorithm in Stan and R. The procedure is described in our previous work (Šmíra, 2014). For 22 participants, MCMC did not converged ($rhat > 1.02$), and they were excluded. All the scripts are available on-line: https://github.com/neekro/NfCC-IGT.

Reliability estimation of PVL-delta parameters

According to our knowledge, there has not been any published procedure for reliability estimation of Bayesian cognitive modeling parameters. We propose two approaches:

(1.) As a reliability is defined as the correlation between two parallel tests, we sampled a pair of each parameter from its posterior distribution for each participants, and correlated these pairs. We repeated it 1000times for each parameter what results to posterior distributions of reliability estimates. All the distributions were approx. normally distributed with means [and 95% CIs]: $r_A = .74$ [.65, .82]; $r_w = .84$ [.79, .89]; $r_a = .89$ [.83, .94]; $r_A = .87$ [.81, .91].

(2.) For each respondent (and each parameter), we computed variance of posterior distribution. Mean of these error variances is a mean-square error (MSE). The reliability is than $r_{xx'} = 1 - MSE/\mathrm{var}(X)$, where $\mathrm{var}(X)$ is a variance of point (means) parameter estimates between subjects. Because this approach assumes normality for posterior distributions, estimates are slightly smaller: $r_A = .64$; $r_w = .81$; $r_a = .87$; $r_A = .84$.

Iowa Gambling Task (IGT): Traditional IGT scores

LTC (learning of long-term consequences) is the difference between number of choices from good and bad decks: (C+D)−(A+B). **IFL** (bias for infrequent losses) is the difference (B+D)−(A+C). **PGDfr** is the preference of good deck among high frequent losses decks (C−A). **PGDinfr** was a preference of good deck among low frequent losses decks (D−B). **TOTAL** was a final total net score.

			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Classical IOWA	1	LTC	.86																									
	2	IFL	-.37**	.91																								
	3	PGDfr	.65**	-.83**	.91																							
	4	PGDinfr	.54**	.36**	-.12	.86																						
	5	FINAL	.93**	-.38**	.63**	.49**	.82																					
Bayesian IOWA	6	A	.30*	-.38**	.31**	.12	.27*	.74																				
	7	w	.58**	-.58**	.76**	.07	.58**	.20*	.84																			
	8	a	-.46*	.22**	-.43**	-.22*	-.48*	-.34**	-.60**	.89																		
	9	c	.65**	-.25*	.52**	.23*	.64**	.37**	.55**	-.87**	.87																	
DDT	10	k	-.10	-.09	.02	-.17	-.06	.04	.00	-.02	-.14	.84																
	11	AUC	.12	.09	.00	.19	.07	.01	.03	-.02	.16	-.99**	.83															
	12	RSE	-.13	.01	-.06	-.16	-.04	-.11	-.08	.09	-.20	.68**	-.74**	.64														
GNG	13	GOwrong	.02	.08	-.12	.02	.03	-.07	-.08	.09	-.08	.08	-.10	.04	.77													
	14	NOGOwrong	-.05	.23*	-.21*	.11	-.03	-.20	-.18	.28*	-.23*	-.01	.00	.05	.32*	.48												
	15	GORT	.08	.08	-.08	.14	.08	-.08	-.04	.05	-.02	.03	-.04	-.02	.81**	.01	.82											
	16	NOGORT	.07	-.07	.02	-.03	.08	-.06	-.01	-.01	.02	.15	-.17	.09	.56**	.05	.57**	.56										
SST	17	COM	-.08	-.02	-.14	-.09	-.08	-.12	-.05	.21	-.24*	-.05	.06	.03	-.02	.30*	-.11	-.22*	.59									
	18	OM	-.11	.04	-.05	-.03	-.09	.07	.08	-.11	.09	.07	-.07	.05	.17	.03	.08	.15	-.36*	.85								
	19	COR	.18	-.10	.13	.07	.15	.03	-.01	.03	.00	-.07	.07	-.08	-.18	-.13	-.08	-.11	.15	-.95**	.85							
	20	RT	-.06	.09	-.04	.02	-.05	.02	.08	-.16	.17	.03	-.03	.00	.18	-.07	.13	.19	-.56**	.84**	-.75**	.93						
	21	SSD	-.04	.05	.00	-.01	-.04	.02	.14	-.23*	.22	.14	-.13	.09	.07	-.21	.09	.15	-.49**	.84**	-.77**	.92**	.94					
	22	SSRT	.02	.05	-.13	.11	.03	.09	-.22	.24*	-.17	-.22*	.22*	-.17	.13	.30*	.05	.00	.27*	-.56**	.55**	-.49**	-.75**	.64				
other	23	N-back	.05	.02	.02	.01	.00	-.05	.03	-.15	.07														.93			
	24	stroop	-.07	.06	-.12	-.05	-.09	-.22	-.12	.12	-.11														-.15	.59		
	25	NLcor	.25*	-.11	.19	.02	.27*	.20	.21	-.15	.20														.08	.00	.94	
	26	NL	.08	-.05	.04	.00	.05	-.07	-.09	-.01	.06														-.11	.05	-.33*	.61

Note: * $p < .05$; ** $p < .05$ after Holm–Bonferroni correction. Other: N-back (N-back score), Stroop task (stroop score) and Number-Letter (NLcor and NL scores). Reliabilities on diagonal.

Delay Discouting Task (DDT)

We used two conditions, 990 CZK (37 EUR) or 24900 CZK (950 EUR), and week/month/three month/one year delays. For both conditions **k**-parameter (e.g. Mazur, 1987) and **AUC** was estimated alongside **RSE** (fit of k parameter to the data; higher value suggests lower fit). Higher values of k and lower values of AUC suggest preference of small amount now compared to higher reward later. Total scores were means of parameters from both conditions.

Go/No-Go Task (GNG)

We used 192 trials with two conditions divided into four blocks. Scores: relative wrong frequency on go (**GOwrong**) and no-go tasks (**NOGOwrong**); mean reaction time on go (**GORT**) and no-go (**NOGORT**) tasks.

Stop Signal Task (SST)

192 trials divided into four blocks were used. Scores: relative frequencies of commission (**COM**), omission (**OM**) and correct (**COR**) responses alongside with mean reaction time (**RT**) on go tasks, stop-signal delay (**SSD**) and stop-signal reaction time (**SSRT** = RT-SSD). We worked also with relative frequency of fail on no-go tasks, but its blocks were not positively correlated ($\omega_{FAIL} = .01$) and therefore this score was excluded.

N-Back

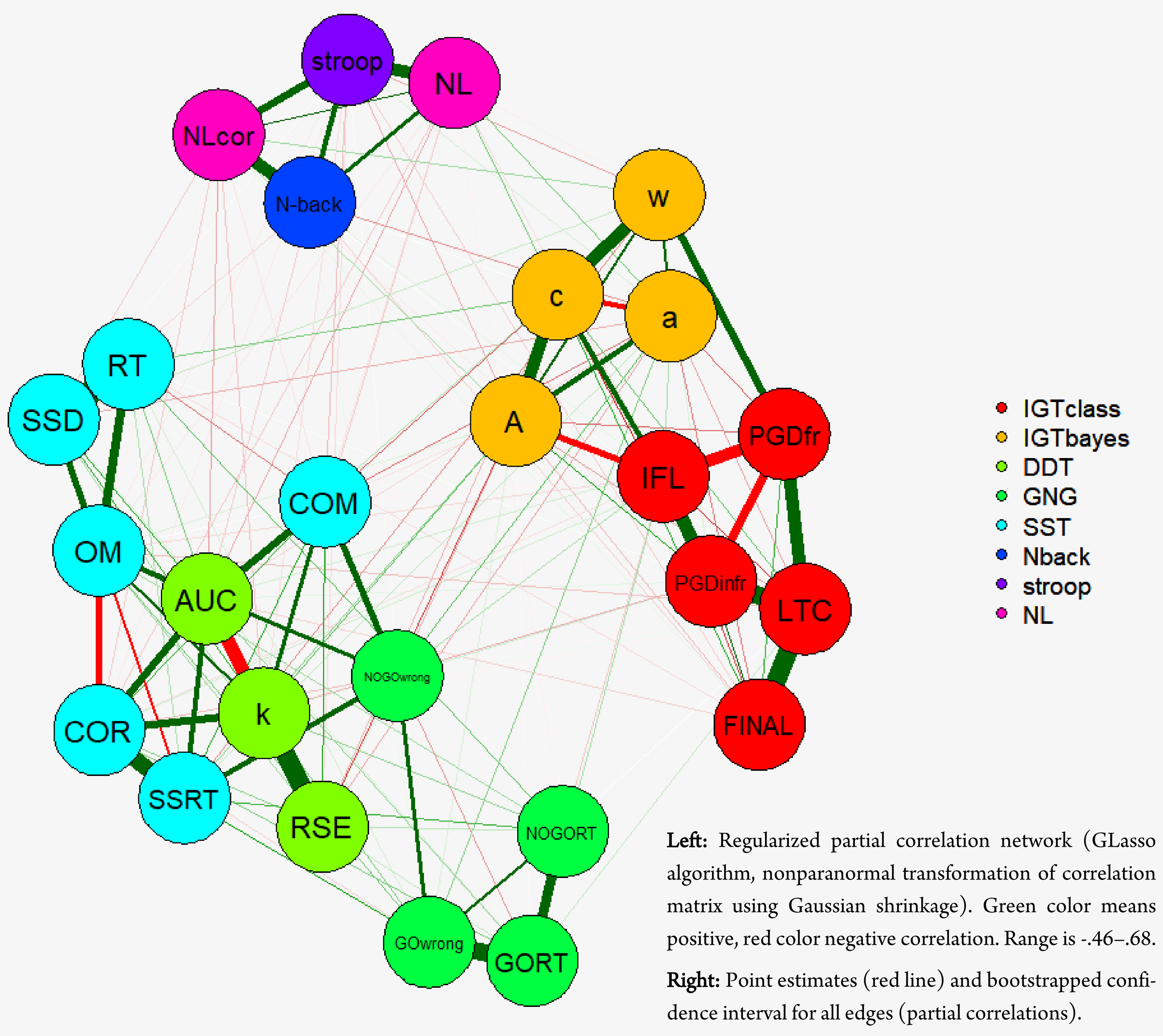
We used 2-back and 3-back tasks, both divided into four blocks with 24 tasks each (6 "yes" conditions). Score (**N-back**) was defined as the total correct of the "yes" trials.

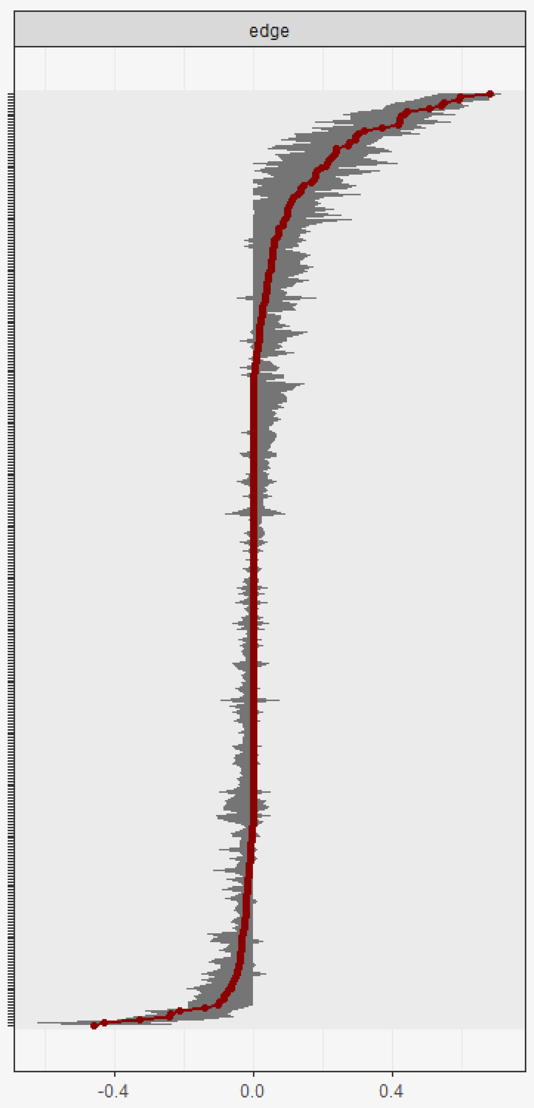
Spatial Stroop Task (Stroop)

Position on the screen was combined with words up/down/left/right, respondents used arrow keys. There were four blocks: neutral spatial, neutral reading, congruent and incongruent, where respondents had to mark the sense of the word, however its position was different. The total score (**Stroop**) was estimated as the difference between the mean reaction time in the 4th and 2nd block.

Number-Letter Task (NL)

Respondents were exposed to the pairs of a numbers and a letters. Based on a top/down position, they had to decide if the letter is consonant/vowel, or if the number is odd/even. One third of 48 trials was congruent (both conditions led to the same right key). Scores: total correct (**NLcor**) and the difference of the mean reaction time between switch and repeat trials (**NL**).





Left: Regularized partial correlation network (GLasso algorithm, nonparanormal transformation of correlation matrix using Gaussian shrinkage). Green color means positive, red color negative correlation. Range is -.46–.68.

Right: Point estimates (red line) and bootstrapped confidence interval for all edges (partial correlations).

Results and Discussion

Left-bottom table contains Spearman correlations and reliabilities of all the scores. We estimated also network model (chart). We used regularized partial correlation network (gLASSO algorithm, EBIC minimized) in bootnet R package (Epskamp, Borsboom, & Fried, 2017) with nonparanormal transformation to Gaussian distribution (huge package). Be aware that the original correlation matrix is not fully interconnected and correlations between the two blocks of methods are just a product of transformation. These results are appropriate for exploratory purposes only.

Bayesian cognitive scores correlates highly with classical test score, but nor classical nor Bayesian IGT score are well correlated to other cognitive tasks; however, correlations between different tests are generally low. Average of Fisher-z-transformated absolute values of correlations with other tests was higher for Bayesian scores (.104) than for traditional scores (.078), $t(123.5) = 2.34$, $p = .021$, though the effect was small, Cohen's $d = .39$.

Conclusion: Bayesian scores seem to have slightly higher concurrent validity to other cognitive tasks, however correlations are generally very small. Also tests which are supposed to measure the same traits, e.g. NL and Stroop or SST and GNG, share only a small amount of variance. However, Bayesian IGT cognitive parameters correlated with classical IGT scores and can be probably used in similar way. Finally, a and c parameters were highly correlated (especially if we take into account their reliabilities), they comparably correlate with other tests, and they probably measure the same traits. We advice to use their mean (after reverting one of them) for practical purposes and probably merge them in next generation of PVL-delta model.

References

Buelow, M. T., & Suhr, J. A. (2009). Construct Validity of the Iowa Gambling Task. *Neuropsychology Review, 19*(1), 102–114.

Steingroever, H., Wetzels, R., & Wagenmakers, E.-J. (2014). Absolute performance of reinforcement-learning models for the Iowa Gambling Task. *Decision, 1*(3), 161–183.

Steingroever, H., Wetzels, R., & Wagenmakers, E.-J. (2013). A Comparison of Reinforcement Learning Models for the Iowa Gambling Task Using Parameter Space Partitioning. *The Journal of Problem Solving, 5*(2), 1–32.

Steingroever, H., Pachur, T., Šmíra, M., & Lee, M. D. (in press). Bayesian techniques for analyzing group differences in the Iowa Gambling Task: A case study of intuitive and deliberate decision makers. *Psychonomic Bulletin & Review.*

Šmíra, M. (2014). *Cognitive modeling of the performance on the Iowa Gambling Task: an application to Parkinson's disease patients' data*. Masaryk University. Retrieved from https://is.muni.cz/th/439671/fss_b/Finalalalal.pdf

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition, 50*(1–3), 7–15.

Thomas, D. R., & Zumbo, B. D. (2012). Difference Scores From the Point of View of Reliability and Repeated-Measures ANOVA. *Educational and Psychological Measurement, 72*(1), 37–43.

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. In M. L. Commons, J. E. Mazur, J. A. Nevin, & H. Rachlin (Eds.), *Quantitative analysis of behavior: Vol. 5. The effect of delay and of intervening events of reinforcement value* (pp. 55–73). Hillsdale, NJ: Erlbaum.

Contact

Institute for Research on Children, Youth and Family, & Department of Psychology
Masaryk University, Faculty of Social Studies
Joštova 10, 602 00 Brno, Czech Republic
web: http://ivdmr.fss.muni.cz
e-mail to the authors:
hynek.cigler@mail.muni.cz

Project was supported by the Czech Science Foundation (GA15-20970S).