OXFORD
UNIVERSITY PRESS

Archives of Clinical Neuropsychology 33 (2018) 502–507

Archives of CLINICAL NEUROPSYCHOLOGY

An Initial Examination of Performance on Two Versions of the Iowa Gambling Task

Melissa T. Buelow[1,*], Wesley R. Barnhart[2]

[1]*Department of Psychology, The Ohio State University Newark, Newark, OH 43055, USA*
[2]*ID/ASD Research Group, Nisonger Center, University Center for Excellence in Developmental Disabilities, Columbus, OH 43210, USA*

*Corresponding author at: Department of Psychology, The Ohio State, University Newark, 1179 University Drive, Newark, OH 43055, USA.
Tel.: +1-740-755-7808; fax: +1-740-366-5047.
E-mail address: buelow.11@osu.edu (M.T. Buelow).

Editorial Decision 28 September 2017; Accepted 3 October 2017

Abstract

Objective: To examine differences between two versions of the Iowa Gambling Task (IGT).
Method: A total of 282 undergraduate students completed one of two versions of the IGT: the original version from 1994 (n = 132), or the 2007 version available through Psychological Assessment Resources (PAR) (n = 150).
Results: PAR (2007 version) IGT participants decided more advantageously (i.e., selected more from the small immediate reward but long-term positive/gain decks than the large immediate reward but long-term negative/loss decks) than original IGT participants during Trials 21–60. This difference was likely due to fewer Deck B selections by the PAR IGT participants during the early (Trials 1–40) and later (Trials 41–100) trials.
Conclusions: The PAR IGT may result in a greater ability to make future-oriented, advantageous decisions more quickly than on the original IGT. Implications for future assessment of decision-making impairments in clinical and research settings are discussed.

Keywords: Decision-making; Validity; Iowa gambling task

Introduction

Decision-making involves a choice between two or more options. Risky decision-making is further defined as continued disadvantageous decisions in spite of knowledge of risks (Bechara, 2007). Evidence of risky decision-making exists across multiple disorders (Buelow & Suhr, 2009); however, to determine decision-making deficits in patient and non-patient samples, it is important to ensure accurate assessment of decision-making. When multiple versions of a task exist, equivalency between those versions is imperative.

The Iowa Gambling Task (IGT; Bechara, 2007; Bechara, Damasio, Damasio, & Anderson, 1994) is the most common decision-making measure used by clinicians and researchers alike (see Buelow & Suhr, 2009, for review). The original IGT (Bechara et al., 1994) utilized four decks of paper cards and a set of play money. Participants were given $2000 and told to maximize profit by selecting 100 cards from four decks (A,B,C,and D). Participants were not told the relative risks/benefits of each deck, learning this information through feedback. Only after each selection did the participant become increasingly aware of the relative risks/benefits and, over time, this information informed about the long-term effects of each deck. Decks A and B are disadvantageous, as they result in high immediate gains ($100 on every selection from A or B) but long-term negative outcomes. In the original IGT, participants experienced five losses per 10 selections from Deck A (with specific losses ranging from $150 to $350), and one loss per 10 selections from Deck B (always a loss of $1250). Decks C and D are advantageous, as they result in lower immediate gains ($50 on every selection from C or D) but long-term positive outcomes. Participants experienced five losses per 10 selections from Deck C (with specific losses ranging from $25 to $75), and one

doi:10.1093/arclin/acx103 Advance Access publication on 27 October 2017

loss per 10 selections from Deck D (always a loss of $250). After 10 selections, participants incurred a net loss of $250 if they selected from A and B, and a net gain of $250 from C and D.

The IGT underwent several changes since its initial conception. The IGT is currently computer administered, and has also been published through Psychological Assessment Resources (PAR) (Bechara, 2007). Of note, an updated version of the PAR IGT was recently released (IGT2), with the primary changes being in the normative data and administration age range. Participants again start with $2000 and maximize profit over 100 trials/selections. Decks A and B are still disadvantageous and Decks C and D advantageous, but the amounts gained/lost vary. At first, selections from Deck A result in an average win of $100 and an average loss of $1250 (5/10 selections result in a loss). But, as participants continue to select from Deck A, the average win increases in increments of $10 while the average loss increases both in increments of $250 and in the total number of losses (6/10 selections, 7/10 selections, etc.). For Deck B, the win average increases by $10 as the task progresses and the losses average $1250 (increasing in increments of $250 while only occurring on 1/10 selections). In sum, as participants continue to select from Decks A and B, the immediate wins increase but the losses also accumulate, indicating long-term negative consequences of continued selections from these decks (Bechara, 2007). A similar process occurs for Decks C and D (average $50 win [increasing in $5 increments] and average $250 loss [increasing in $25 increments]). The number of losses increases for Decks A and C while remaining the same for Decks B and D.

Several differences exist between the original and revised IGT. First, on the original IGT the wins are always set to $50/$100, whereas in the revised (2007 PAR) IGT the wins average $50/$100. On the original IGT, losses are consistent across trials (e.g., after 10 trials at any point in the task, consistently choosing from Deck A results in a loss of $1250 across five random losses). On the revised IGT, the losses instead increase over time both in dollar amount and number (Bechara, 2007). This also occurs for Deck C. For Decks B and D, the average wins and losses increase with continued selections from these decks, but the loss frequency remains the same over time. On the computerized version of the original IGT, participants see a dynamic bar illustrating the cumulative effect of wins and losses. This component was retained in the PAR IGT. Finally, on the original IGT there were an unlimited number of cards in each deck, whereas on the PAR IGT participants were limited to 60 selections from one deck.

One concern in the literature is that healthy controls in more recent studies exhibit lower scores than healthy controls in earlier studies (e.g., Balodis, MacDonald, & Olmstead, 2006; Buelow & Blaine, 2015; Dymond, Cella, Cooper, & Turnbull, 2010; Steingroever, Wetzels, Horstmann, Neumann, & Wagenmakers, 2013; Tchanturia et al., 2012). In addition, differences in IGT versions could account for differences in findings across studies, an issue compounded by the lack of detail provided in many studies regarding the particular IGT version used. Authors do not always provide enough detail to determine the precise IGT version utilized in a study. Yet, inconsistencies are relatively common in the IGT literature, such that not all studies of IGT performance in individuals with schizophrenia, eating disorders, bipolar disorder, and others show impairments (e.g., Adida, Jollant, Clark, Guillaume, Goodwin, Azorin, & Courtet, 2015; Bark, Dieckmann, Bogerts, & Northoff, 2005; Buelow, Frakey, Grace, & Friedman, 2014; Clark, Iversen, & Goodwin, 2002; Euteneuer et al., 2009; Gu et al., 2013; Guillaume et al., 2010; Tchanturia et al., 2012). It is unclear to what extent the particular IGT version utilized could account for these between-study (but within-diagnosis) inconsistencies.

The present study sought to examine differences in the pattern of performance on two versions of the IGT: the original (yet computerized) IGT and the updated (2007) PAR IGT. These two versions differ in terms of task particulars, including the win/loss amounts and the pattern of losses remaining consistent versus increasing over time. These differences could in turn lead to differences in performance and could account for the decrease in healthy control participant performance in more recent studies (e.g., Steingroever et al., 2013). The present study sought to determine differences in task performance based on IGT version. Although inconsistencies have been shown across previous IGT studies, it is unclear in many cases which IGT version was utilized and thus we are unable to predict the exact differences in outcomes based on IGT version.

Methods

Participants

Participants were collected across multiple studies in which the IGT was administered. Some data (n = 118) were previously published (Barnhart & Buelow, 2017). A total of 342 individuals took part (n = 175 PAR [2007] IGT, n = 166 Original IGT); however, 60 were removed from further analyses due to self-reported diagnosis of Attention-Deficit/Hyperactivity Disorder, Traumatic Brain Injury, or a psychiatric diagnosis (specific diagnosis not recorded). The final sample included 282 psychology students (52.9% female, M_{age} = 18.72 [SD_{age} = 1.81], M_{ed} = 12.13 [SD_{ed} = 0.45], 66.4% Caucasian). Participants were randomly assigned to complete one of two versions of the IGT: the original (Bechara et al.,

1994) version or the PAR version (Bechara, 2007). There were no between-group differences in gender, ethnicity, age, or educational level, *p*s > .070.

Materials and Procedure

The study was approved by the university's Institutional Review Board, and participants provided informed consent before completing the IGT. The original IGT (Bechara et al., 1994) was administered via a computerized protocol that followed the original guidelines. The PAR IGT (Bechara, 2007) was administered following the standard computerized protocol. Due to concerns about practice effects (Bechara, 2007) and limitations of the study designs, participants completed only one of the two IGT versions (original, PAR) before being debriefed.

Data Analysis

Previous research has supported three scoring approaches on the IGT: (1) total performance across all 100 trials; (2) performance divided by 20-card blocks of trials (where disadvantageous decks [A, B] were subtracted from advantageous decks [C, D]); and (3) performance divided by early and later trials. Analyzing performance across all 100 trials has fallen out of favor compared to the other analyses. In addition, the PAR IGT score report focuses on advantageous minus disadvantageous selections across 20-card blocks of trials and the total number of selections from each individual deck (there is no score report for the original IGT). Therefore, this overall analysis was not conducted. To address the second scoring approach, a mixed ANOVA was conducted with Trials (Trials 1–20, 21–40, 41–60, 61–80, 81–100) as the within-subject variable and IGT Type (original IGT, PAR [2007] IGT) as the between-subject variable. Finally, previous research has shown IGT performance can be split into early trials, in which decisions are made without adequate knowledge about the risks/benefits of each deck (decision-making under ambiguity; Trials 1–40), and later trials, in which decisions are made after participants have learned the risks of each deck (Trials 41–100; Brand et al., 2007). Recent research has examined differences between individual decks in terms of frequencies of losses (Steingroever et al., 2013), suggesting Decks A and B are not equally disadvantageous. Therefore, we also calculated the percent selections from each individual deck during the early (Trials 1–40) and later (Trials 41–100) trials. A mixed ANOVA was conducted with Deck Type (Deck A Trials 1–40, Deck A Trials 41–100, Deck B Trials 1–40, etc.) as the within-subject variable and IGT Type as the between-subject variable.

Results

Table 1 includes the means and standard deviations for each IGT variable.

Advantageous Minus Disadvantageous Deck Comparisons

Mauchly's Test of Sphericity was significant, $\chi^2(9) = 48.58$, $p < .001$, $\varepsilon = 0.91$, so the Huynh-Feldt correction was applied. The main effect of Trial was significant, $F(3.70, 1035.81) = 19.35$, $p < .001$, partial $\eta^2 = .07$. Performance improved from Trials 1–20 to Trials 61–80 (*p*s < .050), with no difference between Trials 61–80 and Trials 81–100, $p = .539$. The main effect of IGT Version was not significant, $F(1,280) = 2.97$, $p = .086$, partial $\eta^2 = .01$. There was a significant interaction between Trials and IGT Version, $F(3.70, 1035.81) = 5.81$, $p < .001$, partial $\eta^2 = .02$. On Trials 1–20, participants selected more advantageously on the Original than PAR IGT, $p = .017$, Cohen's $d = 0.29$. However, on Trials 21–40 ($p = .004$, $d = 0.34$) and Trials 41–60 ($p = .002$, $d = 0.37$), participants selected more advantageously on the PAR than Original IGT. No task differences emerged on Trials 61–80 ($p = .163$, $d = 0.17$) or Trials 81–100 ($p = .858$, $d = 0.02$).

Individual Deck Selections Comparisons

Mauchly's Test of Sphericity was significant, $\chi^2(27) = 2000.25$, $p < .001$, $\varepsilon = 0.55$, so the Greenhouse–Geisser correction was applied. The main effect of Deck was significant, $F(3.88, 1086.77) = 68.77$, $p < .001$, partial $\eta^2 = .20$ (see Table 1 for significant findings). The main effect of IGT version was not significant, $F(1,280) = 0.98$, $p = .324$, partial $\eta^2 = .00$. There was a significant interaction between Deck and IGT Version, $F(3.88, 1086.77) = 2.79$, $p = .027$, partial $\eta^2 = .01$. On Trials 1–40, participants completing the PAR IGT selected more from Deck A ($p = .014$, $d = 0.29$) and Deck C ($p = .034$, $d = 0.25$), but less from Deck B ($p = .036$, $d = 0.25$) than on the original IGT. No differences emerged on early

Table 1. Demographic information, means, standard deviations, and ANOVA results for study variables

Variable	PAR (2007) IGT		Original IGT	
	M	*SD*	*M*	*SD*
Age	18.52	1.35	18.99	2.28
Gender (% Female)	44.4%		50.0%	
Education (in years)	12.11	0.39	12.15	0.50
Ethnicity (% Caucasian)	66.7%		66.2%	
(C + D)–(A + B)				
Total (1–100)	3.92	24.94	−0.58	26.05
1–20	−2.94	5.06	−1.30	6.36
21–40	0.56	5.96	−1.86	8.11
41–60	2.02	7.64	−0.86	7.73
61–80	2.59	8.32	1.21	8.14
81–100	1.69	9.39	1.50	8.58
% Decks				
A 1–40[a]	20.42	6.38	18.03	9.68
B 1–40[b]	32.55	9.48	35.93	16.90
C 1–40[c]	22.97	7.11	20.61	11.29
D 1–40[d]	24.07	9.56	25.44	11.10
A 41–100[e]	15.40	8.55	14.65	9.88
B 41–100[f]	29.49	15.18	33.81	16.10
C 41–100[g]	23.73	13.83	21.30	16.43
D 41–100	31.17	17.92	30.24	15.03

Note: PAR IGT = Psychological Assessment Resources Iowa Gambling Task, 2007 edition; Original IGT = Bechara et al. (1994) Iowa Gambling Task; (C + D) – (A + B) = Advantageous minus disadvantageous deck selections by 20-card blocks of trials; % Decks = Percent selections from each individual deck on early (Trials 1-40) versus later (Trials 41–100) trials.
[a]A 1–40 < B 1–40, C 1–40, D 1–40, A 41–100, B 41–100, C 41–100, D 41–100.
[b]B 1–40 > C 1–40, D 1–40, A 41–100, B 41–100, C 41–100, D 41–100.
[c]C 1–40 > 41–100; C 1–40 < D 1–40, B 41–100, D 41–100.
[d]D 1–40 > A 41–100, D 41–100; D 1–40 < B 41–100.
[e]A 41–100 < B 41–100, C 41–100, D 41–100.
[f]B 41–100 > C 41–100.
[g]C 41–100 < D 41–100.

Deck D selections, $p = .267$, $d = 0.13$. During the later trials (Trials 41–100), participants who completed the PAR IGT selected less from Deck B than participants who completed the original IGT, $p = .021$, $d = 0.28$. No other differences were significant, $ps > .177$, $ds < 0.17$.

Discussion

The present study examined differences in performance on two versions of the IGT. We found more advantageous performance on Trials 1–20 of the original IGT, but more advantageous performance on Trials 21–60 of the PAR IGT. That significant differences emerged during earlier trials points to a difference in when decision-making switches from under ambiguity to under risk (Brand et al., 2007) on these two IGT versions. Increasing the level of risk associated with disadvantageous deck selections may lead individuals to learn to decide advantageously more quickly.

Examining individual deck selections, we found most significant differences occurred during the earlier trials (Trials 1–40). Participants who completed the PAR IGT selected more from Deck A and C, while also selecting less from Deck B, compared to the original IGT. These individual deck findings on the early trials may explain our previous finding: the early Deck A selections may have allowed participants to more quickly learn to avoid this deck, leading to an earlier transition to the advantageous decks than on the original IGT. Further evidence of this transition comes from the PAR IGT participants selecting less from Deck B during the later trials (Trials 41–100), showing more of a focus on long-term outcomes than frequency of losses (Chiu, Lin, Huang, Lin, Lee, & Hsieh, 2008). This finding points to a potential difference in how participants learn from feedback on the original versus PAR IGT, as well as provides further support for examining individual deck selections regardless of the IGT version. However, it should be noted that our effect sizes fell generally in the small range, indicating the importance of future replication in a more representative sample.

There are several implications. IGT versions are used almost interchangeably in the literature, with few details provided in study procedures to determine the administered version. These version differences could help explain some of the mixed

findings across studies (e.g., Adida et al., 2015; Bark et al., 2005; Euteneuer et al., 2009; Guillaume et al., 2010; Tchanturia et al., 2012), as well as between earlier and more recent studies of healthy control participants (Steingroever et al., 2013). We found significant version differences in task performance in a sample of healthy, non-clinically presenting college students. If these between-version differences are replicated in additional studies, a determination would need to be made as to the clinical and research utility of each version. If future replications confirm an earlier transition to advantageous decision-making on the PAR versus original IGT, utilization of the PAR IGT might lead to greater confidence in a conclusion of decision-making impairment in patient and non-patient samples. However, prior to making this claim, future research should examine version differences in patients of varying ages and backgrounds undergoing neuropsychological evaluation of decision-making processes.

There are several limitations to the present study. We utilized a sample of healthy college student volunteers that limits generalizability to just this population. As the IGT is the most common, and only clinically available, behavioral measure of risky decision-making, it is important to follow-up on the present study with a larger, more representative sample. Such a sample would allow for examination of replicability of our generally small effect sizes. We utilized a between-groups rather than within-groups design. Future research in which both IGT versions were completed in a counterbalanced order would allow for direct comparison of the two versions in the first administration, and then of cross-version practice effects in the second administration.

The present study provided the first evidence that the original and PAR (2007) versions of the IGT may not assess risky decision-making processes in the same way, which could lead to differences in how decision-making impairments are assessed in clinical and non-clinical samples. Evidence points towards the PAR IGT enhancing early learning to avoid disadvantageous decks compared to the original IGT. Utilization of the PAR IGT in research and clinical trials may provide greater evidence of decision-making impairment among those who fail to learn to decide advantageously, but additional research is needed to confirm these version differences.

Conflict of interest

None declared.

References

Adida, M., Jollant, F., Clark, L., Guillaume, S., Goodwin, G. M., Azorin, J.-M., et al. (2015). Lithium might be associated with better decision-making performance in euthymic bipolar patients. *European Neuropsychopharmacology*, *25*, 788–797. doi:10.1016/j.euroneuro.2015.03.003.

Balodis, I. M., MacDonald, T. K., & Olmstead, M. C. (2006). Instructional cues modify performance on the Iowa gambling task. *Brain and Cognition*, *60*, 109–117. doi:10.1016/j.bandc.2005.05.007.

Bark, R., Dieckmann, S., Bogerts, B., & Northoff, G. (2005). Deficit in decision-making in catatonic schizophrenia: An exploratory study. *Psychiatry Research*, *134*, 131–141. doi:10.1016/j.psychres.2004.04.013.

Barnhart, W. R., & Buelow, M. T. (2017). Risky decision-making in college students as a function of self-reported eating behaviors. *Journal of Undergraduate Research at Ohio State*, *7*, 1–7.

Bechara, A. (2007). *Iowa gambling task professional manual*. Lutz, FL: Psychological Assessment Resources.

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, *50*, 7–15. doi:10.1016/0010-0277(94)90018-3.

Brand, M., Recknor, E. C., Grabenhorst, F., & Bechara, A. (2007). Decisions under ambiguity and decisions under risk: Correlations with executive functions and comparisons of two different gambling tasks with implicit and explicit rules. *Journal of Clinical and Experimental Neuropsychology*, *29*, 86–99. doi:10.1080/13803390500507196.

Buelow, M. T., & Blaine, A. L. (2015). The assessment of risky decision-making: A factor analysis of performance on the Iowa gambling task, balloon analogue risk task, and Columbia card task. *Psychological Assessment*, *27*, 777–785. doi:10.1037/a0038622.

Buelow, M. T., Frakey, L. L., Grace, J., & Friedman, J. H. (2014). The contribution of apathy and increased learning trials to risky decision-making in Parkinson's disease. *Archives of Clinical Neuropsychology*, *29*, 100–109. doi:10.1093/arclin/act065.

Buelow, M. T., & Suhr, J. A. (2009). Construct validity of the Iowa gambling task. *Neuropsychology Review*, *19*, 102–114. doi:10.1007/s11065-009-9083-4.

Chiu, Y. C., Lin, C. H., Huang, J. T., Lin, S., Lee, P. L., & Hsieh, J. C. (2008). Immediate gain is long-term loss: Are there foresighted decision makers in the Iowa gambling task? *Behavioral and Brain Functions*, *4*, 13. doi:10.1186/1744-9081-4-13.

Clark, L., Iversen, S. D., & Goodwin, G. M. (2002). Sustained attention deficit in bipolar disorder. *British Journal of Psychiatry*, *180*, 313–319. doi:10.1192/bjp.180.4.313.

Dymond, S., Cella, M., Cooper, A., & Turnbull, O. H. (2010). The contingency-shifting variant Iowa gambling task: An investigation with young adults. *Journal of Clinical and Experimental Neuropsychology*, *32*, 239–248. doi:10.1080/13803390902971115.

Euteneuer, F., Schaefer, F., Stuermer, R., Boucsein, W., Timmermann, L., Barbe, M. T., et al. (2009). Dissociation of decision-making under ambiguity and decision-making under risk in patients with Parkinson's disease: A neuropsychological and psychophysiological study. *Neuropsychologia*, *47*, 2882–2890. doi:10.1016/j.neuropsychologia.2009.06.014.

Gu, H., Liu, C., Liu, C., Chen, M., Zhang, Q., Zhai, J., et al. (2013). The combined effects of the 5-HTTLPR and HTR1A rs6295 polymorphisms modulate decision-making in schizophrenia patients. *Genes, Brain, and Behavior*, *12*, 133–139. doi:10.1111/j.1601-183X.2012.00866.x.

Guillaume, S., Sang, C. N., Jaussent, I., Raingeard, I., Bringer, J., Jollant, F., et al. (2010). Is decision-making really impaired in eating disorders? *Neuropsychology*, *24*, 808–812. doi:10.1037/a0019806.

Steingroever, H., Wetzels, R., Horstmann, A., Neumann, J., & Wagenmakers, E. J. (2013). Performance of healthy participants on the Iowa gambling task. *Psychological Assessment*, *25*, 80–93. doi:10.1037/a0029929.

Tchanturia, K., Liao, P.-C., Forcano, L., Fernandez-Aranda, F., Uher, R., Treasure, J., et al. (2012). Poor decision-making in male patients with anorexia nervosa. *European Eating Disorders Review*, *20*, 169–173. doi:10.1002/erv.1154.