Psychological Assessment
2013, Vol. 25, No. 1, 180–193

1040-3590/13/$12.00 DOI: 10.1037/a0029929

Performance of Healthy Participants on the Iowa Gambling Task

Helen Steingroever and Ruud Wetzels
University of Amsterdam

Annette Horstmann and Jane Neumann
Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany, and Leipzig University Medical Center

Eric-Jan Wagenmakers
University of Amsterdam

The Iowa Gambling Task (IGT; Bechara, Damasio, Damasio, & Anderson, 1994) is often used to assess decision-making deficits in clinical populations. The interpretation of the results hinges on 3 key assumptions: (a) healthy participants learn to prefer the good options over the bad options; (b) healthy participants show homogeneous choice behavior; and (c) healthy participants first explore the different options and then exploit the most profitable ones. Here we test these assumptions using 2 extensive literature reviews and analysis of 8 data sets. The results show that all 3 assumptions may be invalid; that is, (a) healthy participants often prefer decks with infrequent losses; (b) healthy participants show idiosyncratic choice behavior; and (c) healthy participants do not show a systematic decrease in the number of switches across trials. Our findings question the prevailing interpretation of IGT data and suggest that, in future applications of the IGT, key assumptions about performance of healthy participants warrant close scrutiny.

Keywords: choice behavior, frequency of losses, between-group variability, within-group variability, exploration–exploitation

Patients with lesions to the ventromedial prefrontal cortex (vmPFC) suffer from severe decision-making deficits; even though their cognitive, memory, and problem-solving abilities remain intact, they are often unable to consider the consequences of their actions and to learn from their mistakes (Bechara, Damasio, Damasio, & Anderson, 1994; Bechara, Damasio, Damasio, & Lee, 1999; Bechara, Tranel, & Damasio, 2000; Yechiam, Busemeyer, Stout, & Bechara, 2005). The most famous vmPFC patient, Phineas Gage, survived an accident in which an iron rod lesioned his frontal cortex, causing a profound change in his decision-making abilities and social behavior. Prior to the accident, Gage was a hard-working, efficient, and responsible railroad construction foreman. However, after his accident he could not resume work because his behavior was impulsive, impatient, and antisocial.[1] These symptoms are characteristic for vmPFC patients in general (Boes et al., 2011); vmPFC patients often make decisions that are irresponsible, risky, and go against their own interests. When confronted with a decision problem where the gains of immediate reward need to be weighted against the risks of long-term loss, vmPFC patients tend to focus on the immediate reward and disregard possible negative future outcomes of their choices, such as separation from family and friends and loss of reputation and job (Bechara & Damasio, 2002; Dunn, Dalgleish, & Lawrence, 2006).

To study the decision-making deficits of vmPFC patients, Bechara et al. (1994) developed the Iowa Gambling Task (IGT). The purpose of this task is to mimic real-life decision making in an experimental context. In the IGT, participants are presented with four decks of cards; each card yields a reward and, occasionally, also a loss. Participants are told to repeatedly choose cards to optimize their long-term outcomes. Unbeknownst to the participants, two decks contain risky cards that have high immediate rewards but negative long-term outcomes, whereas the other two decks contain safe cards that have small immediate rewards but positive long-term outcomes. Participants can succeed on the IGT only when they learn to forgo high immediate rewards and prefer the safe options over the risky options. This choice behavior is assumed to be characteristic for healthy participants and is taken as evidence that they base their choices on the long-term outcomes of the decks (Bechara et al., 1994). In contrast, vmPFC patients

This article was published Online First September 17, 2012.

Helen Steingroever and Ruud Wetzels, Department of Psychology, University of Amsterdam, Amsterdam, the Netherlands; Annette Horstmann and Jane Neumann, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany, and Integrated Research and Treatment Center Adiposity Diseases, Leipzig University Medical Center, Leipzig, Germany; Eric-Jan Wagenmakers, Department of Psychology, University of Amsterdam.

Annette Horstmann's and Jane Neumann's work is supported by the Federal Ministry of Education and Research, Germany (FKZ: 01EO1001). We thank Fernie and Tunney (2006); Fridberg et al. (2010); Kjome et al. (2010); Premkumar et al. (2008); Rodríguez-Sánchez et al. (2005); Toplak et al. (2005); and Wood et al. (2005) for providing the data used in this article.

Correspondence concerning this article should be addressed to Helen Steingroever, Department of Psychology, University of Amsterdam, Weesperplein 4, 1018 XA Amsterdam, the Netherlands. E-mail: helen.steingroever@gmail.com

[1] For more information on Phineas Gage, see http://www.deakin.edu.au/hmnbs/psychology/gagepage/.

Table 1
Payoff Scheme of the Traditional Iowa Gambling Task as Developed by Bechara et al. (1994)

Variable	Deck A: Bad deck with frequent losses	Deck B: Bad deck with infrequent losses	Deck C: Good deck with frequent losses	Deck D: Good deck with infrequent losses
Reward/trial	100	100	50	50
Number of losses/10 cards	5	1	5	1
Loss/10 cards	−1250	−1250	−250	−250
Net outcome/10 cards	−250	−250	250	250

presumably prefer the risky options because they are insensitive to the future consequences of their choices. The result of this "myopia for the future" is that the choice behavior of vmPFC patients is controlled primarily by immediate prospects, positive or negative (Bechara et al., 1994, 2000).

Over the last 2 decades, the IGT has become what is arguably the most popular neuropsychological paradigm to measure decision-making deficits in clinical populations (Toplak, Sorge, Benoit, West, & Stanovich, 2010). To illustrate, a search for *Iowa Gambling Task* yields about 178,000 hits on Google, and about 3,640 hits on Google Scholar.[2] Prior to 2005, the IGT had already been used in almost 100 neurological and psychiatric studies (Bowman, Evans, & Turnbull, 2005). Studies that use the IGT have involved a broad variety of clinical populations such as patients with vmPFC lesions (Bechara et al., 1999; Bechara, Damasio, Tranel, & Anderson, 1998; Bechara et al., 2000), pathological gambling (Cavedini, Riboldi, Keller, D'Annucci, & Bellodi, 2002), obsessive-compulsive disorder (Cavedini, Riboldi, D'Annucci, et al., 2002), psychopathic tendencies (Blair, Colledge, & Mitchell, 2001), and schizophrenia (Bark, Dieckmann, Bogerts, & Northoff, 2005; Martino, Bucay, Butman, & Allegri, 2007). In addition, the IGT has been applied to cocaine addicts (Stout, Busemeyer, Lin, Grant, & Bonson, 2004), traffic offenders (Lev, Hershkovitz, & Yechiam, 2008), and inmates (Yechiam, Kanz, et al., 2008). The traditional way to compare performance of a clinical group to that of a control group is based either on the overall proportion of choices from the good decks or on a difference score between the overall proportion of choices from the good and bad decks. Note that these procedures collapse choice proportions over the two good decks and over the two bad decks, leading to a loss of potentially diagnostic information (Chiu & Lin, 2007; Dunn et al., 2006; Lin, Chiu, Lee, & Hsieh, 2007), a point to which we will return later.

In addition to the broad application of the IGT in neuropsychological studies, Bechara (2007) promoted to use the IGT in clinical assessment of patients with decision-making deficients. Bechara therefore developed IGT software including a professional manual. The software yields different normative scores, such as the total net gain and the total number of cards chosen from each deck. These normative scores can be compared to those of a demographically corrected sample or U.S. census-matched sample. According to Bechara, the IGT is "ideal for assessing patients who exhibit poor decision-making skills in the presence of otherwise normal or unaffected intelligence because of head injury or insult or any other condition thought to impact the function of the prefrontal cortex" (PAR, 2012).

The validity of IGT scores as a diagnostic instrument hinges on at least three assumptions about the performance of healthy participants on the IGT. First, there is the explicit assumption that healthy participants learn to prefer the good options over the bad options (Bechara et al., 1994). Second, there is the implicit assumption that choice performance for healthy participants is homogeneous—a clinical test in which healthy participants have widely diverging scores is of limited use. Third, there is the explicit assumption that healthy participants first explore the different options and then settle down and exploit the most profitable ones (i.e., the exploration–exploitation trade-off; Bechara et al., 1994).

Our goal in this article is to assess the validity of the above assumptions using two extensive reviews of the literature and detailed analysis of eight experimental data sets. To anticipate the conclusion, our results call into question each of the above assumptions. First, the choices of healthy participants are primarily driven by the frequency of losses rather than the long-term outcomes of the decks (i.e., the frequency-of-losses effect). Second, the performance of healthy participants is characterized by a considerable variability, both across groups and across participants within groups. Third, healthy participants fail to show a pronounced exploration–exploitation trade-off.

The outline of this article is as follows. The first section explains the IGT, outlines the assumptions about the IGT performance of healthy participants, and reviews current criticism on the IGT. The second section tests each of the three key assumptions above by reviewing findings from 39 groups of healthy participants and by analyzing data from eight IGT studies. Section 3 summarizes our findings and discusses their ramifications.

The Iowa Gambling Task

Description of the Iowa Gambling Task

The purpose of the IGT is to measure decision-making deficits of clinical populations in an experimental setting. The IGT mimics real-life decision making in the sense that participants are required to integrate rewards and losses, weight benefits and risks associated with each possible option, remember that information, and use it in subsequent decisions. In the traditional IGT, participants are initially given $2,000 facsimile money and are presented with four decks of cards. Participants are told to successively choose cards from the four decks to maximize their long-term outcome (Bechara et al., 1994; Bechara, Damasio, Tranel, & Damasio, 1997).

In the IGT, each card is associated with a specific amount of monetary reward and potentially also of monetary loss. The dif-

[2] Retrieved February 2, 2012.

ference between the four decks lies in the payoff scheme: Two decks (Decks A and B) are associated with high immediate, constant rewards, but with even higher unpredictable, occasional losses resulting in negative long-term outcomes. Decks A and B are therefore called bad or disadvantageous decks. The other two decks (Decks C and D) are associated with lower immediate, constant rewards, but with even lower unpredictable, occasional losses and thus result in positive long-term outcomes. Decks C and D are therefore called good or advantageous decks. In addition to the different magnitude of the immediate rewards and occasional losses resulting in different long-term outcomes, the decks differ in the frequency of losses: Two decks yield frequent losses (Decks A and C), and two decks yield infrequent losses (Decks B and D). The payoff scheme of the traditional IGT as developed by Bechara et al. (1994) is presented in Table 1. Each time participants choose a card, they receive feedback on the reward and the loss (if any) associated with that card, and the running tally. Figure 1 shows an example screen shot of the IGT. Participants are not told how long the task takes, but typically it contains 100 trials. The task aims to determine whether participants learn through trial and error to prefer the good, safe decks (i.e., Decks C and D) over the bad, risky decks (i.e., Decks A and B).

IGT performance is typically recorded for a clinical group and a control group. Different outcome measures have been reported, but most studies compute either the overall proportion of choices from the good decks or a difference score between the overall proportion of choices from the good and bad decks. An overall proportion of choices from the good decks larger than .50 or a positive difference score are considered as nonimpaired performance (Bechara et al., 1999, 1998; Bowman & Turnbull, 2003). The disadvantage of these outcome measures is that they do not reflect the change in the deck preferences across trials. Therefore, many studies also report these outcome measures in blocks of 10–20 trials. These trial-dependent outcome measures are often entered in a Block × Group analysis of variance. Note that these procedures collapse choice proportions over the two good decks and over the two bad decks (Chiu & Lin, 2007; Dunn et al., 2006; Lin et al., 2007).

Assumptions Underlying Performance on the Iowa Gambling Task

Bechara et al. (1994) developed the IGT to measure decision-making deficits of patients with vmPFC lesions. According to

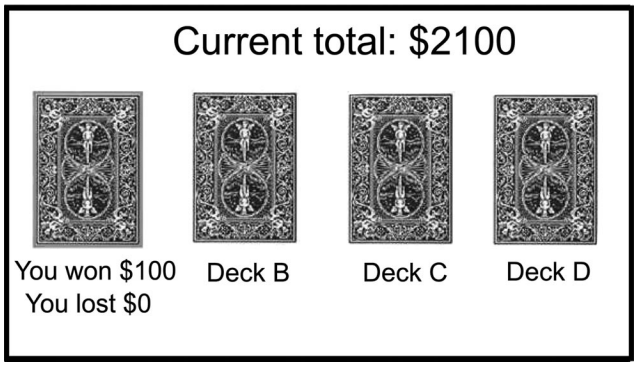

Figure 1. Example screen shot of the Iowa Gambling Task.

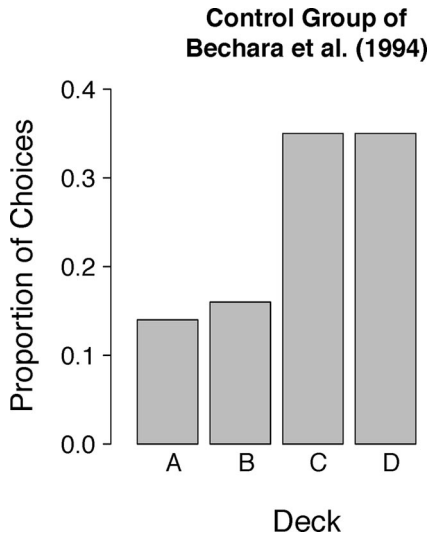

Figure 2. Mean proportion choices from each deck in Bechara et al.'s (1994) control group. This control group of healthy participants shows a pronounced overall preference for the good decks (i.e., Decks C and D).

Bechara et al. (1994, 2000), these brain-damaged patients perform poorly on the IGT because they are insensitive to the future consequences of their actions ("myopia for the future"). Performance of vmPFC patients is therefore assumed to be guided by immediate prospects, positive or negative, rather than the accumulation of long-term positive outcomes (Bechara et al., 1994, 2000). Such myopic choice behavior results in a preference for the bad, risky decks, that is, those decks that yield constant high immediate rewards but even higher unpredictable losses (i.e., Decks A and B); hence, this choice behavior leads to long-term negative outcomes. Healthy participants, on the other hand, are assumed to prefer the good, safe decks (i.e., Decks C and D) because they correctly infer that these decks yield positive long-term outcomes,

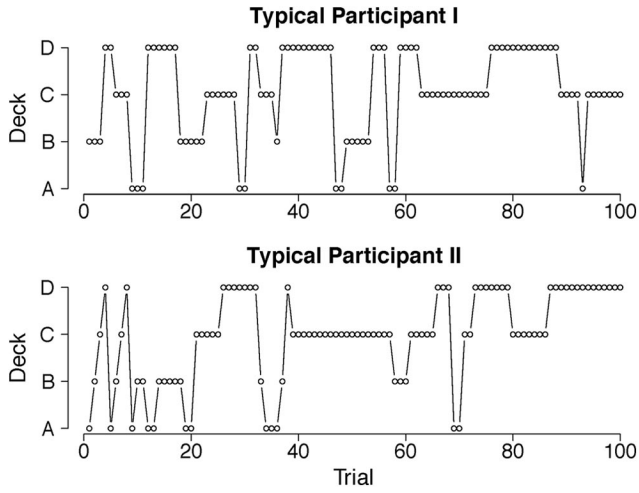

Figure 3. Deck selection profiles of two typical participants in Bechara et al.'s (1994) control group. These two participants first explore all decks, then gradually switch to the good decks, and only occasionally return to the bad decks.

despite rather small immediate rewards (Bechara et al., 1994). Figure 2 displays the mean proportion of choices from each deck in Bechara et al.'s (1994) control group of healthy participants ($N = 44$). This figure illustrates that the control group showed a strong overall preference for the good decks, and thus clearly supports the assumption of Bechara et al. that healthy participants prefer the good decks due to their positive long-term outcomes. It is important to note that this group of healthy participants chose both good decks and both bad decks about equally often, indicating that these participants do not prefer the decks with infrequent losses (Decks B and D) over the decks with frequent losses (Decks A and C).

To support their assumptions about the characteristic choice behavior of healthy participants across trials, Bechara et al. (1994) presented deck selection profiles of two typical control participants (cf. Figure 3). These profiles illustrate that these control participants "initially sampled all decks . . . , but eventually switched to more and more selections from the good Decks C and D, with only occasional returns to decks A and B" (p. 12). In the case of Bechara et al.'s two typical control participants, the last 50 trials featured only five and six choices from the bad decks (cf. Figure 3).

The difference in the performance of healthy participants and vmPFC patients has been frequently interpreted as support of the somatic marker hypothesis. This hypothesis states that emotion-related signals generated from the body, so-called somatic markers, regulate decision making. Especially in complex, uncertain situations, somatic markers are assumed to mark the possible options with an emotional signal related to the goodness or badness of the associated outcome. Based on these emotional signals, healthy participants anticipate certain bodily states associated with each response option (Bechara & Damasio, 2002; Bechara et al., 1999; Damasio, Everitt, & Bishop, 1996). These expectations are assumed to guide decisions by restricting the decision-making space to options associated with positive affective states. Crucially, vmPFC patients are found to be unable to rely on somatic markers when facing multiple response options (Bechara & Dama-

Table 2
Overview of the Groups of Healthy Participants Included in the Literature Reviews

Study	Sample size	Age[a]	Literature review[b]	Data received?
Bark et al. (2005)	26	$M = 29.8$, $SD = 9.4$	1, 2	Not available
Bechara et al. (1994)	44	20–79	1, 2	No answer
Bechara et al. (1998)	21	24–68	2	Not available
Bechara et al. (1999)	13	22–58	2	Not available
Blair et al. (2001)	23	10.5–13.9	2	Not available
Caroselli et al. (2006)	141	$M = 21.7$, $SD = 4.6$	1, 2	No answer
Cavedini, Riboldi, D'Annucci, et al. (2002)	34	$M = 29.5$, $SD = 8.9$	2	Not available
Cavedini, Riboldi, Keller, et al. (2002)	40	$M = 30.3$, $SD = 9.6$	2	No answer
Fernie & Tunney (2006)	20	Undergraduates	1, 2	Received
Fridberg et al. (2010)	15	$M = 29.6$, $SD = 7.6$	1, 2	Received
Johnson et al. (2006)	14	$M = 15.9$, $SD = 2.4$	2	No answer
Kester et al. (2006)	25	$M = 17.1$, $SD = 1.8$	1, 2	Not available
Kjome et al. (2010)	20	$M = 33.9$, $SD = 11.2$	2	Received
Lehto & Elorinne (2003)	51	7–11	2	Not available
Lehto & Elorinne (2003)	40	19–53	2	Not available
Lev et al. (2008)	36	$M = 36.9$, $SD = 12.2$	2	Not available
Martino et al. (2007)	15	$M = 35.0$, $SD = 10.9$	1, 2	No answer
North & O'Carroll (2001)	20	$M = 30.8$, $SD = 1.9$	1, 2	Not available
O'Carroll & Papps (2003)	11	$M = 20.0$, $SD = 3.1$	1, 2	Not available
Overman et al. (2004)	35	6th grade (~11 years)	2	Not available
Overman et al. (2004)	60	7th grade (~12 years)	2	Not available
Overman et al. (2004)	57	8th grade (~13 years)	2	Not available
Overman et al. (2004)	59	9th grade (~14 years)	2	Not available
Overman et al. (2004)	60	10th grade (~15 years)	2	Not available
Overman et al. (2004)	60	11th grade (~16 years)	2	Not available
Overman et al. (2004)	60	12th grade (~17 years)	2	Not available
Overman et al. (2004)	60	17–23	2	Not available
Petry (2001)	21	$M = 36.1$, $SD = 11.5$	1, 2	Not available
Premkumar et al. (2008)	25	$M = 35.4$, $SD = 11.9$	2	Received
Ritter et al. (2004)	15	$M = 47.1$, $SD = 10.2$	1, 2	Not available
Rodríguez-Sánchez et al. (2005)	22	$M = 26.1$, $SD = 6.5$	1, 2	Received
Sevy et al. (2007)	20	$M = 33.0$, $SD = 10.0$	1, 2	Not available
Shurman et al. (2005)	10	$M = 32.1$, $SD = 4.5$	1, 2	Not available
Tomb et al. (2002)	10	Undergraduates	1, 2	Not available
Toplak et al. (2005)	34	$M = 15.4$, $SD = 1.5$	1, 2	Received
Wilder et al. (1998)	30	18–52	1, 2	No answer
Wood et al. (2005)	88	$M = 22.1$, $SD = 4.5$	2	Received
Wood et al. (2005)	67	$M = 77.3$, $SD = 4.6$	2	Received
Yechiam, Hayden, et al. (2008)	25	$M = 39.2$, $SD = 13.3$	2	Not available

[a] Age (in years) as mean and standard deviation, range, profession, or school grade with the approximate age in parentheses. [b] 1 = inclusion in Literature Review 1 (choices from each deck separately); 2 = inclusion in Literature Review 2 (choices from the good and bad decks).

sio, 2002; Bechara et al., 1999). This inability is assumed to lead to poor decision making (but see Dunn et al., 2006; Maia & McClelland, 2004, for critiques).

In sum, the basic assumption underlying the IGT is that patients with vmPFC lesions prefer the bad decks because of their "myopia for the future," whereas healthy participants base their choices on the long-term outcomes and therefore gradually learn to prefer both good decks.

Current Criticism on the Iowa Gambling Task

Despite its frequent use, the IGT has been confronted by a growing body of criticism. It is increasingly apparent that even healthy participants perform poorly because they fail to develop a preference for both good options, contradicting the assumptions of Bechara et al. (1994). More specifically, many studies have shown that healthy participants prefer the decks with infrequent losses (Decks B and D), indicating that healthy participants base their choices on the frequency of losses rather than on the long-term outcomes (Caroselli, Hiscock, Scheibel, & Ingram, 2006; Dunn et al., 2006; Lin et al., 2007; MacPherson, Phillips, & Della Sala, 2002; Wilder, Weinberger, & Goldberg, 1998; Yechiam & Busemeyer, 2005). Caroselli et al. (2006) even concluded that "the card selection preferences of undergraduates were more similar to those of Bechara et al.'s seven patients with frontal lobe damage than to those of their 44 normal controls" (p. 208).

Studies applying the Soochow Gambling Task—a variation of the IGT in which both good options yield high-frequent losses and both bad options yield low-frequent losses—show that healthy participants prefer the bad options over the good options, indicating again that healthy participants base their choices on the frequency of losses (Ahn, Busemeyer, Wagenmakers, & Stout, 2008; Chiu et al., 2008). The frequency-of-losses effect has also been found in a different version of the IGT in which Chiu and Lin (2007) established a higher contrast between rewards and losses on each trial by increasing their magnitude while keeping the traditional long-term outcomes (i.e., −250 for 10 cards from the bad decks, and +250 for 10 cards from the good decks). Chiu and Lin concluded that "the IGT contains some redundant procedures, confounding features, and problems in interpretation" and that "these problems should be refined to make the IGT a truly useful assessment tool" (Conclusion section, para. 1). Further evidence for the frequency-of-losses effect in healthy participants has been reported by Huizenga, Crone, and Jansen (2007), who showed that the dominant IGT strategy in the considered age groups ranging from 6 to 25 years is to focus on the frequency of losses, a dominance that increases with age.

It is important to note that for many published IGT studies, the selection proportions for the two good decks and the two bad decks have been collapsed. This presentation method obscures the impact of the frequency-of-losses effect (Chiu & Lin, 2007; Dunn et al., 2006; Lin et al., 2007). In this article we aim to overcome this complication by reviewing the literature that does contain information from each individual deck, by reanalyzing raw data of previous IGT studies that we received upon request, and by collecting data from a relatively large sample of healthy participants.

Literature Reviews and Data Analyses

In this section we assess the three key assumptions about the performance of healthy participants on the IGT. Our analyses below provide evidence for three findings that run counter to the key IGT assumptions. First, the primary factor that drives deck selections is frequency of losses instead of an appreciation of long-term outcomes. Second, IGT performance is characterized by a considerable variability, both across groups and across participants within groups. Third, healthy participants fail to progress from an initial stage of exploration to a later a stage of exploitation.

Table 3
Literature Review 1: Mean Proportion of Choices From Each Deck Separately of Healthy Participants

Study	Sample size	Deck A: Bad deck with frequent losses	Deck B: Bad deck with infrequent losses	Deck C: Good deck with frequent losses	Deck D: Good deck with infrequent losses
Bark et al. (2005)	26	.23	.29	.24	.24
Bechara et al. (1994)	44	.14	.16	.35	.35
Caroselli et al. (2006)	141	.22	.35	.20	.23
Fernie & Tunney (2006)	20	.20	.33	.22	.25
Fridberg et al. (2010)	15	.13	.30	.14	.43
Kester et al. (2006)	25	.21	.25	.26	.28
Martino et al. (2007)	15	.15	.27	.21	.37
North & O'Carroll (2001)	20	.10	.20	.36	.34
O'Carroll & Papps (2003)	11	.15	.28	.20	.37
Petry (2001)	21	.16	.25	.26	.33
Ritter et al. (2004)	15	.18	.25	.24	.33
Rodriguez-Sánchez et al. (2005)	22	.17	.32	.20	.31
Sevy et al. (2007)	20	.19	.31	.23	.27
Shurman et al. (2005)	10	.16	.18	.34	.32
Tomb et al. (2002)	10	.15	.19	.34	.32
Toplak et al. (2005)	34	.23	.30	.20	.27
Wilder et al. (1998)	30	.20	.27	.24	.29
Total sample size[a] (*M*)	479	.19	.29	.24	.29

Note. The weighted mean proportion of choices of the 17 studies support the frequency-of-losses effect. All studies are based on a 100-trial Iowa Gambling Task except Fridberg et al. (2010; 95 trials).
[a] Weighted by the sample size of each study.

Method

To test the assumptions of Bechara et al. (1994) about the choice behavior of healthy participants on the IGT, we undertook three steps: We conducted two literature reviews, analyzed seven published data sets, and analyzed a large data set collected by ourselves ($N = 162$).

Literature Review 1: Choices from each deck separately. Our first literature review focuses on the mean proportion of choices from each of the four decks. The aim of this literature review was to quantify the size of the frequency-of-losses effect in healthy participants, defined here as participants who do not have any neurological impairments and who are typically used as control participants. We tried to locate any study that reported the mean proportion of choices from each deck separately. These studies were located by searching for IGT studies published before July 1, 2011, in ISI Web of Knowledge and Google Scholar (search term: *Iowa Gambling Task*), and by searching the references of relevant articles.

We restricted this literature review to studies that presented the mean proportion of choices from each of the four decks in the text or tables, or provided figures that enabled us to extract this information. For maximum compatibility, we only included IGT studies that used a procedure and payoff scheme close to the ones described by Bechara et al. (1994). This means that we only included studies that used facsimile money, that yielded a net outcome of −250 for the bad decks and +250 for the good decks in every 10 card unit, that used two high-frequent losses decks (i.e., Decks A and C) and two low-frequent losses decks (i.e., Decks B and D), and that used the same reward structure within the good decks and within the bad decks. Our first literature review involves 17 studies covering 479 healthy participants.

Literature Review 2: Choices from the good and bad decks. Our second literature review focuses on the mean proportion of choices from the good and bad decks. The aim of this literature review was to quantify the variability between groups of healthy

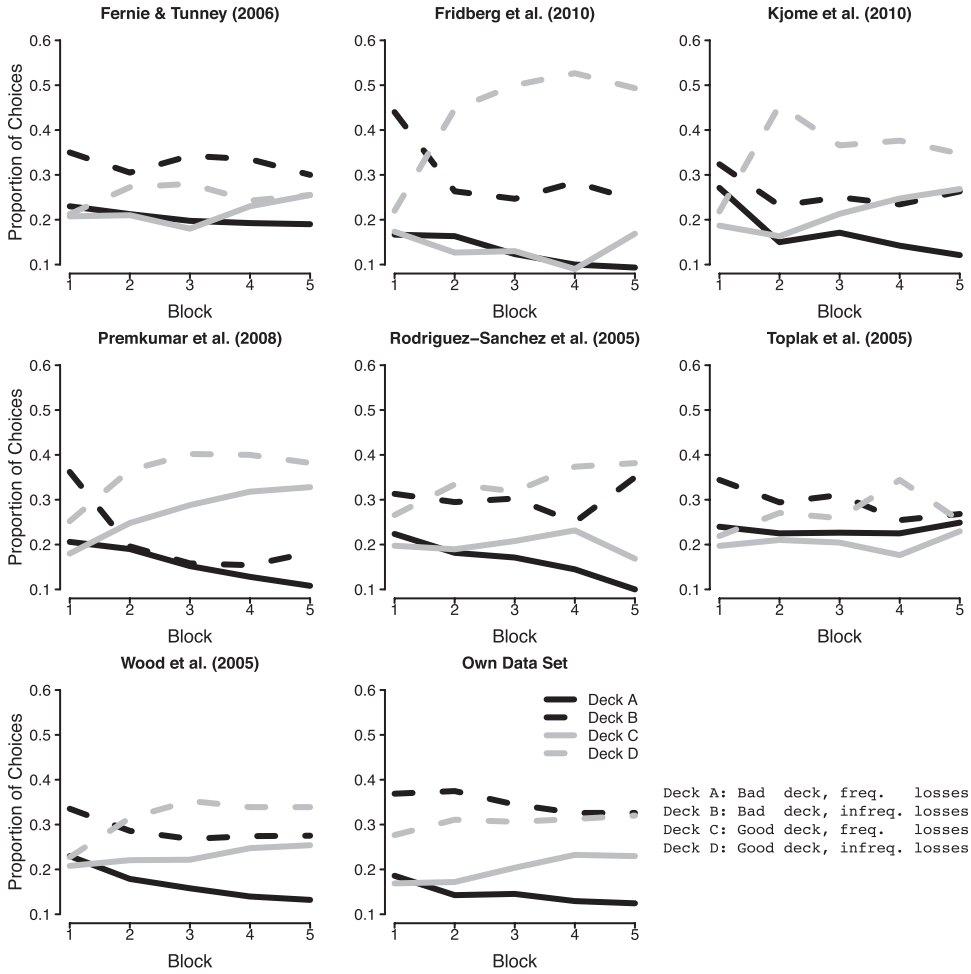

Figure 4. Mean proportion of choices from each deck within five blocks in each data set that was sent to us upon request, and in our own data set. Each block contains 20 trials, except the last block of Fridberg et al.'s (2010) data set (15 trials). Note that Kjome et al. (2010), Premkumar et al. (2008), and Wood et al. (2005) did not use the exact payoff scheme as reported by Bechara et al. (1994); these studies nevertheless met our inclusion criteria because they maintained the traditional net outcomes and the traditional frequency-of-losses structure, and the reward structure stayed the same within the good decks and within the bad decks. freq. = frequent; infreq. = infrequent.

participants. We therefore searched for IGT studies reporting the mean proportion of choices from the good and bad decks. These studies were located by searching for IGT studies published before July 1, 2011, in ISI Web of Knowledge and Google Scholar (search term: *Iowa Gambling Task*).

We restricted this literature review to studies that presented the mean proportion of choices from the good and bad decks in the text or tables, or provided figures that enabled us to extract this information. The inclusion criteria are the same as described for the first literature review, such as only IGT studies that used a procedure and payoff scheme close to the ones described by Bechara et al. (1994). Our second literature review involves 39 groups with 1,427 healthy participants. Because of the vast number of published IGT studies, this review is not exhaustive; however, we considered it large enough to address our research question. To the best of our knowledge, our review of healthy participants' performance on the IGT is the most comprehensive to date.

Published data sets. We e-mailed all corresponding authors of the studies included in at least one of the literature reviews and requested the data of the healthy participants. Despite three attempts to contact the authors, we only received seven data sets that contained the information we needed. Table 2 gives an overview of all groups of healthy participants included in the literature reviews and specifies in which literature review each group is included and whether the data have been obtained. Data are labeled as not available if the data could no longer be located, if the complete format of the data was not available (e.g., only available collapsed over the two good decks and the two bad decks), or if the authors did not send us the data by the time we completed this article. The data provided by Wood, Busemeyer, Koling, Cox, and Davis (2005; $N = 153$) were collapsed into a single group (age: $M = 45.25$ years, $SD = 27.21$).

Own data set. We conducted an experiment that featured 162 healthy participants (82 female, 80 male) performing a 100-trial IGT. We used a computerized version of the IGT with facsimile money and the payoff scheme as developed by Bechara et al. (1994). The mean age of all participants was 25.56 years ($SD = 4.86$). Female participants had a mean age of 25.52 years ($SD = 5.32$), and male participants of 25.60 years ($SD = 4.37$). All participants reported an unremarkable neurological history and were paid for their participation.

Evidence Against Assumption 1: Frequency-of-Losses Effect

Literature Review 1: Choices from each deck separately. To assess whether the choices of healthy participants are guided primarily by the long-term outcomes of the decks, we conducted a literature review on the overall proportion of choices from each of the four decks. Only 17 studies with 479 participants met our inclusion criteria; Table 3 shows the mean proportion of choices from each deck for each of the studies separately. Out of these 17 studies, 13 studies reported that Deck B (i.e., the bad deck with infrequent losses) was chosen more often or about as often as Deck C or D (i.e., the two good decks). The remaining four studies showed that healthy participants tended to prefer the good decks. One of these four studies is the study of Bechara et al. (1994); note that none of the remaining three studies yielded an overall preference for the good decks that is stronger than the one reported by Bechara et al.

The frequency-of-losses effect is evident from a consideration of the weighted mean proportion of choices from the different decks (i.e., Deck A: .19; Deck B: .29; Deck C: .24; Deck D: .29)—participants prefer the decks with infrequent losses over the decks with frequent losses. These findings suggest that the choices of healthy participants are primarily influenced by the frequency of losses rather than the expected valence of the decks.

Experimental data. The seven received data sets and our own data set enabled us to consider the mean proportion of choices from each deck as a function of trial number, as displayed in Figure 4. This figure shows that only the data set of Premkumar et al. (2008) supports the importance of long-term outcomes for deck selection frequencies; in this data set, healthy participants gradually developed a preference for both good decks (i.e., Decks C and D) over both bad decks. The remaining data sets instead support

Table 4
Mean Proportion of Choices From Each Deck Separately of Healthy Participants in Each Data Set That Was Sent to Us Upon Request, and in Our Own Data Set

Study	Sample size[a]	Deck A: Bad deck with frequent losses	Deck B: Bad deck with infrequent losses	Deck C: Good deck with frequent losses	Deck D: Good deck with infrequent losses
Fernie & Tunney (2006)	20	.20	.33	.22	.25
Fridberg et al. (2010)	15	.13	.30	.14	.43
Kjome et al. (2010)[b]	19	.17	.26	.22	.35
Premkumar et al. (2008)[b]	25	.16	.21	.27	.36
Rodriguez-Sánchez et al. (2005)	19	.16	.30	.20	.34
Toplak et al. (2005)	34	.23	.30	.20	.27
Wood et al. (2005)[b]	153	.17	.29	.23	.31
Own data set	162	.15	.35	.20	.30
Total sample size[c] (M)	447	.17	.31	.21	.31

Note. All studies are based on a 100-trial Iowa Gambling Task except Fridberg et al. (2010; 95 trials).
[a] Sample sizes might be lower than in Tables 2, 3, and 5 due to incomplete received data sets. [b] Studies did not use exactly the same payoff scheme as reported by Bechara et al. (1994); these studies nevertheless met our inclusion criteria because they maintained the traditional net outcomes and the traditional frequency of losses structure, and the reward structure stayed the same within the good decks and within the bad decks. [c] Weighted by the sample size of each study.

the frequency-of-losses effect. Some data sets show a pronounced frequency-of-losses effect with a clear preference for both decks with infrequent losses (i.e., Decks B and D) over both decks with frequent losses (i.e., Fridberg et al., 2010; Rodríguez-Sánchez et al., 2005; our own data set). Other data sets show a less pronounced frequency-of-losses effect, indicating that, at the end of the IGT, participants chose about equally often from Decks B, C, and D, while clearly avoiding Deck A (i.e., Fernie & Tunney, 2006; Kjome et al., 2010; Wood et al., 2005). In the data set of Toplak, Jain, and Tannock (2005), participants show a slight preference for decks with infrequent losses over the decks with frequent losses during the entire sequence of trials; however, at the end of the IGT, participants chose about equally often from all decks. Table 4 summarizes the information of Figure 4 by presenting the overall mean proportion of choices from each deck. The weighted mean proportion of choices from each deck of the seven received data sets and our own data set underscore the frequency-of-losses effect (i.e., Deck A: .17; Deck B: .31; Deck C: .21; Deck D: .31) and thus corroborate the findings of the first literature review (see Table 3).

Evidence Against Assumption 2: High Variability in Performance Between and Within Groups

Literature Review 2: Choices from the good and bad decks. To assess whether the choices of healthy participants are relatively consistent between groups, we conducted a literature review on the mean proportion of choices from the good decks (i.e., Decks C and D) and the bad decks (i.e., Decks A and B). Table 5 presents the mean proportion of choices from good and bad decks for all groups included in our second literature review, and Figure 5 presents the histogram of the corresponding mean proportion of choices from the good decks. From the table and the figure, it is evident that the majority of the studies reported a weak overall preference for the good decks (i.e., between 50% and 60% choices from the good decks), and that some studies even reported an overall preference

Table 5
Literature Review 2: Mean Proportion of Choices From Good Decks (C and D) and Bad Decks (A and B) of Healthy Participants

Study	Sample size	Age[a]	Good decks	Bad decks	Good and bad decks
Caroselli et al. (2006)	141	$M = 21.7$, $SD = 4.6$	.43	.57	−.14
Fernie & Tunney (2006)	20	Undergraduates	.47	.53	−.06
Toplak et al. (2005)	34	$M = 15.4$, $SD = 1.5$	.47	.53	−.06
Bark et al. (2005)	26	$M = 29.8$, $SD = 9.4$	.48	.52	−.04
Sevy et al. (2007)	20	$M = 33.0$, $SD = 10.0$	.50	.50	.00
Rodríguez-Sánchez et al. (2005)	22	$M = 26.1$, $SD = 6.5$	.51	.49	.02
Yechiam, Hayden, et al. (2008)	25	$M = 39.2$, $SD = 13.3$	.52	.48	.04
Overman et al. (2004)	60	7th grade (~12 years)	.53	.47	.06
Wilder et al. (1998)	30	18–52	.53	.47	.06
Kester et al. (2006)	25	$M = 17.1$, $SD = 1.8$	.54	.46	.08
Lehto & Elorinne (2003)	51	7–11	.54	.46	.08
Wood et al. (2005)	88	$M = 22.1$, $SD = 4.5$	.54	.46	.08
Cavedini, Riboldi, D'Annucci, et al. (2002)	34	$M = 29.5$, $SD = 8.9$	.55	.45	.10
Cavedini, Riboldi, Keller, et al. (2002)	40	$M = 30.3$, $SD = 9.6$	.55	.45	.10
Wood et al. (2005)	67	$M = 77.3$, $SD = 4.6$	.55	.45	.10
Lehto & Elorinne (2003)	40	19–53	.56	.44	.12
Lev et al. (2008)	36	$M = 36.9$, $SD = 12.2$	.56	.44	.12
Fridberg et al. (2010)	15	$M = 29.6$, $SD = 7.6$	.57	.43	.14
Kjome et al. (2010)	20	$M = 33.9$, $SD = 11.2$	.57	.43	.14
O'Carroll & Papps (2003)	11	$M = 20.0$, $SD = 3.1$	.57	.43	.14
Ritter et al. (2004)	15	$M = 47.1$, $SD = 10.2$	.57	.43	.14
Johnson et al. (2006)	14	$M = 15.9$, $SD = 2.4$	.58	.42	.16
Martino et al. (2007)	15	$M = 35.0$, $SD = 10.9$	.58	.42	.16
Overman et al. (2004)	35	6th grade (~11 years)	.58	.42	.16
Overman et al. (2004)	57	8th grade (~13 years)	.59	.41	.18
Petry (2001)	21	$M = 36.1$, $SD = 11.5$	.59	.41	.18
Blair et al. (2001)	23	10.5–13.9	.60	.40	.20
Bechara et al. (1998)	21	24–68	.62	.38	.24
Bechara et al. (1999)	13	22–58	.63	.37	.26
Premkumar et al. (2008)	25	$M = 35.4$, $SD = 11.9$	.63	.37	.26
Overman et al. (2004)	59	9th grade (~14 years)	.64	.36	.28
Overman et al. (2004)	60	10th grade (~15 years)	.64	.36	.28
Shurman et al. (2005)	10	$M = 32.1$, $SD = 4.5$	.66	.34	.32
Tomb et al. (2002)	10	Undergraduates	.66	.34	.32
Overman et al. (2004)	60	11th grade (~16 years)	.67	.33	.34
Overman et al. (2004)	60	17–23	.67	.33	.34
Overman et al. (2004)	60	12th grade (~17 years)	.68	.32	.36
Bechara et al. (1994)	44	20–79	.70	.30	.40
North & O'Carroll (2001)	20	$M = 30.8$, $SD = 1.9$	.70	.30	.40

Note. All studies are based on a 100-trial Iowa Gambling Task except Fridberg et al. (2010; 95 trials) and Overman et al. (2004; 200 trials).
[a] Age (in years) as mean and standard deviation, range, profession, or school grade with the approximate age in parentheses.

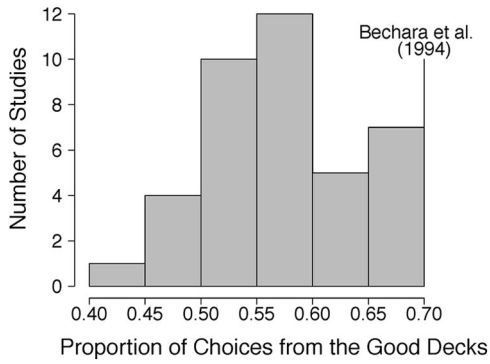

Figure 5. Literature Review 2: Histogram of the mean proportion of choices from good decks (i.e., Decks C and D) of 39 groups of healthy participants. The majority of the groups show a weak overall preference for the good decks, whereas some groups show an overall preference for the bad decks and a few groups show a pronounced overall preference for the good decks, as reported by Bechara et al. (1994).

for the bad decks. Few studies reported a pronounced overall preference for the good decks, and none reported a preference for the good decks that is as strong as the one reported by Bechara et al. (1994).

Our second literature review reveals substantial between-group variability in deck preference. In addition, the review underscores that the performance of healthy participants is generally much worse than initially reported by Bechara et al. (1994). It is especially striking that the majority of the studies reported a weak overall preference for the good decks. The strong evidence for the frequency-of-losses effect, discussed in the last section, suggests that a weak overall preference for the good decks may only be due to a summing effect. For example, the data of Fridberg et al. (2010) show that healthy participants prefer the good decks over the bad decks (i.e., .57 vs. .43; see Table 5). This finding is underscored by the left panel of Figure 6, which presents the mean proportion of good and bad choices as a function of trial number (cf. Fridberg et al., 2010, Figure 1). This panel shows that the proportion of good choices increases as the number of trials increases and leads the authors to conclude that "controls subsequently learned to select from the advantageous decks" (p. 31). However, this conclusion is mainly due to a summing effect, and hides the frequency-of-losses effect, as can be seen in our first literature review, where the mean proportion of choices from each deck were presented separately (i.e., Deck A: .13; Deck B: .30; Deck C: .14; Deck D: .43; see Table 3). The right panel of Figure 6 (cf. Fridberg et al., 2010, Figure 2) displays the mean proportion of choices from each deck as a function of trial number. This figure shows that the decks with infrequent losses (i.e., Decks B and D) are at any point during the IGT preferred over the decks with frequent losses (i.e., Decks A and C). The reason why the control participants appear to learn to prefer the good decks is that the choices from Deck D strongly increase, whereas the choices from the other three decks slowly decrease; nevertheless, Deck B is still chosen much more often than Deck A or C at any point during the IGT. This example illustrates how collapsing the choices from the good decks and from the bad decks might hide the frequency-of-losses effect and highlights the importance of analyzing the choice behavior of each deck separately.

While reviewing the literature, we also noticed that some studies reported a high within-group variability in the performance of healthy participants on the IGT. We found six studies that explicitly reported the number of healthy participants who showed impaired performance (i.e., overall preference for the bad decks). Table 6 lists these six studies and the corresponding number and percentage of healthy participants with impaired performance. This table points out that a substantial number of healthy participants perform poorly on the IGT, suggesting that there is high within-group variability in the performance of healthy participants in these six studies. Moreover, the table suggests that a substantial number of participants violate the assumption that healthy participants prefer the good decks over the bad decks.

In sum, the performance of healthy participants on the IGT is characterized by high variability between groups: In some studies healthy participants seem to behave according to the assumptions of Bechara et al. (1994) by expressing a strong preference for the good decks; however, in most studies the preference for the good decks is relatively weak, and in some studies healthy participants express a preference for the bad decks, in complete contradiction

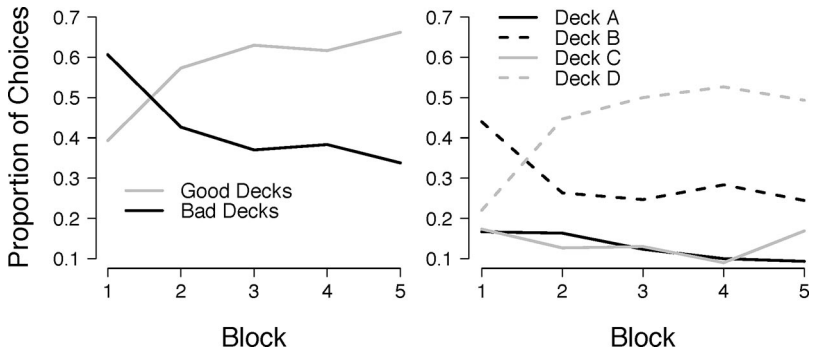

Figure 6. Choice behavior of healthy participants in Fridberg et al. (2010). The left panel shows the mean proportion of choices from good and bad decks within five blocks. Each block contains 20 trials, except the last block (15 trials). The right panel shows the mean proportion of choices from each deck within five blocks. This data set illustrates how collapsing the choices from the good decks and from the bad decks might hide the frequency-of-losses effect.

to the assumptions that underlie the IGT. In addition, a few data sets suggest that there may also be substantial variability across participants within the same study, a finding we will return to in more detail in the next section.

Experimental data. The extent of within-group variability in performance of healthy participants can be understood best by an individual-participants analysis of deck selection profiles. Figure 7 shows the deck selection profiles of five participants in our own data set. This figure illustrates that choice behavior differs dramatically between the five healthy participants: The first participant keeps switching between all decks and does not develop a preference for any deck; the second participant seems to develop a preference for the bad decks with occasional switches to Deck C; the third and fourth participants both develop a preference for Decks B and D, but the third participant tends to stay longer at the same deck than the fourth participant; finally, the fifth participant develops a preference for Deck B and only rarely switches to Deck C or D. These individual deck selection profiles suggest that each participant displays a highly idiosyncratic choice behavior. The deck selection profiles of all available data sets further underscore the substantial within-group variability, as can be seen in the online appendix containing 394 individual deck selection profiles from six studies.[3]

The extent of within-group variability can be further assessed by determining the number of participants who perform in the deficient range (i.e., >50% choices from the bad decks). Table 7 presents the number and percentage of healthy participants with impaired performance in each data set that was sent to us upon request, and in our own data set. From this table, it is apparent that a substantial number of healthy participants—in some studies, even the majority of the participants—perform deficiently on the IGT. This table corroborates the findings from Table 6 and Figure 7: The performance of healthy participants is characterized by high within-group variability, and a substantial number of participants violate the assumption that healthy participants prefer the good decks over the bad decks.

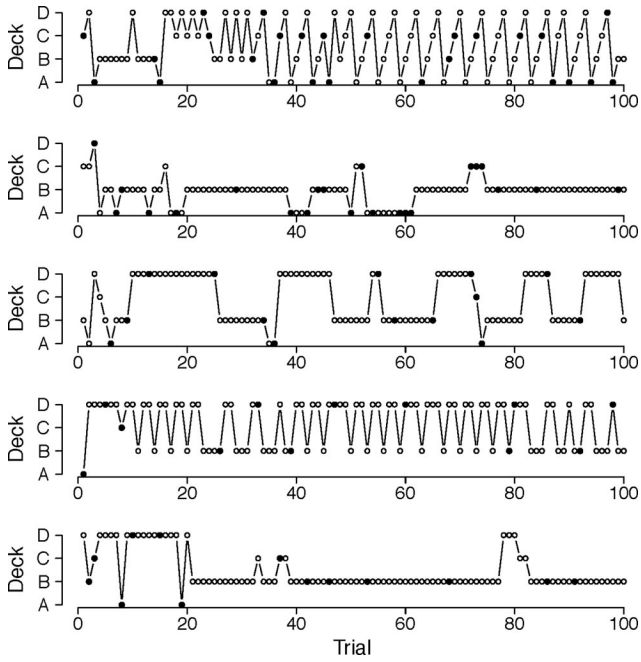

Figure 7. Deck selection profiles of five participants in our own data set express the high within-group variability in the performance of healthy participants. The filled circles indicate the occurrence of rewards and losses together; the empty circles indicate the occurrence of only rewards.

Evidence Against Assumption 3: Absence of a Pronounced Exploration–Exploitation Trade-Off

Bechara et al. (1994) assumed that healthy participants first explore all decks, then switch to the good decks, and only occasionally return to the bad decks. In a reinforcement learning context (Sutton & Barto, 1998), this means that healthy participants progress from an initial stage of exploration to a later stage of exploitation. This assumption implies that the number of switches decreases as the number of trials increases. However, the available data sets fail to support this assumption. Figure 8 shows the mean proportion of switches as a function of trial number. The data sets of Premkumar et al. (2008) and Wood et al. (2005) show a slight decrease in the mean proportion of switches across trials; however, this decrease is not as pronounced as assumed by Bechara et al., who showed that, in the case of two typical healthy participants, the last 10 trials featured only one and no switch (cf. Figure 3). In the remaining data sets, the mean proportion of switches remains surprisingly stable. Therefore, Figure 8 supports the conclusion that healthy participants fail to progress from an initial stage of exploration to a later stage of exploitation.[4] In a reinforcement learning context (Sutton & Barto, 1998), this suggests that participants failed to learn about the relative value of the choice options that they had available.

Table 6
Number and Percentage of Healthy Participants With Impaired Performance (>50% Choices From the Bad Decks) as Specified in Six Studies

Study	Impaired performance	
	n	%
Adinoff et al. (2003)	4	27.0
Bechara et al. (1999)	3	23.0
Bechara et al. (2001)	18[a]	37.0
Bechara & Damasio (2002)	13[a]	32.5
Bowman & Turnbull (2003)	2	12.0
Lehto & Elorinne (2003)	3[b]	7.5

[a] These participants were classified as impaired because they performed within the range of ventromedial prefrontal cortex patients. [b] These participants even made ≥70% choices from the bad decks.

[3] The deck selection profiles of the six data sets can be downloaded (http://dl.dropbox.com/u/12798592/DeckSelectionProfiles.zip). Note that the data sets of Rodríguez-Sánchez et al. (2005) and Toplak et al. (2005) could not be included in this analysis because we only obtained these data in bins of 20 trials.

[4] As in the previous section, the data sets of Rodríguez-Sánchez et al. (2005) and Toplak et al. (2005) could not be included in this analysis because we only obtained these data in bins of 20 trials.

Table 7
Number and Percentage of Healthy Participants With Impaired Performance (>50% Choices From the Bad Decks) in Each Data Set That Was Sent to Us Upon Request, and in Our Own Data Set

Study	Impaired performance	
	n	%
Fernie & Tunney (2006)	13	65.0
Fridberg et al. (2010)	5	33.3
Kjome et al. (2010)[a]	6	31.6
Premkumar et al. (2008)[a]	5	20.0
Rodriguez-Sánchez et al. (2005)	9	47.4
Toplak et al. (2005)	20	58.8
Wood et al. (2005)[a]	55	35.9
Own data set	79	48.8

Note. All studies are based on a 100-trial Iowa Gambling Task except Fridberg et al. (2010; 95 trials).
[a] Studies did not use exactly the same payoff scheme as reported by Bechara et al. (1994); these studies nevertheless met our inclusion criteria because they maintained the traditional net outcomes and the traditional frequency-of-losses structure, and the reward structure stayed the same within the good decks and within the bad decks.

Discussion

This article focused on performance of healthy participants on the IGT. Two literature reviews and the analysis of eight data sets together challenge the assumptions of Bechara et al. (1994) about the IGT choice behavior of healthy participants. Our findings reveal the presence of a frequency-of-losses effect in healthy participants, as already indicated by many studies (Ahn et al., 2008; Caroselli et al., 2006; Chiu & Lin, 2007; Chiu et al., 2008; Dunn et al., 2006; Huizenga et al., 2007; Lin et al., 2007; MacPherson et al., 2002; Wilder et al., 1998; Yechiam & Busemeyer, 2005). In addition, we showed that performance of healthy participants is characterized by considerable variability, both across groups and across participants within the same study, and that healthy participants fail to progress from an initial stage of exploration to a later stage of exploitation. These findings clearly contradict the common belief that "most healthy participants sample cards from each deck, and after about 40 or 50 selections are fairly good at sticking to the good decks" ("Iowa Gambling Task," 2012).

This evidence against the key IGT assumptions suggests that IGT scores may not serve as a valid measure of decision-making deficits in clinical populations, for two reasons (cf. Dunn et al., 2006). First, the high variability in healthy participants' performance complicates the prevailing interpretation of IGT data. Specifically, conclusions about deficient choice performance in clinical populations depend strongly on the performance of the control group. For instance, the absence of a difference between clinical and control groups may be due to intact decision-making ability in the clinical group, or it may be due to poor performance of the control group (Dunn et al., 2006).

Second, our findings question the ecological validity of IGT scores. Since a substantial number of healthy participants perform poorly on the IGT, but likely do not show any decision-making deficits in real life, it is unclear to what extent IGT scores measure everyday, real-life decision making (Dunn et al., 2006).

Our analysis demonstrated that in order to reveal the frequency-of-losses effect and accurately describe the learning process, it is important that future studies analyze the choice behavior as a function of the trial number from each deck separately. One way to capture the learning process is to apply trend analyses (e.g., a linear or quadratic contrast) on the choice pattern from each deck (Dunn et al., 2006; Premkumar et al., 2008).

Our results showed that many healthy participants perform poorly on the IGT, begging the question as to why they do not learn to prefer both good decks over both bad decks. Altogether, it seems that many participants find it particularly difficult to figure out that Deck B is a bad deck—after all, Deck B yields high immediate, constant rewards. One explanation for why healthy participants do not learn to prefer both good decks over Deck B is related to the payoff scheme as developed by Bechara et al. (1994). As first pointed out by Lin et al. (2007), cards from Deck C never yield a net loss, but cards from Deck A do. This goes against Bechara et al.'s idea of designing one good deck and one bad deck with high-frequent losses and suggests that the good decks and the bad decks are pseudobalanced. Decks B, C, and D are more similar than expected because they yield either no or very few net losses, whereas Deck A yields frequent and big net losses and is thus the

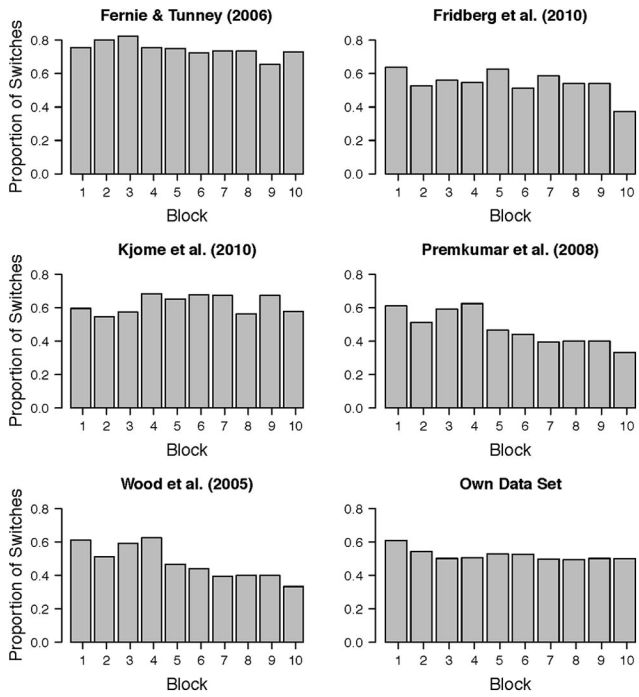

Figure 8. Mean proportion of switches within 10 blocks in each data set that was sent to us upon request, and in our own data set. Each block contains 10 trials, except the first block in all data sets (nine trials) and the last block in Fridberg et al.'s (2010) data set (five trials). The mean proportion of switches remains stable across trials in each data set, indicating the absence of a pronounced exploration–exploitation trade-off. Note that Kjome et al. (2010), Premkumar et al. (2008), and Wood et al. (2005) did not use the exact payoff scheme as reported by Bechara et al. (1994); these studies nevertheless met our inclusion criteria because they maintained the traditional net outcomes and the traditional frequency-of-losses structure, and the reward structure stayed the same within the good decks and within the bad decks.

only deck that clearly differs from the others with respect to the net outcomes. Another explanation is that the payoff scheme encourages participants to focus only on the immediate losses; the losses vary, but the immediate rewards are completely predictable, as they are constant across trials for each deck. The predictability of immediate rewards also contradicts Bechara et al.'s intention of "uncertainty of reward and punishment" (p. 8).

Another explanation for the surprisingly poor performance of healthy participants is a lack of motivation. Indications for a lack of motivation can be seen in the individual deck selection profiles, as shown in Figure 7 and in the online appendix. Studying these profiles, it seems impossible to identify a learning process for the majority of the participants. Many participants keep switching between the decks, irrespective of the occurrence of losses and of the long-term outcomes of the decks. However, Fernie and Tunney (2006) have shown that participants do not perform better when they are more motivated to do well (i.e., by using real money vs. facsimile money); instead, participants did improve when they were given more information about the task (i.e., some decks are worse than others, most money is made when the poor options are avoided). This suggests that it is the difficulty of the IGT rather than the lack of motivation that contributes to a poor performance of healthy participants.

One final explanation for the poor performance of healthy participants is that 100 trials are not enough to learn much about the nature of the decks. In support of this conjecture, Wetzels, Vandekerckhove, Tuerlinckx, and Wagenmakers (2010) showed that healthy participants do learn to prefer the good decks over the bad decks, but that they require at least 100 trials to do so. The reason for the slow learning process might be the infrequent occurrence of losses in Decks B and D (i.e., once in 10 cards), providing participants with too little information to learn quickly that Deck B should be avoided.

It is still unclear what accounts for the relatively poor performance of clinical populations on the IGT. It may be that clinical populations are even more susceptible to the frequency-of-losses effect than healthy participants are. It may also be that clinical populations are primarily driven by the immediate prospects of the decks, as originally proposed by Bechara et al. (1994). In order to understand the factors that determine the choices on the IGT, one possibility is to fit a multivariate normal mixture model to clinical samples and samples of healthy participants (cf. Huizenga et al., 2007). This analysis allows one to identify groups that use distinct IGT decision strategies, such as focusing on the frequency of losses, focusing on immediate rewards, or focusing on long-term outcomes. Such an analysis does not explain the learning process and therefore requires a relatively long IGT. Another possibility to determine the factors driving performance on the IGT is to apply a formal model that explicitly describes how the learning process unfolds over time (e.g., Busemeyer & Stout, 2002).

In sum, we showed that the IGT is characterized by several anomalous phenomena that go against the assumptions of the prevailing theory (Bechara et al., 1994). Specifically, performance of healthy participants is primarily influenced by the frequency of losses rather than the long-term outcomes; moreover, performance of healthy participants is highly idiosyncratic, and often lacks a tendency to first explore decks and then exploit those that are most attractive. These anomalies question the extent to which IGT scores measure real-life decision-making deficits and complicate the prevailing interpretation of IGT data. Our findings suggest that the IGT needs to be reevaluated before being accepted as convincing tool to measure decision-making deficits in clinical populations.

References

Adinoff, B., Devous, M. D., Cooper, D. B., Best, S. E., Chandler, P., Harris, T., . . . Cullum, M. (2003). Resting regional cerebral blood flow and gambling task performance in cocaine-dependent subjects and healthy comparison subjects. *American Journal of Psychiatry, 160,* 1892–1894. doi:10.1176/appi.ajp.160.10.1892

Ahn, W.-Y., Busemeyer, J. R., Wagenmakers, E.-J., & Stout, J. C. (2008). Comparison of decision learning models using the generalization criterion method. *Cognitive Science, 32,* 1376–1402. doi:10.1080/03640210802352992

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

Bechara, A. (2007). *Iowa Gambling Task (IGT) 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

Bechara, A., & Damasio, H. (2002). Decision-making and addiction (Part I): Impaired activation of somatic states in substance dependent individuals when pondering decisions with negative future consequences. *Neuropsychologia, 40,* 1675–1689. doi:10.1016/S0028-3932(02)00015-5

Bechara, A., Damasio, H., Damasio, A. R., & Lee, G. P. (1999). Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making. *Journal of Neuroscience, 19,* 5473–5481.

Bechara, A., Damasio, H., Tranel, D., & Anderson, S. W. (1998). Dissociation of working memory from decision making within the human prefrontal cortex. *Journal of Neuroscience, 18,* 428–437.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997, February 28). Deciding advantageously before knowing the advantageous strategy. *Science, 275*, 1293–1295. doi:10.1126/science.275.5304.1293

Bechara, A., Dolan, S., Denburg, N., Hindes, A., Anderson, S. W., & Nathan, P. E. (2001). Decision-making deficits, linked to a dysfunctional ventromedial prefrontal cortex, revealed in alcohol and stimulant abusers. *Neuropsychologia, 39,* 376–389. doi:10.1016/S0028-3932(00)00136-6

Bechara, A., Tranel, D., & Damasio, H. (2000). Characterization of the decision-making deficit of patients with ventromedial prefrontal cortex lesions. *Brain, 123,* 2189–2202. doi:10.1093/brain/123.11.2189

Blair, R. J. R., Colledge, E., & Mitchell, D. G. V. (2001). Somatic markers and response reversal: Is there orbitofrontal cortex dysfunction in boys with psychopathic tendencies? *Journal of Abnormal Child Psychology, 29,* 499–511. doi:10.1023/A:1012277125119

Boes, A. D., Grafft, A. H., Joshi, C., Chuang, N. A., Nopoulos, P., & Anderson, S. W. (2011). Behavioral effects of congenital ventromedial prefrontal cortex malformation. *BMC Neurology, 11,* 151–162. doi:10.1186/1471-2377-11-151

Bowman, C. H., Evans, C. E. Y., & Turnbull, O. H. (2005). Artificial time constraints on the Iowa Gambling Task: The effects on behavioural performance and subjective experience. *Brain and Cognition, 57,* 21–25. doi:10.1016/j.bandc.2004.08.015

Bowman, C. H., & Turnbull, O. H. (2003). Real versus facsimile reinforcers on the Iowa Gambling Task. *Brain and Cognition, 53,* 207–210. doi:10.1016/S0278-2626(03)00111-8

Busemeyer, J. R., & Stout, J. C. (2002). A contribution of cognitive decision models to clinical assessment: Decomposing performance on the Bechara gambling task. *Psychological Assessment, 14,* 253–262. doi:10.1037/1040-3590.14.3.253

Caroselli, J. S., Hiscock, M., Scheibel, R. S., & Ingram, F. (2006). The simulated gambling paradigm applied to young adults: An examination

of university students' performance. *Applied Neuropsychology, 13,* 203–212. doi:10.1207/s15324826an1304_1

Cavedini, P., Riboldi, G., D'Annucci, A., Belotti, P., Cisima, M., & Bellodi, L. (2002). Decision-making heterogeneity in obsessive-compulsive disorder: Ventromedial prefrontal cortex function predicts different treatment outcomes. *Neuropsychologia, 40,* 205–211. doi:10.1016/S0028-3932(01)00077-X

Cavedini, P., Riboldi, G., Keller, R., D'Annucci, A., & Bellodi, L. (2002). Frontal lobe dysfunction in pathological gambling patients. *Biological Psychiatry, 51,* 334–341. doi:10.1016/S0006-3223(01)01227-6

Chiu, Y.-C., & Lin, C.-H. (2007). Is Deck C an advantageous deck in the Iowa Gambling Task? *Behavioral and Brain Functions*, *3,* 37. doi:10.1186/1744-9081-3-37

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

Damasio, A. R., Everitt, B. J., & Bishop, D. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philosophical Transactions of the Royal Society B: Biological Sciences, 351,* 1413–1420. doi:10.1098/rstb.1996.0125

Dunn, B. D., Dalgleish, T., & Lawrence, A. D. (2006). The somatic marker hypothesis: A critical evaluation. *Neuroscience & Biobehavioral Reviews, 30,* 239–271. doi:10.1016/j.neubiorev.2005.07.001

Fernie, G., & Tunney, R. J. (2006). Some decks are better than others: The effect of reinforcer type and task instructions on learning in the Iowa Gambling Task. *Brain and Cognition, 60,* 94–102. doi:10.1016/j.bandc.2005.09.011

Fridberg, D. J., Queller, S., Ahn, W.-Y., Kim, W., Bishara, A. J., Busemeyer, J. R., . . . Stout, J. C. (2010). Cognitive mechanisms underlying risky decision-making in chronic cannabis users. *Journal of Mathematical Psychology, 54,* 28–38. doi:10.1016/j.jmp.2009.10.002

Huizenga, H. M., Crone, E. A., & Jansen, B. J. (2007). Decision-making in healthy children, adolescents and adults explained by the use of increasingly complex proportional reasoning rules. *Developmental Science, 10,* 814–825. doi:10.1111/j.1467-7687.2007.00621.x

Iowa Gambling Task. (2012, February 3). *in Wikipedia*. Retrieved from http://en.wikipedia.org/wiki/Iowa_gambling_task

Johnson, S. A., Yechiam, E., Murphy, R. R., Queller, S., & Stout, J. C. (2006). Motivational processes and autonomic responsivity in Asperger's disorder: Evidence from the Iowa Gambling Task. *Journal of the International Neuropsychological Society, 12,* 668–676. doi:10.1017/S1355617706060802

Kester, H. M., Sevy, S., Yechiam, E., Burdick, K. E., Cervellione, K. L., & Kumra, S. (2006). Decision-making impairments in adolescents with early-onset schizophrenia. *Schizophrenia Research, 85,* 113–123. doi:10.1016/j.schres.2006.02.028

Kjome, K. L., Lane, S. D., Schmitz, J. M., Green, C., Ma, L., Prasla, I., . . .Moeller, F. G. (2010). Relationship between impulsivity and decision making in cocaine dependence. *Psychiatry Research, 178,* 299–304. doi:10.1016/j.psychres.2009.11.024

Lehto, J. E., & Elorinne, E. (2003). Gambling as an executive function task. *Applied Neuropsychology, 10,* 234–238. doi:10.1207/s15324826an1004_5

Lev, D., Hershkovitz, E., & Yechiam, M. (2008). Decision making and personality in traffic offenders: A study of Israeli drivers. *Accident Analysis and Prevention, 40,* 223–230. doi:10.1016/j.aap.2007.05.009

Lin, C.-H., Chiu, Y.-C., Lee, P.-L., & Hsieh, J.-C. (2007). Is Deck B a disadvantageous deck in the Iowa Gambling Task? *Behavioral and Brain Functions*, *3,* 16. doi:10.1186/1744-9081-3-16

MacPherson, S. E., Phillips, L. H., & Della Sala, S. (2002). Age, executive function, and social decision making: A dorsolateral prefrontal theory of cognitive aging. *Psychology and Aging, 17,* 598–609. doi:10.1037/0882-7974.17.4.598

Maia, T. V., & McClelland, J. L. (2004). A reexamination of the evidence for the somatic marker hypothesis: What participants really know in the Iowa Gambling Task. *Proceedings of the National Academy of Sciences of the United States of America, 101,* 16075–16080. doi:10.1073/pnas.0406666101

Martino, D. J., Bucay, D., Butman, J. T., & Allegri, R. F. (2007). Neuropsychological frontal impairments and negative symptoms in schizophrenia. *Psychiatry Research, 152,* 121–128. doi:10.1016/j.psychres.2006.03.002

North, N. T., & O'Carroll, R. E. (2001). Decision making in patients with spinal cord damage: Afferent feedback and the somatic marker hypothesis. *Neuropsychologia, 39,* 521–524. doi:10.1016/S0028-3932(00)00107-X

O'Carroll, R. E., & Papps, B. P. (2003). Decision making in humans: The effect of manipulating the central noradrenergic system. *Journal of Neurology, Neurosurgery & Psychiatry, 74,* 376–378. doi:10.1136/jnnp.74.3.376

Overman, W. H., Frassrand, K., Ansel, S., Trawalter, S., Bies, B., & Redmond, A. (2004). Performance on the Iowa Card Task by adolescents and adults. *Neuropsychologia, 42,* 1838–1851. doi:10.1016/j.neuropsychologia.2004.03.014

PAR. (2012). Iowa Gambling Task (IGT). Retrieved from http://www4.parinc.com/Products/Product.aspx?ProductID=IGT

Petry, N. M. (2001). Substance abuse, pathological gambling, and impulsiveness. *Drug and Alcohol Dependence, 63,* 29–38. doi:10.1016/S0376-8716(00)00188-5

Premkumar, P., Fannon, D., Kuipers, E., Simmons, A., Frangou, S., & Kumari, V. (2008). Emotional decision-making and its dissociable components in schizophrenia and schizoaffective disorder: A behavioural and MRI investigation. *Neuropsychologia, 46,* 2002–2012. doi:10.1016/j.neuropsychologia.2008.01.022

Ritter, L. M., Meador-Woodruff, J. H., & Dalack, G. W. (2004). Neurocognitive measures of prefrontal cortical dysfunction in schizophrenia. *Schizophrenia Research, 68,* 65–73. doi:10.1016/S0920-9964(03)00086-0

Rodríguez-Sánchez, J. M., Crespo-Facorro, B., Perez Iglesias, R., González-Blanch Bosch, C., Álvarez, M., Llorca, J., & Vázquez-Barquero, J. L. (2005). Prefrontal cognitive functions in stabilized first-episode patients with schizophrenia spectrum disorders: A dissociation between dorsolateral and orbitofrontal functioning. *Schizophrenia Research, 77,* 279–288. doi:10.1016/j.schres.2005.04.023

Sevy, S., Burdick, K. E., Visweswaraiah, H., Abdelmessih, S., Lukin, M., Yechiam, E., & Bechara, A. (2007). Iowa Gambling Task in schizophrenia: A review and new data in patients with schizophrenia and co-occurring cannabis use disorders. *Schizophrenia Research, 92,* 74–84. doi:10.1016/j.schres.2007.01.005

Shurman, B., Horan, W. P., & Nuechterlein, K. H. (2005). Schizophrenia patients demonstrate a distinctive pattern of decision-making impairment on the Iowa Gambling Task. *Schizophrenia Research, 72,* 215–224. doi:10.1016/j.schres.2004.03.020

Stout, J. C., Busemeyer, J. R., Lin, A., Grant, S. J., & Bonson, K. R. (2004). Cognitive modeling analysis of decision-making processes in cocaine abusers. *Psychonomic Bulletin & Review, 11,* 742–747. doi:10.3758/BF03196629

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.

Tomb, I., Hauser, M., Deldin, P., & Caramazza, A. (2002). Do somatic markers mediate decisions on the gambling task? *Nature Neuroscience, 5,* 1103–1104. doi:10.1038/nn1102-1103

Toplak, M. E., Jain, U., & Tannock, R. (2005). Executive and motivational processes in adolescents with attention-deficit-hyperactivity disorder (ADHD). *Behavioral and Brain Functions*, *1,* 8. doi:10.1186/1744-9081-1-8

Toplak, M. E., Sorge, G. B., Benoit, A., West, R. F., & Stanovich, K. E. (2010). Decision-making and cognitive abilities: A review of associations between Iowa Gambling Task performance, executive functions, and intelligence. *Clinical Psychology Review, 30,* 562–581. doi:10.1016/j.cpr.2010.04.002

Wetzels, R., Vandekerckhove, J., Tuerlinckx, F., & Wagenmakers, E.-J. (2010). Bayesian parameter estimation in the expectancy valence model of the Iowa Gambling Task. *Journal of Mathematical Psychology, 54,* 14–27. doi:10.1016/j.jmp.2008.12.001

Wilder, K. E., Weinberger, D. R., & Goldberg, T. E. (1998). Operant conditioning and the orbitofrontal cortex in schizophrenic patients: Unexpected evidence for intact functioning. *Schizophrenia Research, 30,* 169–174. doi:10.1016/S0920-9964(97)00135-7

Wood, S., Busemeyer, J., Koling, A., Cox, C. R., & Davis, H. (2005). Older adults as adaptive decision makers: Evidence from the Iowa Gambling Task. *Psychology and Aging, 20,* 220–225. doi:10.1037/0882-7974.20.2.220

Yechiam, E., & Busemeyer, J. (2005). Comparison of basic assumptions embedded in learning models for experience-based decision making. *Psychonomic Bulletin & Review, 12,* 387–402. doi:10.3758/BF03193783

Yechiam, E., Busemeyer, J. R., Stout, J. C., & Bechara, A. (2005). Using cognitive models to map relations between neuropsychological disorders and human decision-making deficits. *Psychological Science, 16,* 973–978. doi:10.1111/j.1467-9280.2005.01646.x

Yechiam, E., Hayden, E. P., Bodkins, M., O'Donnell, B. F., & Hetrick, W. P. (2008). Decision making in bipolar disorder: A cognitive modeling approach. *Psychiatry Research, 161,* 142–152. doi:10.1016/j.psychres.2007.07.001

Yechiam, E., Kanz, J. E., Bechara, A., Stout, J. C., Busemeyer, J. R., Altmaier, E. M., & Paulsen, J. S. (2008). Neurocognitive deficits related to poor decision making in people behind bars. *Psychonomic Bulletin & Review, 15,* 44–51. doi:10.3758/PBR.15.1.44

Received March 29, 2012
Revision received June 12, 2012
Accepted July 25, 2012 ■