The Bangor Gambling Task: Computerized replication and reappraisal of an emotion-based decision task

Rojas-Líbano, Daniel; Zúñiga, Javiera; Corrales, Vanessa; Pino, Paulina; Infante, Macarena; Turnbull, Oliver H.; Bowman, Caroline; Salas, Christian

Applied Neuropsychology: Adult

DOI:
10.1080/23279095.2022.2162403

E-pub ahead of print: 13/01/2023

Peer reviewed version

The Bangor Gambling Task: computerized replication and reappraisal of an Emotion-based decision task.

Daniel Rojas-Líbano[a], Javiera Zúñiga[a,b], Vanessa Corrales[c], Paulina Pino[c], Macarena Infante[c], Oliver Turnbull[e], Caroline Bowman[e], Christian Salas[a,c,d]*

[a]*Centro de Estudios en Neurociencia Humana y Neuropsicología, Facultad de Psicología, Universidad Diego Portales, Santiago, Chile;* [b]*Programa de Magister en Neurociencia Social, Facultad de Psicología, Universidad Diego Portales, Santiago, Chile;* [c]*Facultad de Psicología, Universidad Diego Portales, Santiago, Chile;* [d]*Unidad de Neuropsicología Clínica, Facultad de Psicología, Universidad Diego Portales, Santiago, Chile;* [e]*School of Psychology, Bangor University, Wales, UK*

Daniel Rojas-Líbano ORCID ID: 0000-0002-9517-5545

*corresponding author:

Christian Salas

christian.salas@udp.cl

Centro de Estudios en Neurociencia Humana y Neuropsicología

Facultad de Psicología

Universidad Diego Portales

Vergara 275

Santiago, Chile

WORD COUNT: 10971 (including references)

Emotion-based decision making (EBDM) is the capacity to make decisions based on prior emotional consequences of actions. Several neuropsychological tasks, using different gambling paradigms and with different levels of complexity, have been designed to assess EBDM. The Bangor Gambling Task (BGT) was created as a brief and simple card gambling-task to assess EBDM. BGT contains a single-card deck and requires participants to decide whether to gamble or not, which can result in wins or losses. Unknown to the participant, the winning probabilities decrease throughout the task (from 0.75 in the first block to 0.25 in the fifth block), requiring participants to reduce their gambling probability to avoid long-term losses. A few studies have offered evidence regarding the BGT convergent validity. However, there are no computerized versions of BGT available, thus slowing the process of gathering information to explore the EBDM mechanisms behind the task, its validity, and clinical usefulness. In this article, we present a computerized version of the BGT using the Matlab environment and make all our code available. We explore BGT's replicability and analyze its probabilistic structure, providing trial-level and block-level analyses. Eighty-one participants performed the computerized version, which followed the same structure as the original version. It took participants 8.5 ± 3.3 minutes to complete the task, which is faster than the paper version. Replicating previous studies, participants diminished their gambling probability throughout the task, learning to inhibit the initially rewarded gambling behavior. This change in gambling probability could be considered a proxy for EBDM. Our analyses suggest that the last blocks are especially sensitive to capturing deficits in EBDM, and we propose some modifications to BGT's original version to enhance the initial exploratory and learning phase. Our results show that the BGT constitutes a quick and simple task to evaluate EBDM capacities.

Keywords: Emotion-based decision making; Emotion-based learning; Bangor Gambling Task; Replications

Introduction

Emotion-based decision making (EBDM) has been defined as the capacity to make decisions based on prior experience of the emotional consequences of actions (Bowman & Turnbull, 2004). This capacity is especially relevant in complex and uncertain situations. EBDM demands learning an association between complex situations and the emotional state generated by them in terms of visceral and somatosensory markers -commonly known as "somatic markers". Based on previous experience, these markers can signal the prospective consequences of an action and assist in selecting advantageous response options (Bechara et al., 1996, 1997; Bechara & Damasio, 2005; Wright et al., 2017). It has been proposed that somatic markers would allow an "emotional hunch" or "gut feeling", which guides what traditionally has been labeled as "cognitive" decision making (Bechara et al., 1997). The conscious or unconscious nature of the information processed by the somatic marker is still a matter of debate (Maia & McClelland, 2004). Regarding the neuroanatomical basis of these mechanisms, evidence suggests that EBDM arises from large-scale systems that include cortical and subcortical structures, such as the ventromedial PFC, amygdala, somatosensory and insular cortices, and the peripheral nervous system (Poppa & Bechara, 2018).

EBDM has been considered a critical psychological process in understanding not only healthy development (Cauffman et al., 2010; S. Wood et al., 2005) but also psychiatric and neuropsychiatric disorders (Franken et al., 2008; Haaland & Landrø, 2007; Lee et al., 2007; Smoski et al., 2008), economic decision making (Bechara & Damasio, 2005) and socio-emotional changes after brain injury (Barrash et al., 2011; McDonald et al., 2013). EBDM has been particularly important in understanding personality and neurobehavioral changes after brain damage, specifically in patients with traumatic brain injury (TBI) or focal damage to the frontal lobes, who can exhibit impairments in decision making despite preserving the knowledge necessary to make a decision (Milner, 1963). These patients can exhibit typical performance on classic executive tests, which are simple, highly structured, and emotion-free (R. Wood, 2013). However, at the same time, these patients can also present aberrant social behavior, make impulsive and risky choices, or perseverate in a behavior despite its negative consequences (R. Wood, 2013). This dissociation between knowing and doing has been fundamental in the search for new tools that can capture decision making in the real world, where contingencies change, uncertainty prevails, and choices can have an emotional cost (Burgess & Stuss, 2017; Manchester et al., 2004).

Researchers and clinicians have traditionally assessed EBDM using the Iowa Gambling Task (IGT) (Bechara et al., 1994). This task was developed to measure real-world decision-making problems in individuals with ventromedial prefrontal damage who exhibited average performance on classic executive tasks. The task simulated decision-making by introducing reward, punishment, and elements of uncertainty to a gambling task. Participants could win or lose money by selecting cards from four separate decks: two ‘risky’ or ‘bad’ decks with high wins and losses and two ‘safe’ or ‘good’ decks with small wins and losses. A central element of this task is that participants must defer (or inhibit) short-term benefits for long-term profit throughout 100 trials with varying contingencies. In the long-term (i.e., across several trials), selecting from the ‘good’ decks is advantageous for the participant. In contrast, ‘bad’ decks are disadvantageous, given the long-term losses associated with them. The IGT’s estimated duration is around 20-30 minutes (Bowman & Turnbull, 2004).

Following the IGT, researchers have developed alternative tasks to measure EBDM. The structure and type of stimuli used by these tasks, as well as the level of cognitive demand required, have been heterogeneous. Some have preserved a card gambling paradigm (*Cards and Lottery Task*, (Mueller et al., 2017); *Probability Associated Gambling Task*, (Zamarian et al., 2008); *Columbia Card Task*, (Figner et al., 2009)), while others have innovated using new gambling games (*Game of Dice Task*, (Brand et al., 2005); *Cups Task*, (Weller et al., 2007) or designs that are more appealing to younger populations such as the *Balloon Analogue Risk Task* (Lejuez et al., 2002). Concerning card gambling paradigms, in general, they have become more complex in terms of the number of visible items (the *Columbia Card Task* shows a grid of 32 cards), structure (*the Cards and Lottery* combines two games), and amount of information provided about winning probability (*Probability Associated Gambling Task).* In 2004, Bowman and Turnbull developed the Bangor Gambling Task (BGT), a task with a similar structure to the IGT (financial reward and punishment, varying contingencies) but simpler and more straightforward in design. The BGT follows card-based gambling paradigms developed outside the field of neuropsychology, which are simpler because they use only one deck where contingencies are manipulated (Newman et al., 1987). This particular type of simple card gambling paradigm can be extremely helpful in assessing EBDM in populations where cognitive impairment is common and severe (stroke, TBI, brain tumors, etc).

In the BGT, participants choose whether to gamble or not with a *single* deck of cards. The decision not to gamble is inconsequential, but the gambling choice could result in monetary gains or losses. Unbeknownst to the participant, the winning probability diminishes across the five blocks that constitute the task. There are five consecutive 20-trial blocks, and each block contains a fixed winning probability that decreases step-wise. Participants are initially reinforced to bet, with a 75% chance of winning (first block), but by the end of the task (fifth and last block), this probability is only 15%. Therefore, to succeed, participants have to readjust their gambling probability to match the decreasing gains continually. The BGT structure and temporal dynamics resemble "affective shifting" paradigms that measure the capacity to adjust responses when the reinforcement value of stimuli changes (Fellows & Farah, 2003). The BGT also resembles what the literature has called the "outcome devaluation" paradigm, where a participant overcomes a previously trained action after outcome devaluation (Friedel et al., 2014).

A handful of studies have employed the BGT, offering preliminary data on its validity as a measure of EBDM. In their original study, Bowman and Turnbull (2004) provided valuable data suggesting that the IGT and BGT had a similar structure (i.e., both have 100 consecutive trials consisting of a gambling choice that results in gains or losses) as well as good concurrent validity. In the IGT, they measured performance as the following score: the number of good-deck gamble trials minus the number of bad-deck gamble trials computed over each of the 20-trial blocks. Analogously, the BGT score corresponded to the number of gamble trials ('bad' choices) minus the number of no-gamble ('good' choices), also over each of the five 20-trial blocks. Therefore, an adequate EBDM process in IGT and BGT should manifest as an *increase* in these scores across the trial blocks. Comparing the IGT and BGT in an undergraduate student sample (n = 40), the authors showed a similar type of incremental learning in both tasks and a positive correlation of large size in performance scores (r^2 = .93; $p < 0.001$). In a more recent study, performance on the BGT was compared between individuals with TBI (n = 30) and controls (n = 39) (Adlam et al., 2017). Here, the authors found that across the task, survivors of TBI made more gambling choices than controls, resulting in different BGT scores between groups. They found that as a group, the TBI patients could change their initial gambling response across the task, although displaying a large group variability. Preliminary cluster analysis suggested different patient groups regarding BGT scores (Adlam et al., 2017). This finding is indicative that the BGT could potentially be used to assess the different levels of remnant EBDM capacity in TBI survivors.

Recently, Heninga and collaborators explored the relationship between BGT and IGT in a prospective study with 176 adolescents (Heininga et al., 2019). In contrast to the Bowman and Turnbull study (2004), where both tasks were administered within 10 minutes apart, here, the BGT and IGT were administered at two different time points separated by 3 years. The authors reported a similar incremental pattern in both tasks but no correlation between overall scores (Heininga et al., 2019). It was suggested that differences in the design of both studies, particularly about the time of data collection, could account for such discrepancies. The authors highlighted the need for future studies to replicate these findings and explore the test-retest reliability of both tasks in different populations.

Despite this promising data and the potential advantages of the BGT in terms of time administration and task simplicity, there is little research analyzing the structure and properties of the task. This has important implications for the task's potential broader applicability, particularly in construct validity, reliability, and learning effects. Finally, in a context where empirical results from psychology have been questioned (Earp & Trafimow, 2015; Goodman et al., 2016; Simmons et al., 2011; Yong, 2012), the replicability of findings is essential to have a good understanding of the scope and validity of our generalizations. In Clinical Neuropsychology, this is particularly important since patient diagnosis and care should stem from the best available evidence (Gelman & Geurts, 2017). Most research reports suggest that EBDM problems can emerge at different moments of the lifespan (Smith et al., 2012), and EBDM impairments can occur in a wide range of clinical populations, such as brain injury (Bechara, 2004), dementia (Darby & Dickerson, 2017), schizophrenia, (Shurman et al., 2005), personality disorder (Haaland & Landrø, 2007), substance dependence (Whitlow et al., 2004), and pathological gambling (Brevers et al., 2013). Therefore, it is necessary to have a variety of well-validated assessment tools.

In this study, we present the development of a computerized version of the BGT and the replication of the original task. A computerized version is helpful in several ways. It facilitates data generation to replicate the task and allows exploring its construct validity and clinical usefulness across different patient groups and cultural contexts. Computerized versions of classic neuropsychological tasks present several strengths, such as an increase in ease and standardization of administration (Fillit et al., 2008), a reduction in errors during scoring and interpretation (Mataix-Cols & Bartrés-Faz, 2002), and less time used in the preparation of materials (Koski et al., 2011). Based on the above-mentioned considerations, the main goals of this study were: 1) to develop a computerized version of the original BGT; 2) to compare results from the computerized and previously published non-computerized BGT versions in terms of replicability and reliability; and 3) to analyze the evolution of EBDM through the task, as a means to reassess the task's probabilistic structure and properties.

Materials and Methods

Participants and Procedure

We recruited eighty-one participants (41 female; Age Median = 26; Age Mean = 32.9; SD = 14.52 years old). The data collection process occurred in two different periods.

The first occurred between December 2018 and September 2019 (n = 39 participants). The second occurred between April and May 2022 (n = 42 participants). Participant recruitment happened through researchers' social networks and printed posts in the University's Psychology Department. Our inclusion criteria were: To be older than 18 years old and to have completed secondary education. The exclusion criteria were: Refusal to sign the written consent and a diagnosis of a neurological condition. The study was reviewed and approved by the institutional Ethics Committee at Universidad Diego Portales, Facultad de Psicología. All participants signed a written consent form for participation. Participants performed the task individually in a sound-proof, dimly lit experimental room. We used a computer screen View Sonic XG2402, with a spatial resolution of 1920 x 1080 pixels and dimensions of 53.4 cm (width) and 30.1 cm (height) to present the images throughout the task.

Task

The entire task structure, including winning probabilities for each block, the sequence of cards presented, and the set of instructions given to the participants, was the same as the one used in the original task version (Bowman & Turnbull, 2004). In the original version, a deck of 100 regular playing cards, consisting of 38 high cards (Jack, King, Queen, or Ace) and 62 low cards (cards between 2 and 10), were sequenced to create a pattern of winning and losing streaks. We followed exactly this organization to present the cards on the computer screen. The complete predetermined sequence of cards is presented in Table 1. The predetermined sequence of wins and losses associated with a "yes" decision is presented in Table 1 and Figure 1B. We used the protocol from the original version to write the computer code that ran the task, controlled/monitored the hardware, and acquired behavioral data during the task. We wrote all hardware control and data acquisition routines in Matlab, using the Psychophysics Toolbox extension

(Brainard, 1997). All the Matlab code written to run the task is available for download at https://github.com/dirl75/BGT. The task has a simple structure, and the code contains several comments, therefore it can easily be translated into any programming language. Most task parameters, such as the number of trials, inter-trial interval, wins/losses sequence, cards presented, amount of money, and the like, are modifiable task parameters in the code.

Experimental Procedures

During the task, the participant sat in front of the screen at a distance of 60 cm from it. The first three screens contained the following written instructions:

- Your goal is to make as much money as possible
- You have $20000 to gamble.
- This is not a regular deck of cards.
- Every card will either win or lose some money.
- A ‘face’ card will always be a winning card, and a ‘number’ card will always be a losing card.
- It is up to you whether to gamble on the card before you turn it over.
- If you decide to gamble, you will either win or lose the amount stated on the card.
- If you decide not to gamble, you will neither win nor lose the amount stated on the card.
- You may gamble as often or as little as you like.
- Some cards are worse than others, but you will make money if you gamble wisely.

We told participants that the amount of money was a virtual one and they should imagine it corresponded to the national currency. Participants did not receive actual cash for participation in the study. As an incentive, we offered participants that they could enter a lottery to win one of three gift cards (prepaid stored-value money cards issued by a local retailer). The participants were aware that the lottery was not related to task performance. Nevertheless, participants' performance showed that most of them understood the task and took decisions accordingly, learning to diminish losses. After the participant read the instructions and decided he or she was ready, the task began. On each trial, a deck's image appeared on the screen, and the participant had to choose whether to gamble by pressing one out of two keys on a computer keyboard. There were no imposed time constraints on trial completion, and participants could take as long as they wanted to finish the task. After the keypress, the image changed, revealing the trial's card, the gamble outcome, and the new amount of total money, if applicable (see Fig. 1A). The card showed up on the screen regardless of choice during a fixed interval of 2.5 seconds. If the participant chose not to gamble, no change in money occurred. If the participant decided to gamble, it could either win or lose depending on the card type ('High Cards', i.e., J, Q, K, A, meant winning, and 'Low Cards', i.e., the rest, meant losing). For a detailed description of the number of wins and losses, see Fig. 1B and Table 1. There were no constraints to the gambling behavior of participants. For instance, they could choose not to gamble across the entire task if they decided to do so.

The task ended after the participant had completed 100 trials. The schedule of gains and losses assigned to each trial (which applied if the participant chose to gamble) was predefined and was the same across participants (see Fig. 1B). The structure of gains and losses was such that the probability of winning if gambling was a decreasing function of the five 20-trial blocks of the task. These win probabilities were: Block 1, 0.75; Block 2, 0.5; Blocks 3 and 4, 0.25; Block 5, 0.15 (Fig. 1C). Participants had no way of telling when a block change took place since block information was neither communicated during task induction nor displayed on the screen. The overall probability of winning throughout the 100 trials of the task if gambling in every single trial was P(win) = (1/5) · (0.75+0.5+0.25+0.25+0.15) = 0.38.

We also analyzed the evolution of the amount of money across the task. Albeit indirectly, total money earned is related to gambling probability. It is, therefore, a variable worth exploring to assess its potential as a different data proxy of the EBDM process. Throughout the article, when we refer to money amounts, we use either thousands (to avoid many-digit figures) or a percentage of the initial amount. If gambling in every single trial, the expected amount of money to get from the task (i.e., the sum of all gains and losses, plus the initial amount) was (in thousands) -$30, or a decrease of 250% relative to the initial amount ($20). As shown in Figure 1B, possible winning and losing amounts were of two types (in thousands): +/- $1 or +/- $0.5.

The overall change in the amount of money across the task, $\Delta Money$, was calculated as the amount of money accrued by the last trial minus the amount of money in the first trial.

Statistics and Data Analysis

We performed all data analysis offline using Matlab (version R2014b) and wrote custom code to import, store, and analyze data. We implemented all the statistical comparisons with participants as the unit of analysis.

a) Calculation of the BGT score

The BGT score is the metric of EBDM developed in the original study (Bowman & Turnbull, 2004) that quantifies task performance. We calculated it as the number of trials in which the participant decided not to gamble N_i, minus the number of trials in which she/he decided to gamble Y_i, over each of the twenty trials of a given block.

Thus, for block i:

$(BGT\ score)_i = N_i - Y_i$

Since the block contains twenty trials, and the only possible responses are yes or no, we have $N_i = 20 - Y_i$. Therefore:

$(BGT\ score)_i = (20 - Y_i) - Y_i$

$(BGT\ score)_i = 20 - 2 \cdot Y_i$

b) Calculation of change in Gambling Probability

The gambling probability was calculated per block, as the fraction of trials within block i in which the participant decided to gamble (i.e., pressed the button "yes"). Thus,

$P(Gambling)_i = \frac{Y_i}{20}$

Consequently, the overall *change* in gambling probability elicited by the task was the difference in gambling probability between last and first blocks:

$\Delta P(Gambling) = P(Gambling)_5 - P(Gambling)_1$

$\Delta P(Gambling) = \frac{Y_5 - Y_1}{20}$

c. Replication metric

To assess the extent to which our data replicated previous results, we calculated the effect size between ours and previously reported data. We used data from three previous studies (Adlam et al., 2017; Bowman & Turnbull, 2004; Heininga et al., 2019). For each 20-trial block of the task, we calculated Hedges' g -the difference between the means of two groups, divided by the pooled standard deviation to estimate the effect size. The formula is as follows: $g = \frac{mBGT_{this} - mBGT_{previous}}{s_p}$

where $mBGT_{previous}$ and $mBGT_{this}$ indicate BGT score means from 'previous study' and 'this study', respectively, and s_p indicates the pooled standard deviation. Therefore, a small effect size, could be considered indicative of successful replication. Given the nature of the metric, where differences are expressed as a fraction of the standard deviation, we estimate that values below 0.5 (i.e., the difference is equal to half of the variability) are acceptable as replicates.

d. Reliability estimate

We were not able to obtain raw data from previously published implementations of the task, in order to assess reliability. Therefore, to obtain an estimate of the reliability of our results, we implemented an alternative resampling method within our data, similar to previously published procedures to assess reliability of cognitive tasks (Pronk et al., 2022; Williams & Kaufmann, 2012). We partitioned the dataset containing all participants in two randomly selected groups ($n = 40$ each) and calculated the correlation coefficient of the mean BGT score across blocks between the two groups. This correlation coefficient constitutes an estimate of reliability since it compares two groups of subjects that performed the task at different times, days, and, in some cases, years. We then repeated this procedure 20,000 times, to draw a large number of possible data partitions, collecting the correlation coefficient from each resample. We then built a distribution of the correlation coefficient values thus obtained.

e. Null model of the task

To explore the task's probabilistic structure and properties, we compared the participants' performance with a model of chance behavior. We constructed a null (random) model of the task responses. To do so, we simulated a process consisting of a random choice, implemented through Matlab's function *randi*, independently for each of the 100 trials, with both options, gamble, and not-gamble, being equiprobable (probability = 0.5). We ran 10^4 simulations (i.e., 'trials') of this process, saving from each trial the simulated choice and the amount of money obtained (taken from Table 1), given the simulated choice. The distributions of money and gambling across simulations constituted our null models, i.e., the expected gambling or money sequences if the process generating the choices was a random one.

f. Statistical tests

To compare the money gained or lost and the probability of gambling between the actual data and the null model, we implemented nonparametric, cluster-corrected statistical comparisons (Maris & Oostenveld, 2007). For every sample, i.e., each (money/choice, trial) value, we computed the t statistic, obtaining the non-permuted t values. We then implemented a permutation, i.e., a random assignment of the data to each of the two groups, and re-calculated t-values. We repeated the permutation procedure 10^4 times, obtaining a distribution of *t-values* for each trial. We then selected all the distribution samples with non-permuted t values corresponding to $p < 0.05$. As the last step, we corrected the t-values by clustering the selected samples by adjacency in the trial axis and calculated cluster-level statistics by taking the sum of the *t-values* within a cluster.

We computed the time course of the net amount of money kept by participants on a per-trial basis. We expressed the raw amount of money as a percentage of the initial quantity, which was the same for all participants. As described above, we used a null model as a comparison. We simulated a random process in which each choice (i.e., gamble or no-gamble) was equiprobable at any given trial. As an additional comparison, we computed the money per trial that would result from gambling in every task trial. Concerning gambling probability, we estimated the probability as a binomial fit of the across-participants data on any given trial using the Matlab function *binofit*.

Results

Descriptive Results. All participants completed the 100 trials constituting the BGT. Participants finished the task in 8.5 ± 3.3 (M ± SD) minutes (Fig. 1D). To estimate how long they took to decide, we measured, for each trial, the time elapsed between deck presentation to the moment of the button press signaling the choice. This decision time was 0.43 ± 2.2 seconds (Fig. 1D). Even when there were quite long decision times (e.g., the maximum value was 41.9 seconds), 95% of the data was below 3 seconds (Fig. 1D).

Given that a previous report had found some age dependence on the BGT score (Adlam et al., 2017), we ran a linear model using our demographic variables, sex and age, as predictors of mean (across-blocks) BGT scores and found that only age was a significant predictor (t = -3.69, p < 0.01), albeit with a relatively low explained variance (R^2 = 0.28). The coefficient value (β = -0.19) implied that overall BGT scores slightly diminish as age increases in our sample. Given that younger ages were much more represented than older ones in our sample, it is difficult to draw a firm conclusion from these results, but it may reflect a tendency to find lower BGT scores at older ages. However, when we used the across-task *change* in gambling probability ΔP (see methods), none of the predictors were significant, with a poorly explained variance (R^2 = 0.1). Given that ΔP quantifies behavioral change across the task, and therefore the adequacy of the participant's decisions, ΔP is more indicative of EBDM than the overall BGT score. Thus, we found no evidence of a relationship between age and EBDM in our sample.

Descriptive data from participants' gambling behavior in each block suggested that, as the blocks progressed, and the winning probability diminished, participants learned to inhibit their initial tendency to gamble. Participants varied widely in their strategies at the beginning of the task. However, by the end, they converged around a 'no-gamble' approach. As in previous studies, we calculated the BGT score as the number of 'no' (no gamble) responses minus the number of 'yes' (gamble) responses in a block. In the first block, participants started with a BGT score closer to zero (Mean = -1.64; Median = 0; SD = 8.08; Min = -20; Max = 16), indicating no preference over a single choice type, accompanied by a relatively large variability. It is worth noting that in the first block, 50 participants (61.7%) displayed negative BGT scores, indicative of an initial preference for gambling. In contrast, by the end of the task (fifth block), the average score was positive, and variability had reduced considerably (Mean = 11.6 ; Median = 10; SD = 6.27; Min = -2; Max = 20) (See Fig. 2A-B). Here only 8 participants (9.8%) displayed negative BGT scores. We ran a repeated-measures ANOVA to evaluate the change in BGT score across blocks, finding a significant effect of block ($F(4,320)=87.68$, $p < 0.001$) and an effect size of $\eta^2 = 0.61$. We also performed a paired t-test to compare BGT scores between the first and last blocks, obtaining $t(80)= - 12.06$, $p < 0.001$.

These results showed a decrease in the tendency to gamble across the task. However, we found no relationship between first- and last-block BGT scores (r = 0.13, p = 0.26; n = 81), which indicates that the learning occurred independently of initial scores.

In addition, we noted that, across the task blocks, the largest mean difference D in BGT scores, corresponding to an increase indicative of an EBDM adequate to task changes, occurred between blocks 3 and 4 ($D_{2\text{-}1}$ = 0.86; $D_{3\text{-}2}$ = 2.1; $D_{4\text{-}3}$ = 7.48; $D_{5\text{-}4}$ = 2.64; see Figure 2A).

Reliability: One of our estimates of task reliability was a resampling procedure (see Methods) that produced a distribution of Pearson's correlation coefficient between across-block BGT scores. The distribution obtained had the following parameters: Mean = 0.97; Median = 0.98; SD = 0.02; 5th Percentile = 0.96; 95th Percentile = 0.99. Therefore, and given that we ran sessions at different periods between 2018 and 2022 (see "Participants and Procedure"), we conclude that our task's dataset has high reliability.

Replication: Concerning replication, we compared the BGT scores with three previous reports (Adlam et al., 2017; Bowman & Turnbull, 2004; Heininga et al., 2019). In general, we found a close agreement between our results and previous studies, thus suggesting a successful replication and adaptation from the pencil and paper to the computerized version. Figure 2C shows the across-participants mean BGT scores, per block, alongside data from previous publications, with similar time courses across studies. The difference in BGT score between the last and first blocks was also similar (This study: 13.3; Bowman et al. 2004: 13; Adlam et al. 2017: 9.2; Heininga et al. 2019:

10.9. We will refer to these studies as B04, A17 and H09, respectively) (see Fig. 2C). To quantify the similarity between ours and previous versions (replication's success), we used a measure of effect size. Our estimate was Hedges' g, a standardized mean difference between two conditions. In this case, the 'effect' is constituted by the change in time, space, and task modality (i.e., computerized) given by our implementation of the task in a new context. A successful replication, then, would be indicated by small effect sizes. The entire set of g values per block and each of the three previous publications are presented in Table 2. The overall (across-blocks) mean values were: for B04, $g = 0.35 \pm 0.18$; for A17, $g = 0.39 \pm 0.18$, and for H19, $g = 0.50 \pm 0.25$. When computed per block, the maximum value $g = 0.71$ (H19, Block 5) and the minimum value is $g = 0.06$ (B04, Block 1). Consistently, the largest values of g, indicative of larger differences, were associated with H19, which was also the only study that used adolescents as subjects.

Probabilistic structure and properties of the task*: *Money amounts. Our results show that, for most task blocks, participants' amount of money resembles a person gambling randomly. That is to say, the time course of money changes seems to reflect a sequence of random choices. A statistically significant divergence between the data and the null model only emerges around the mid-fourth block (Fig. 3A). In other words, only during the last 25% of trials participants' amounts of money was different from what would be expected by chance alone. Interestingly, around the same time, both the participants' data and the null model started to differ from the money time series that would have been obtained had one decided to gamble on every single trial (Fig. 3A). Regarding actual amounts, at the end of the first 20 trials, participants displayed little change compared to the first trial regarding the amount of money (Mean = 111.46%; SD =

6.8%; Min = 97.4%; Max = 126.7% of the initial amount). In contrast, at the end of the task, there was an overall decrease but also a much larger dispersion of the data (Mean = 43%; SD = 42.77%; Min = -72.6; Max = 133.3% of the initial amount).

Probabilistic structure and properties of the task*: *Gambling probability. As for gambling probability, it started with wide variability among participants (centered at 0.5 in the first block) and progressively decreased, finishing close to 0.25 in the last block. A first result from our analysis is that the gambling probability time series was adequately fitted by a linear model using the trial number as a predictor (Beta = -0.004 probability unit/trial; $R^2 = 0.59$; $F(1,98) = 145.8$; p-value <0.0001). This analysis shows that gambling probability is thus a decreasing function of the trial number (see Fig 4A), which suggests that participants learn to inhibit their initial gambling behavior throughout the task. This is consistent with the diminishing winning probability imposed by the task structure. Importantly, our statistical comparisons show that participants' probability of gambling only differed statistically from the null model in the last 25% of trials (Figure 4A), similar to the case of the amount of money retained per trial. When analyzed by looking at the participants' mean gambling probability in the first and last blocks, we found an initial wide distribution, clustering roughly around equiprobability, that evolved to end much narrower and centered around a lower value (Block1: Mean = 0.58; SD = 0.18; Min = 0.1; Max = 1. Block 5: Mean = 0.25; SD = 0.18; Min = 0; Max = 0.9).

Given that learning in this task corresponds to a change (reduction) in the decision to gamble, we also looked at the overall *change* in gambling probability throughout the task (see Figure 4B). As a metric, we computed the difference ΔP in gambling probability between the last and first blocks (see Methods). Therefore, we expect that learning will show up as a negative difference ($\Delta P < 0$), indicating the above mentioned reduction in gambling probability. For ΔP, we found: Mean = - 0.33; Median = - 0.35; SD = 0.24. In our sample, 9 participants (11%) did not learn (that is, displayed a $\Delta P \geq 0$). The distribution of ΔP values from our sample is shown in figure 5A. In this distribution, 95% of the data lies between -0.72 and 0.15. There is an overlap with the null model, whose 95% runs between -0.3 y 0.3.

When we analyzed the relationship between money across the task $\Delta Money$ and ΔP (see Figure 5B), we found an association (r = -0.45, $p < 0.01$, n = 81). Therefore, there is potentially relevant and additional information in the money variable. For example, for a specific ΔP value of -0.4, indicative of EBDM, there are four participants. Their $\Delta Money$ values vary between -85.7% (large loss) to 0% (no loss).

Discussion

As described in the introduction, during the last decade, EBDM has increasingly become a relevant psychological process to understand development, mental health, and the neural basis of human decision making. Such interest requires the availability of a wide range of tools to assess EBDM in normal and patient populations. The BGT was introduced in 2004 as a simple task to evaluate EBDM (Bowman & Turnbull, 2004) and had also been used to evaluate brain injury patients in 2017 (Adlam et al., 2017). However, we lack additional replications and in-depth analyses of its structure and properties. This study reports the successful development of a computerized version of the BGT by replicating previous findings in a Spanish-speaking sample of adult participants. These results reflect the task's robustness since it offers similar results in different application formats and when testing individuals from different cultural backgrounds. Furthermore, using a novel approach to analyze the task data (trial by trial), we could also explore the task structure and properties. The main finding here was that the last two blocks of the task appear highly informative regarding gambling behavior inhibition, which is our estimate of EBDM.

Given that this is the first computerized replication, we decided to implement the same task (sequence of cards, wins, and losses) as the original version (Bowman & Turnbull, 2004). However, in terms of time, our participants took less time (8.5 ± 3.3 mins), compared to the Bowman and Turnbull study (15-20 mins). Most likely, this difference is due to our computerized implementation, whose trial structure had an intrinsic pace due to a fixed 2.5 seconds period of feedback after gambling, after which the next trial started.

There appears to be some gain in analyzing the BGT by looking at its blocks separately and not averaging performance across the entire task. Such an approach, never reported before, offers valuable information to understand the contribution of each block to the task's main goal, as well as the potential usefulness of modifying the task structure to enhance its capacity to measure EBDM. In the BGT, this is particularly important concerning the initial block. Theoretically speaking, this block is relevant since it is supposed to reinforce gambling behavior based on the high probability of winning (.75). However, our data analysis shows that this may not be the case since there is a large variability in gambling probability amongst participants during this block (see Fig. 4). We draw similar conclusions regarding the last block, which is the only one where participants show a significant difference from the null model. This block theoretically assesses individuals' capacity to inhibit gambling behavior based on learning from the progressive decrease in winning probabilities. Our data suggest that, compared to other blocks in the task, this one could be particularly sensitive to capture difficulties in the capacity to withhold a previously learned emotional response or inhibit gambling behavior when consequences are emotionally negative. It is important to note that, besides the increase in losing probability, the task structure contains much *larger* losses in the second half of the task than in the first (see Figure 1B). Therefore, we expect that these larger negative stimuli should promote faster EBDM in the second half. We suggest that clinical neuropsychologists should pay particular attention to the performance of patients during these two last blocks since it may reflect EBDM deficits related to inhibition.

Evidence from the IGT supports our findings and stresses the need to analyze different BGT stages separately. For example, studies exploring the relationship between the IGT and risk-taking have also suggested that the early and later stages of the task should be considered independently (Brand et al., 2007) and that the propensity for risk-taking appears to be more related to the last section of the task (Brand et al., 2005). Other authors have interpreted these findings as suggesting that the early stages of the IGT cannot offer information regarding risk-taking since there is no explicit knowledge regarding the rules that organize the task (Upton et al., 2011). This is also consistent with authors describing that gambling behavior during the early (20-40) trials of the IGT must be considered as “exploratory behavior” (Dunn et al., 2006). On the contrary, in the later stages of the task, because participants have developed a more explicit knowledge of the risk profile of alternatives, it would only be possible to observe participants’ propensity for risk-taking. Brand and colleagues (2005) have conceptualized this difference by suggesting that the IGT taps into two mechanisms of decision making: decision under ambiguity (early stages) and decision under risk (later stages). It is sensible to propose, considering our data and reports from the IGT, that the BGT also presents a structure that allows measuring both types of decision making.

There are noticeable structural differences between the IGT and the BGT. Perhaps the most important here is that probabilities are not spatially allocated to specific decks in the BGT but dynamically change over time. However, it is likely that in both tasks, the initial gambling trials can be considered exploratory behavior, thus not reflecting risk-taking propensity or inhibition ability/impairment. Our data shows a roughly 50% probability of gambling in the BGT first block, although there is a .75 probability of winning. This block's exploratory nature may pose a limitation to the goal of the task itself, which is first to reinforce gambling behavior and only later demand from participants to inhibit gambling when probabilities turn disadvantageous. In other words, it may be that the block with the higher probability of reward overlaps with the block where exploratory behavior takes place, thus compromising the consolidation of emotional learning. Stout and colleagues have commented about the IGT that learning from experience should be considered when interpreting findings, particularly considering that some populations (e.g. brain injury, neurodegenerative disorders) may present difficulties in learning the task contingencies and the associated risks (Stout et al., 2005). An indicator that participants have learnt the probabilities is that their gambling behavior 'probability matches' the task structure. In other words, learners should respond with the same probability that they are obtaining gains (Knowlton et al., 1994). For example, participants of the BGT should decide to gamble 75% of the time in the first block. We did not observe that probability match in our sample. As mentioned previously, in the first block, participants seemed committed to more exploratory behavior and were not thoroughly familiarized with the task yet. Therefore, future studies should explore how changes to the BGT structure can enhance the learning process considering this probability match, for example, by adding an extra block with high winning probabilities (0.75) at the beginning of the task.

Now that we have offered evidence that the computerized version of the BGT captures a similar process to the paper and pencil version, future studies should explore its reliability in time (test-retest) as well as its association with other measures commonly related to EBDM, such as impulsivity or risk taking -BIS/BAS, Balloon Analogue Risk Taking- (concurrent validity). It would also be interesting to study the association between task performance and physiological response (e.g. skin conductance, heart rate variability), thus exploring the potential usefulness of the BGT to offer data that can contribute to the Somatic Marker Hypothesis Model. Following the steps of the IGT, it will be important to explore the multiple factors that can account for the performance of subjects in the BGT. We have learnt from the IGT that there are both neuropsychological and personality variables that relate to individuals' performance in EBDM tasks. Future studies should consider this literature and explore how traits such as impulsiveness, sensation seeking (Buelow & Suhr, 2009; Crone et al., 2003; Davis et al., 2007; Franken et al., 2008), negative mood (Miu et al., 2008; Must et al., 2007; Suhr & Tsanadis, 2007) and executive ability (Gansler et al., 2011; Toplak et al., 2010) relate with participants' performance during the exploratory/learning (initial blocks) and inhibiting phases of the task (last blocks). A computerized version such as the one presented here will facilitate both the collection of large sets of data and the analysis of the task structure and its relationship with other variables of interest. Large sets of population data will make possible the development of normative data, thus allowing clinicians to determine whether the performance of an individual falls within the normal range. Considering our preliminary findings, the absence of a reduction in gambling probability throughout the task or an increase in gambling probability could be informative of an EBDM impairment.

The potential of the task BGT is necessary to highlight. The BGT displays a simple structure (a single deck and only two choices) and a short application time, all while allowing to study the EBDM process of the participant. The computerized version of the BGT also offers the flexibility to be employed in different settings since it runs on a laptop or tablet. This makes the BGT a quick and versatile task that can be employed in clinical and research settings. Future studies should address the sensitivity of the task in capturing EMBD difficulties in specific clinical populations.

Acknowledgments

We thank Waldemar Ancamilla for assistance in data collection during earlier versions of the task. We thank Romina Kuppe for assistance with Figure 1.

Funding details

We acknowledge funding by Proyecto Fondecyt 11190604 (Daniel Rojas-Líbano) and Proyecto Fondecyt 11170508 (Christian Salas).

Disclosure statement

The authors report no potential competing interest.

References

Adlam, A.-L. R., Adams, M., Turnbull, O., Yeates, G., & Gracey, F. (2017). The Bangor Gambling Task: Characterising the Performance of Survivors of Traumatic Brain Injury. *Brain Impairment: A Multidisciplinary Journal of the Australian Society for the Study of Brain Impairment*, *18*(1), 62–73.

Barrash, J., Asp, E., Markon, K., Manzel, K., Anderson, S. W., & Tranel, D. (2011). Dimensions of personality disturbance after focal brain damage: investigation with the Iowa Scales of

Personality Change. *Journal of Clinical and Experimental Neuropsychology*, *33*(8), 833–852.

Bechara, A. (2004). The role of emotion in decision-making: evidence from neurological patients with orbitofrontal damage. *Brain and Cognition*, *55*(1), 30–40.

Bechara, A., & Damasio, A. R. (2005). The somatic marker hypothesis: A neural theory of economic decision. *Games and Economic Behavior*, *52*(2), 336–372.

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

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, *275*(5304), 1293–1295.

Bechara, A., Tranel, D., Damasio, H., & Damasio, A. R. (1996). Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex. *Cerebral Cortex* , *6*(2), 215–225.

Bowman, C. H., & Turnbull, O. H. (2004). Emotion-based learning on a simplified card game: the Iowa and Bangor Gambling Tasks. *Brain and Cognition*, *55*(2), 277–282.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.

Brand, M., Kalbe, E., Labudda, K., Fujiwara, E., Kessler, J., & Markowitsch, H. J. (2005). Decision-making impairments in patients with pathological gambling. *Psychiatry Research*, *133*(1), 91–99.

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*(1), 86–99.

Brevers, D., Bechara, A., Cleeremans, A., & Noël, X. (2013). Iowa Gambling Task (IGT): twenty years after - gambling disorder and IGT. *Frontiers in Psychology*, *4*, 665.

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

Burgess, P. W., & Stuss, D. T. (2017). Fifty Years of Prefrontal Cortex Research: Impact on

Assessment. *Journal of the International Neuropsychological Society: JINS*, *23*(9-10), 755–767.

Cauffman, E., Shulman, E. P., Steinberg, L., Claus, E., Banich, M. T., Graham, S., & Woolard, J. (2010). Age differences in affective decision making as indexed by performance on the Iowa Gambling Task. *Developmental Psychology*, *46*(1), 193–207.

Crone, E. A., Vendel, I., & van der Molen, M. W. (2003). Decision-making in disinhibited adolescents and adults: insensitivity to future consequences or driven by immediate reward? *Personality and Individual Differences*, *35*(7), 1625–1641.

Darby, R. R., & Dickerson, B. C. (2017). Dementia, Decision Making, and Capacity. *Harvard Review of Psychiatry*, *25*(6), 270–278.

Davis, C., Patte, K., Tweed, S., & Curtis, C. (2007). Personality traits associated with decision-making deficits. *Personality and Individual Differences*, *42*(2), 279–290.

Dunn, B. D., Dalgleish, T., & Lawrence, A. D. (2006). The somatic marker hypothesis: a critical evaluation. *Neuroscience and Biobehavioral Reviews*, *30*(2), 239–271.

Earp, B. D., & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in Psychology*, *6*, 621.

Fellows, L. K., & Farah, M. J. (2003). Ventromedial frontal cortex mediates affective shifting in humans: evidence from a reversal learning paradigm. *Brain: A Journal of Neurology*, *126*(Pt 8), 1830–1837.

Figner, B., Mackinlay, R. J., Wilkening, F., & Weber, E. U. (2009). Affective and deliberative processes in risky choice: age differences in risk taking in the Columbia Card Task. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*(3), 709–730.

Fillit, H. M., Simon, E. S., Doniger, G. M., & Cummings, J. L. (2008). Practicality of a computerized system for cognitive assessment in the elderly. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *4*(1), 14–21.

Franken, I. H. A., van Strien, J. W., Nijs, I., & Muris, P. (2008). Impulsivity is associated with behavioral decision-making deficits. *Psychiatry Research*, *158*(2), 155–163.

Friedel, E., Koch, S. P., Wendt, J., Heinz, A., Deserno, L., & Schlagenhauf, F. (2014).

Devaluation and sequential decisions: linking goal-directed and model-based behavior. *Frontiers in Human Neuroscience*, *8*, 587.

Gansler, D. A., Jerram, M. W., Vannorsdall, T. D., & Schretlen, D. J. (2011). Does the iowa gambling task measure executive function? *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *26*(8), 706–717.

Gelman, A., & Geurts, H. M. (2017). [Formula: see text]The statistical crisis in science: how is it relevant to clinical neuropsychology? *The Clinical Neuropsychologist*, *31*(6-7), 1000–1014.

Goodman, S. N., Fanelli, D., & Ioannidis, J. P. A. (2016). What does research reproducibility mean? *Science Translational Medicine*, *8*(341), 341ps12.

Haaland, V. Ø., & Landrø, N. I. (2007). Decision making as measured with the Iowa Gambling Task in patients with borderline personality disorder. *Journal of the International Neuropsychological Society: JINS*, *13*(4), 699–703.

Heininga, V. E., Nederhof, E., van Roekel, E., Laceulle, O., & Oldehinkel, A. (2019). Emotion-Based Learning by Iowa and Bangor Gambling Tasks: Similar but not the same? In *PsyArXiv*. https://doi.org/10.31234/osf.io/3nkfj

Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning & Memory* , *1*(2), 106–120.

Koski, L., Brouillette, M.-J., Lalonde, R., Hello, B., Wong, E., Tsuchida, A., & Fellows, L. K. (2011). Computerized testing augments pencil-and-paper tasks in measuring HIV-associated mild cognitive impairment. *HIV Medicine*, *12*(8), 472–480.

Lee, Y., Kim, Y.-T., Seo, E., Park, O., Jeong, S.-H., Kim, S. H., & Lee, S.-J. (2007). Dissociation of emotional decision-making from cognitive decision-making in chronic schizophrenia. *Psychiatry Research*, *152*(2-3), 113–120.

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., Strong, D. R., & Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology. Applied*, *8*(2), 75–84.

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*, *101*(45), 16075–16080.

Manchester, D., Priestley, N., & Jackson, H. (2004). The assessment of executive functions: coming out of the office. *Brain Injury: [BI]*, *18*(11), 1067–1081.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190.

Mataix-Cols, D., & Bartrés-Faz, D. (2002). Is the use of the wooden and computerized versions of the Tower of Hanoi puzzle equivalent? *Applied Neuropsychology*, *9*(2), 117–120.

McDonald, S., Togher, L., & Code, C. (Eds.). (2013). *Social and Communication Disorders Following Traumatic Brain Injury* (2nd ed.). Psychology Press.

Milner, B. (1963). Effects of Different Brain Lesions on Card Sorting: The Role of the Frontal Lobes. *Archives of Neurology*, *9*(1), 90–100.

Miu, A. C., Heilman, R. M., & Houser, D. (2008). Anxiety impairs decision-making: psychophysiological evidence from an Iowa Gambling Task. *Biological Psychology*, *77*(3), 353–358.

Mueller, S. M., Schiebener, J., Stöckigt, G., & Brand, M. (2017). Short- and long-term consequences in decision-making under risk: immediate feedback about long-term prospects benefits people tending to impulsive processing. *Journal of Cognitive Psychology* , *29*(2), 217–239.

Must, A., Juhász, A., Rimanóczy, A., Szabó, Z., Kéri, S., & Janka, Z. (2007). Major depressive disorder, serotonin transporter, and personality traits: why patients use suboptimal decision-making strategies? *Journal of Affective Disorders*, *103*(1-3), 273–276.

Newman, J. P., Patterson, C. M., & Kosson, D. S. (1987). Response perseveration in psychopaths. *Journal of Abnormal Psychology*, *96*(2), 145–148.

Poppa, T., & Bechara, A. (2018). The somatic marker hypothesis: revisiting the role of the "body-loop" in decision-making. *Current Opinion in Behavioral Sciences*, *19*, 61–66.

Pronk, T., Molenaar, D., Wiers, R. W., & Murre, J. (2022). Methods to split cognitive task data for estimating split-half reliability: A comprehensive review and systematic assessment.

Psychonomic Bulletin & Review, *29*(1), 44–54.

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*(2-3), 215–224.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, *22*(11), 1359–1366.

Smith, D. G., Xiao, L., & Bechara, A. (2012). Decision making in children and adolescents: impaired Iowa Gambling Task performance in early adolescence. *Developmental Psychology*, *48*(4), 1180–1187.

Smoski, M. J., Lynch, T. R., Rosenthal, M. Z., Cheavens, J. S., Chapman, A. L., & Krishnan, R. R. (2008). Decision-making and risk aversion among depressive adults. *Journal of Behavior Therapy and Experimental Psychiatry*, *39*(4), 567–576.

Stout, J. C., Rock, S. L., Campbell, M. C., Busemeyer, J. R., & Finn, P. R. (2005). Psychological processes underlying risky decisions in drug abusers. *Psychology of Addictive Behaviors: Journal of the Society of Psychologists in Addictive Behaviors*, *19*(2), 148–157.

Suhr, J. A., & Tsanadis, J. (2007). Affect and personality correlates of the Iowa Gambling Task. *Personality and Individual Differences*, *43*(1), 27–36.

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*(5), 562–581.

Upton, D. J., Bishara, A. J., Ahn, W.-Y., & Stout, J. C. (2011). Propensity for risk taking and trait impulsivity in the Iowa Gambling Task. *Personality and Individual Differences*, *50*(4), 492–495.

Weller, J. A., Levin, I. P., Shiv, B., & Bechara, A. (2007). Neural correlates of adaptive decision making for risky gains and losses. *Psychological Science*, *18*(11), 958–964.

Whitlow, C. T., Liguori, A., Livengood, L. B., Hart, S. L., Mussat-Whitlow, B. J., Lamborn, C.

M., Laurienti, P. J., & Porrino, L. J. (2004). Long-term heavy marijuana users make costly decisions on a gambling task. *Drug and Alcohol Dependence*, *76*(1), 107–111.

Williams, B. J., & Kaufmann, L. M. (2012). Reliability of the Go/No Go Association Task. *Journal of Experimental Social Psychology*, *48*(4), 879–891.

Wood, R. (2013). Disorders of Impulse Control. In T. M. McMillan & R. L. Wood (Eds.), *Neurobehavioral Disability and social Handicap following Traumatic Brain Injury* (pp. 43–55). Psychology Press.

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*(2), 220–225.

Wright, R. J., Rakow, T., & Russo, R. (2017). Go for broke: The role of somatic states when asked to lose in the Iowa Gambling Task. *Biological Psychology*, *123*, 286–293.

Yong, E. (2012). Replication studies: Bad copy. *Nature*, *485*(7398), 298–300.

Zamarian, L., Sinz, H., Bonatti, E., Gamboz, N., & Delazer, M. (2008). Normal aging affects decisions under ambiguity, but not decisions under risk. *Neuropsychology*, *22*(5), 645–657.

Figner, B., Mackinlay, R. J., Wilkening, F., & Weber, E. U. (2009). Affective and deliberative processes in risky choice: Age differences in risk taking in the Columbia Card Task. Journal of Experimental Psychology: Learning, Memory, and Cognition, 35(3), 709-730. doi: 10.1037/a0014983

Lejuez CW, Read JP, Kahler CW, Richards JB, Ramsey SE, Stuart GL, Strong DR, Brown RA (2002) Journal of Experimental Psychology: Applied, 8, 75-84. PubMed ID 12075692

Müller, S. Schiebener, J. Stöckigt, G. & Brand, M. (2017) Short- and long-term consequences in decision-making under risk: immediate feedback about long-term prospects benefits people tending to impulsive processing. J Cognit Psychol, 29:217–39. doi: 10.1080/20445911.2016.1245660

Weller JA, Levin IP, Shiv B, Bechara A. 2007. Neural correlates of adaptive decision making for risky gains and losses. Psychol Sci.18:958--964.

Zamarian, L., Sinz, H., Bonatti, E., & Delazer, M. (2007). Aging affects decision processes under uncertainty, but not when game rules are explicit. Zeitschrift für Neuropsychologie, 18, 14

Tables

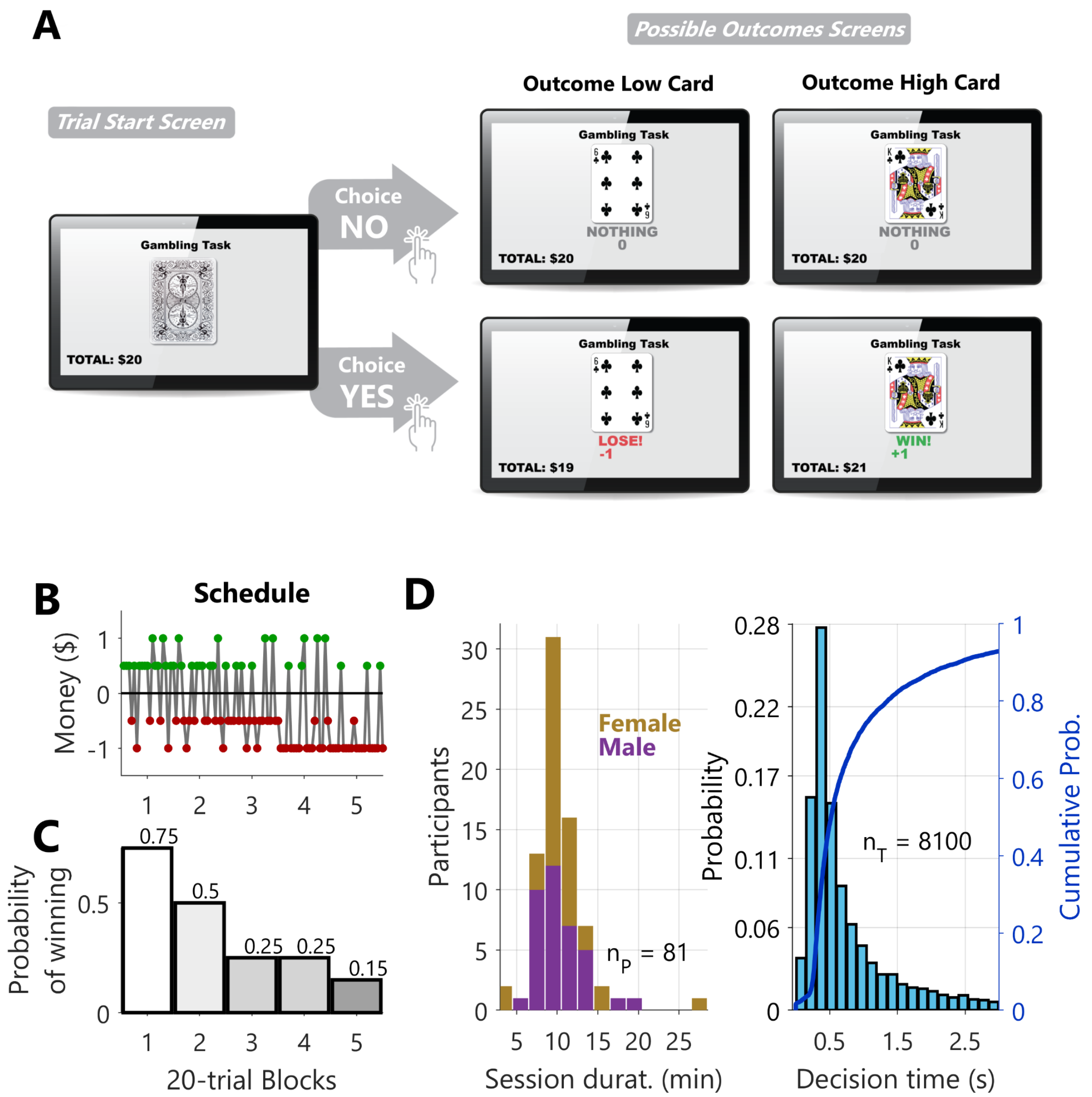
Trial	Block 1	Block 2	Block 3	Block 4	Block 5
1	J Spade [+0.5]	8 Club [-0.5]	6 Diamond [-0.5]	3 Club [-1]	2 Heart [-1]
2	J Club [+0.5]	A Spade [+1]	8 Heart [-0.5]	6 Diamond [-1]	3 Spade [-1]
3	Q Spade [+0.5]	J Heart [+0.5]	9 Spade [-0.5]	2 Heart [-1]	5 Spade [-1]
4	6 Club [-0.5]	6 Diamond [-0.5]	J Club [+0.5]	Q Spade[+0.5]	Q Club [+0.5]
5	Q Diamond [+0.5]	3 Diamond [-1]	8 Club [-0.5]	4 Club [-1]	3 Heart [-1]
6	2 Diamond [-1]	9 Club [-0.5]	J Heart [+0.5]	3 Spade [-1]	6 Heart [-1]
7	J Heart [+0.5]	J Club [+0.5]	10 Diamond [-0.5]	5 Heart [-1]	2 Spade [-1]
8	Q Diamond [+0.5]	10 Spade [-0.5]	4 Heart [-1]	6 Club [-1]	4 Diamond [-1]
9	Q Club [+0.5]	Q Diamond [+0.5]	6 Spade [-0.5]	J Spade [+0.5]	10 Club [-0.5]
10	J Diamond [+0.5]	Q Club [+0.5]	Q Club [+0.5]	K Diamond [+1]	5 Heart [-1]
11	9 Heart [-0.5]	J Diamond [+0.5]	9 Diamond [-0.5]	5 Club [-1]	4 Spade [-1]
12	A Heart [+1]	7 Heart [-0.5]	6 Club [-1]	3 Diamond [-1]	5 Diamond [-1]
13	J Spade [+0.5]	7 Club [-0.5]	8 Diamond [-0.5]	6 Spade [-1]	2 Club [-1]
14	Q Heart [+0.5]	Q Heart [+0.5]	7 Club [-0.5]	7 Diamond [-0.5]	J Club [+0.5]
15	7 Spade [-0.5]	J Spade [+0.5]	K Spade [+1]	K Heart [+1]	2 Heart [-1]
16	K Club [+1]	6 Spade [-0.5]	6 Heart [-0.5]	2 Diamond [-1]	6 Spade [-1]
17	J Diamond [+0.5]	A Diamond [+1]	8 Spade [-0.5]	4 Heart [-1]	4 Diamond [-1]
18	4 Club [-1]	10 Heart [-0.5]	A Heart [+1]	A Club [+1]	3 Club [-1]
19	Q Spade [+0.5]	5 Spade [-1]	10 Spade [-0.5]	6 Heart [-0.5]	Q Diamond [+0.5]
20	Q Heart [+0.5]	J Heart [+0.5]	9 Diamond [-0.5]	5 Spade [-1]	6 Diamond [-1]

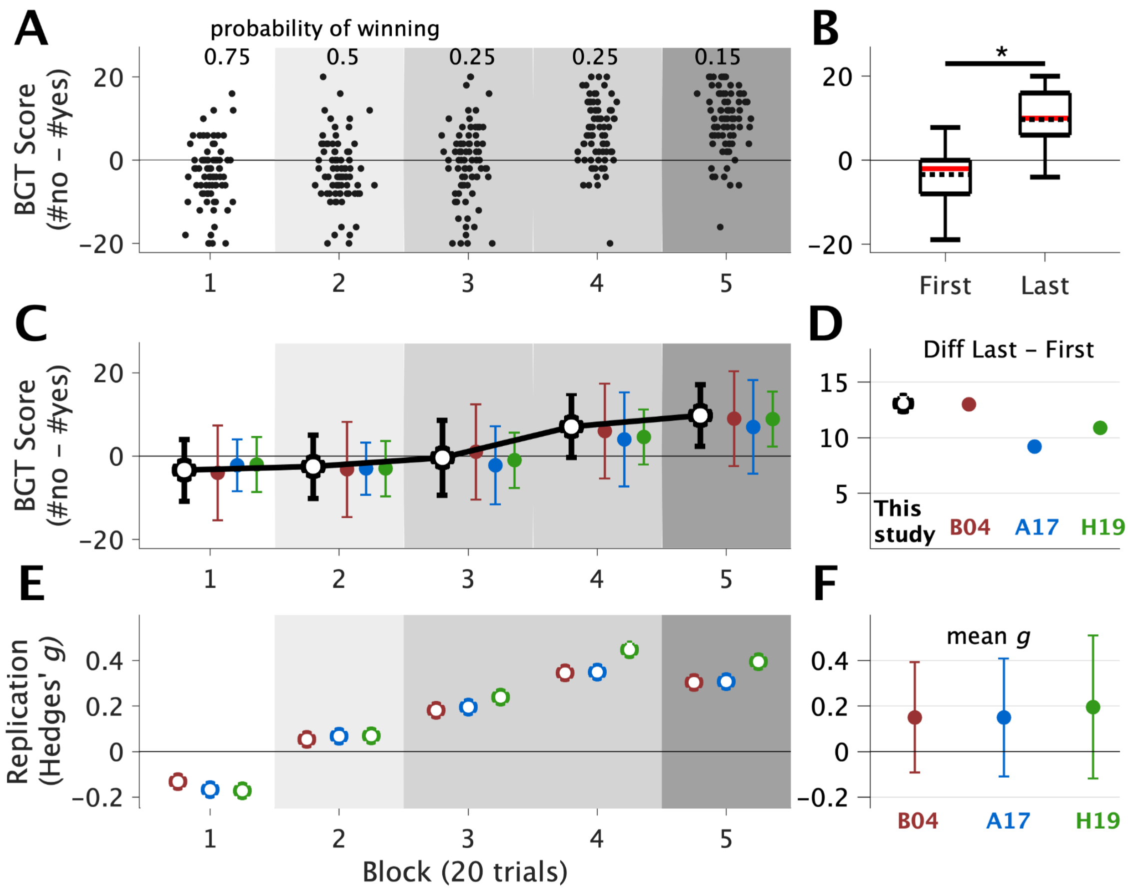
Table 1. Predetermined sequence of cards with wins and losses associated. Table rows correspond to successive trials and columns to successive blocks. On each table entry, we show first the card's number and suit and, in square brackets, the win/loss associated (in thousands of CLP) in case the participant decided to gamble. The set and sequence of cards, as well as the wins/losses, are identical to the one used by Bowman & Turnbull (2004).

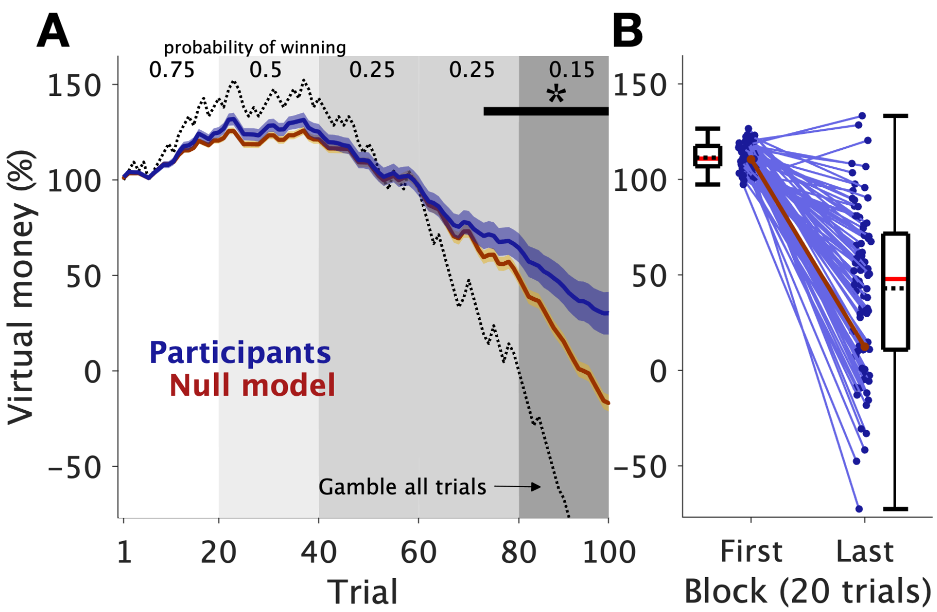
	Bowman & Turnbull, 2004	**Adlam et al., 2017**	**Heininga et al., 2019**
Block 1	0.0565	0.0774	0.081
Block 2	0.3407	0.4749	0.4861
Block 3	0.3932	0.439	0.5692
Block 4	0.4734	0.4791	0.6722
Block 5	0.5032	0.5099	0.7081
Task Mean	0.3534	0.3962	0.5033
Task St Dev	0.178	0.1799	0.2517

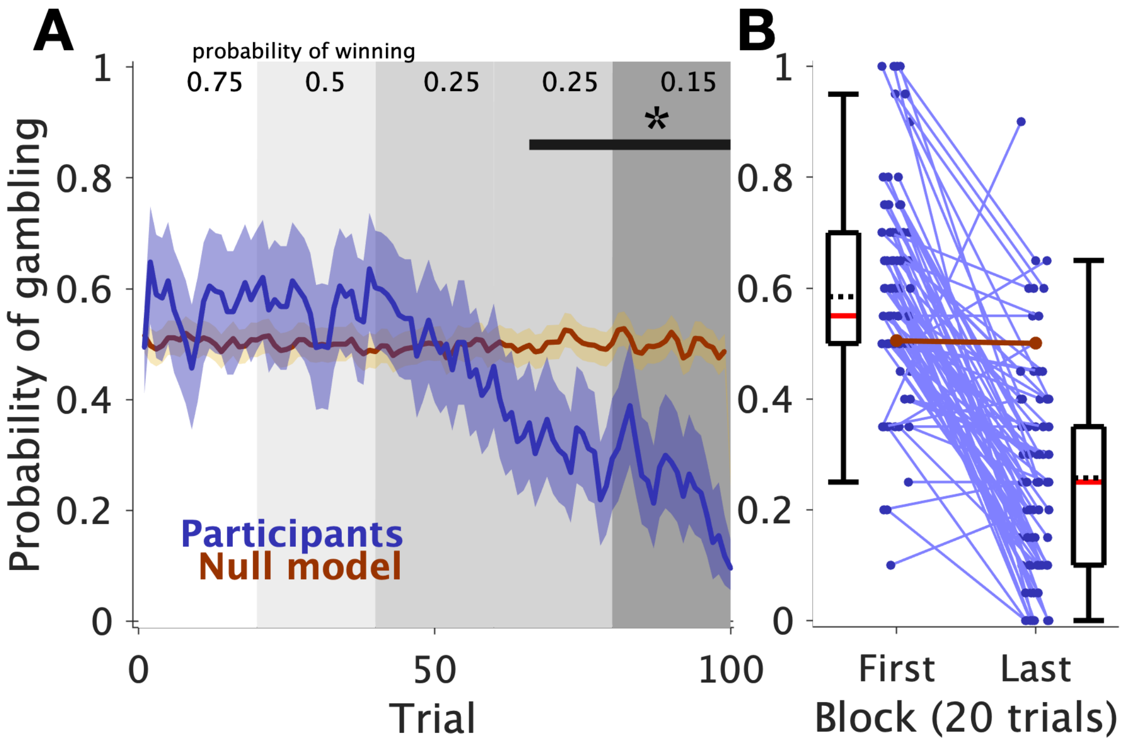
Table 2. Values of Hedges' *g* metric comparing the task scores from the present study with previously published studies. Each column represents a different study, and each row corresponds to a given 20-trial block of the task. The table entries are *g* values, a standardized mean difference between our study and each previous one.

Figures


A
Possible Outcomes Screens
Outcome Low Card
Outcome High Card
Trial Start Screen
Gambling Task
TOTAL: $20
Choice
NO
Choice
YES
NOTHING
0
TOTAL: $20
NOTHING
0
TOTAL: $20
LOSE!
-1
TOTAL: $19
WIN!
+1
TOTAL: $21
B
Schedule
Money ($)
C
Probability of winning
0.75
0.5
0.25 0.25
0.15
20-trial Blocks
D
Participants
Female
Male
n_P = 81
Session durat. (min)
Probability
Cumulative Prob.
n_T = 8100
Decision time (s)


A
probability of winning
0.75
0.5
0.25
0.25
0.15
BGT Score (#no – #yes)
20
0
−20
1
2
3
4
5
B
*
First
Last
C
D
Diff Last – First
15
10
5
This study
B04
A17
H19
E
Replication (Hedges' g)
0.4
0.2
−0.2
Block (20 trials)
F
mean g


A
probability of winning
0.75
0.5
0.25
0.25
0.15
*
Virtual money (%)
150
100
50
0
−50
Participants
Null model
Gamble all trials
1
20
40
60
80
100
Trial
B
First
Last
Block (20 trials)


A
B
probability of winning
0.75
0.5
0.25
0.25
0.15
*
Probability of gambling
Participants
Null model
Trial
First
Last
Block (20 trials)

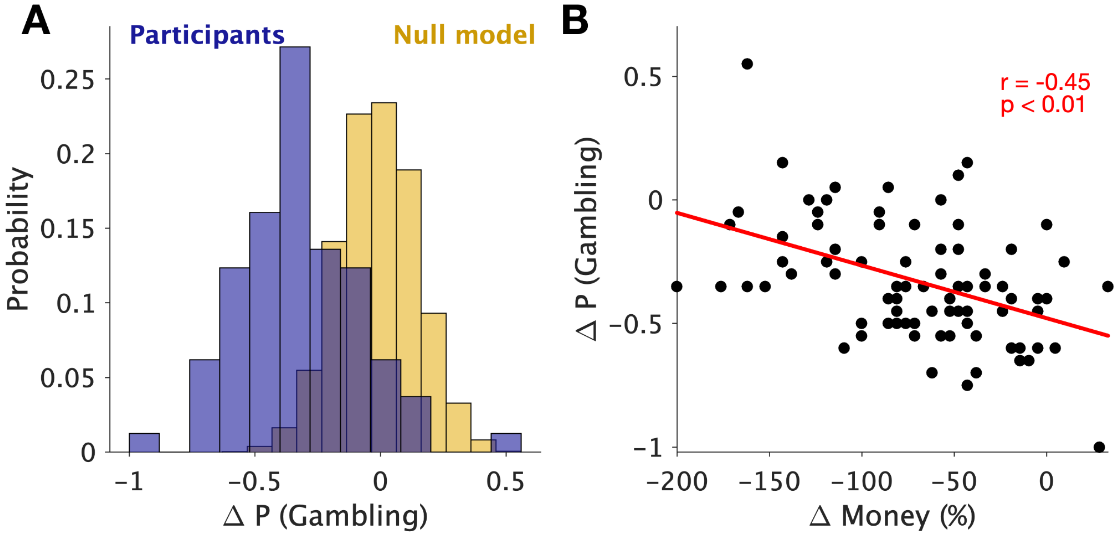

A
Participants
Null model
Probability
Δ P (Gambling)
B
r = -0.45
p < 0.01
Δ P (Gambling)
Δ Money (%)

Figure captions

Figure 1. Structure of the task and descriptive general behavioral results. **A.** Schematic of task screens. The box at the left shows a diagram of the screen shown at the start of the trials, containing the actual deck image (center), the title ('Gambling Task', top center), and the total amount of money (bottom left corner). Participants had to decide whether to gamble or not by pressing one of two keys on a computer keyboard. We show the two possible outcomes on the right side, either a 'low' (i.e., numbered) or a 'high' (i.e., J, Q, K, or A) card. Only high cards were associated with winning, and the probability of getting a high card was a decreasing function of the trial number. If the participant decided to gamble, the card and the amount of money (won or lost) showed up. If the participant chose not to gamble, this involved neither winning nor losing money, and the card also showed up. The total amount of money was also changed and displayed right after the participant's decision. Not drawn to scale. **B.** Schedule of gains and losses. We used the same schedule for every participant, indicated here as the amount gained (green dots) or lost (red dots) in a given trial. The plot shows the gains/losses sequence that would have been obtained had the participant gambled in every single trial. Money is in thousands. Note that the second half of the trial contains not only more frequent losses but also larger losses than the first half. **C.** Winning probabilities. The gains/losses schedule defined a fixed winning probability for each 20-trial block. The task set these probabilities to decrease from 0.75 (first block) to 0.15 (fifth block). **D.** Behavioral descriptive general data. The distribution at the left shows the total task duration, in minutes, for the 81 participants. On the right side, we show the distribution of decision times, in seconds, obtained from the single-trial level (81 participants x 100 trials = 8100). Bars show the distribution function (left y-axis), and

the curve shows the cumulative distribution (right y-axis). Despite having long decision times (e.g., exceeding 20 s), more than 95% of the data consisted of trials with decision times below 2.5 s.

Figure 2. BGT score across blocks and comparisons with data from previous (pencil and paper) versions of the task. **A.** Participants' evolution of BGT scores across blocks. For each participant, we computed the BGT score for each block. The score consists of the subtraction of 'yes' (i.e., gambled trials) choices from 'no' (i.e., not-gambled trials) ones. Each dot represents the BGT score for a participant in a given block. The probability of winning (i.e., the probability of drawing a winning card) decreased from 0.75 (first 20 trials) to 0.15 (last 20 trials). Winning probabilities for each 20-trial block are written in the top portion of the plot and represented as a greyscale of vertical bars. **B.** Distribution of BGT score in the first and last blocks. Boxplots depict the distributions across participants of the first- and last-block BGT scores. Red lines correspond to medians and dotted lines to means. The horizontal bar at the top and the asterisk indicate a significant difference obtained via a t-test ($p<0.001$). **C.** Means and standard deviations of BGT scores across trial blocks from present and previously published data. For each block, we show the mean and standard deviation, computed across-subjects. The black, thick line corresponds to data from the present study. Blue markers correspond to data from Adlam et al., 2017, red markers correspond to data from Bowman et al., 2004, and green markers correspond to data from Heininga et al., 2019. **D.** Difference in BGT scores between the last and first blocks for each study. Colors as in C. B04: Bowman et al., 2004; A17: Adlam et al., 2017; H19: Heininga et al., 2019. **E.** Replication metric between the present and previous studies, across trial blocks. As a replication metric, we computed Hedges' g, a standardized mean difference

between our study and two previous ones, for each task block. **F.** Mean replication metric. We calculated the mean and standard deviation of Hedges' g across blocks for all data sets.

Figure 3. Progression of participants' virtual task money throughout task trials. **A.** Mean amount of money across participants. Money is expressed as a percentage of the initial amount, which was the same for all participants. The grayscale of vertical bars represents the probability of winning, the same as in figure 2. The thick blue curve represents the mean across participants, and the blue shaded area corresponds to the 95% confidence interval. The dark red curve and its associated shaded area represent the mean and 95% confidence interval of the money corresponding to a null model of random gambling (i.e., responses in which gamble and not-gamble have a probability of 0.5). We performed 10^4 simulations of 100-trial sequences of random responses. The dotted black curve represents the money changes obtained if the participant gambled in every trial. The thick black horizontal bar indicates the trials with a statistical difference between the participants' data and the null model. **B.** Change in money from the first to the last block. Each pair of dots connected by a line represents a participant. Dots show the amount of money each participant had at the end of the first and last block. The dark red line corresponds to the corresponding amounts of money calculated for the null model. Boxplots flanking the dots correspond to group distributions for each of the blocks.

Figure 4. Progression of participants' gamble probability throughout task trials. **A.** Mean probability of gambling across participants. The thick blue line represents the mean across participants, computed as a binomial estimate of the probability of success,

and the blue shaded area corresponds to a 95% confidence interval. Probability of winning, null model, and statistical significance as in figure 3. **B.** Change in gambling probability from the first to the last block. Each pair of dots connected by a line represents a participant. The dots show the mean gambling probability of the first and last block, with each dot representing one participant. The dark red line depicts the corresponding probabilities calculated for the null model. Boxplots flanking the dots correspond to group distributions for each of the blocks.

Figure 5. Properties of the task EBDM metric, $\Delta P(Gambling)$. **A**, probability distribution of $\Delta P(Gambling)$ from our data in blue, alongside the distribution obtained from the null model. **B**. Relationship between $\Delta P(Gambling)$ and $\Delta Money$. Money is expressed as a percentage difference, with 100% corresponding to the amount available to the participant at the start of the task. Each dot corresponds to a participant.