# What drive information-seeking in healthy and addicted behaviors

**Irene Cogliati Dezza[1,2,*], Xavier Noel[3], Axel Cleeremans[1], Angela J. Yu[4]**

[1]Centre for Research in Cognition & Neurosciences, ULB Neuroscience Institute, Université Libre de Bruxelles, Belgium

[2]Department of Experimental Psychology, Faculty of Brain Sciences, University College London, London, UK

[3]Faculty of Medicine, Université Libre de Bruxelles, Belgium

[4] Department of Cognitive Science, University of California San Diego, United States

**KEY WORDS**

Novelty-seeking, Information-seeking, Exploration, Problem Gambling, Addiction, Reinforcement learning

**Supplemental Information:** Yes

# ABSTRACT

Information-seeking is an important aspect of human cognition. Despite its adaptive role, we have rather limited understanding of the mechanisms that underlie information-seeking in healthy individuals and in psychopathological populations. Here, we investigate human information-seeking behaviors in healthy individuals and in behavioral addiction by using a novel decision-making task and a novel reinforcement learning model. We compare how healthy humans and addicted individuals differ in the way they trade off a general desire to reduce *uncertainty* (general information-seeking) and a desire for *novelty* (novelty-seeking) when searching for knowledge in the environment. Our results indicate that healthy humans and addicted individuals adopt distinct information-seeking modes. Healthy information-seeking behavior was mostly driven by novelty. Addicted individuals' information-seeking was instead driven by both novelty and general information, with reduced novelty-seeking and increased general information-seeking compared to healthy controls. There are three important implications for our findings: (1) Enhanced novelty-seeking behaviors might be a predictor of wellbeing, (2) behavioral addiction may be marked by a reduction of novelty-seeking and an increase in general information-seeking, (3) the altered information-seeking pattern in addicted individuals may be a compensatory strategy that help them to cope with decision making under uncertainty. By showing healthy humans and addicted individuals adopt distinct information-seeking modes, this study not only sheds light on alterations in decision-making behavior in addiction, but also highlights the likely functional and biological dissociation of novelty-seeking and general information-seeking in the human brain.

# INTRODUCTION

Recent advancements in neuroscience have shown information-seeking to be an essential aspect of human cognition that supports healthy decision-making and goal-directed processing [1 2 3]. Information-seeking is often contraposed to the human tendency of maximizing immediate benefits. A decision-maker who is trying to find out the best restaurant in town may try out all different available options in order to obtain information on the potential benefit of each restaurant, but this information search may be costly or result in unpleasant experiences. Yet, healthy humans finely balance the urge for immediate reward vs. longer-term information gain during repeated choice behavior, thus negotiating an exploration-exploitation trade-off [4 5 6]. Appropriately balancing this tension is a necessary tool for navigating in a world fraught with uncertainty and changeable dynamics. Resolving this tension plays a key role, for instance, in foraging problems [7] and complex decisions in the human daily life [8], and can even boost the performance of artificial agents [9 10]. On the other hand, deficiency in its resolution has been observed in psychopathological conditions such as addiction [11 12]. Previous studies have separately suggested at least two motivational factors that could drive human information-seeking behavior: a general desire to reduce *uncertainty* (or general information-seeking[4 5 13]) and a desire for *novelty* (*novelty*-seeking[14 1]). However, it is still unclear how humans make use of these two motivational factors when seeking information under repeated choices and whether/how general information-seeking and novelty-seeking could independently break down in addiction.

While novelty is only associated with a completely novel item, uncertainty-reduction can promote the exploration of an option beyond the first encounter. The these two motivational factors are however highly related since the uncertainty/information bonus is highest for a novel option. Thus, an uncertainty/information bonus and a novelty bonus can be easily mistaken for each other as statistically significant explanatory factors. To complicate matters, the evidence for general information-seeking[4 5] has come from variants of sequential learning and decision-making tasks (i.e., the bandit tasks), while novelty-seeking has been seen in other types of tasks[14 1], leaving open the possibility that general information-seeking is more important for repeated choice scenario elicited by bandit tasks, while novelty seeking is more important for other scenarios. Here, we explicitly compare general information-seeking and novelty-seeking in a modified version of the bandit task that dissociates the relative contribution of expected reward, novelty, and general information as motivating factors in choice behavior. We also implement a reinforcement-learning type model to quantitatively separate out the importance of these three factors in driving human choice behavior.

In addition to healthy controls (HCs), we also include a sample of individuals with gambling disorder (PGs[15]), a form of addiction without the confound of substance consumption. This allows us to investigate how general information-seeking and novelty-seeking may be compromised in addiction. The focus on problem gambling, as opposed to substance abuse, allows us to target the behaviors underlying addiction without the confounding effects of chronic substance use and abuse [16]. We expect insight on the distinction between novelty-seeking and general information-seeking could be particularly relevant to understanding psychopathologies such as addiction, whereby individuals are trapped into the same behavioral routines (e.g., gambling, substance taking, binge eating) despite the negative consequences associated with them (e.g., financial loss, healthy problems [15]). For example, addicted behaviors may be sustained by a reduced desire for exploring novel opportunities and engaging in novel behavioral patterns, or conversely it may be due to a general reduction in the desire to reduce uncertainty about the environment [12], including previously encountered but imperfectly explored alternatives.

Beyond identifying the processes and mechanisms that are altered in behavioral addiction, the inclusion of the problem gambling group may also reveal modular processes that operate semi-autonomously in the healthy human brain and thus can independently break down in pathological conditions. Indeed, as our study will demonstrate, HCs' information-seeking is mostly driven by novelty, while PGs' information-seeking is characterized by both a reduction in novelty-seeking and increase in general information-seeking. This implies that novelty-seeking and general information-seeking may be supported by separable neural substrates in the human brain.

# METHODS AND MATERIAL

## Participants

Forty (40) unmedicated PGs (mean age = 30.1, 4 females) and twenty-two (22) HCs (mean age = 29.0, 4 females) were recruited from the local communities. The sample size of both groups was based on previous studies [17] [5]. We excluded participants having comorbidity with substance abuse and alcohol use disorder or undergoing psychological and pharmacological treatment and with injuries involving the brain (Table 1; Supplement). Gamblers were selected among those who were gambling at least once per week, while HCs were those without gambling experience in the year preceding experimental participation (Table 1; Supplement). The two groups statistically differed only in terms of gambling severity and years of education (years of education did not correlate with any of the behavioral measures considered in this study and removing PGs with lower years of education did not change the main results reported in the text).

| | **PGs** n=40 | **HCs** n=22 | **Test Statistic** |
|---|---|---|---|
| Gender (M/F) | 36\|4 | 18\|4 | $p = 0.601$ |
| Age | 30.1(9.3) | 29.0(6.6) | $p = 0.982$ |
| Years of Education | 14.7(2) | 16.2(2.2) | $p = 0.037$ * |
| IQ (WAIS block) | 8.4(2.6) | 9.3(1.9) | $p = 0.131$ |
| Gambling Severity (CPGI) | 8.8(6.1) | 0 | $p < 10^{-10}$ * |
| Alcohol use (AUDIT) | 4.6(3.9) | 5.3(3.1) | $p = 0.48$ |
| Drug use (DAST) | 0.225(0.423) | 0.227(0.429) | $p = 0.992$ |
| Smoking dependence (FTND) | n=4 | n=1 | NA |
| Memory Capacity (WAIS) | 10.3(3.5) | 9.7(4.1) | $p = 0.483$ |
| Attentional Control (ACS) | 35.4(9) | 37.5(7) | $p = 0.312$ |
| Depression (BDI) | 5.6(4.9) | 4.2(4.8) | $p = 0.137$ |
| Anxiety (STAI-S) | 35.1(10.9) | 37.9(9.5) | $p = 0.173$ |

| | | | |
|---|---|---|---|
| Anxiety (STAI-T) | 39.6(12.4) | 43.1(11) | $p = 0.2$ |
| Positive Mood (PANAS) | 35.4(6.3) | 36.3(5.3) | $p = 0.701$ |
| Negative Mood (PANAS) | 21.1(7.9) | 19.8(4.8) | $p = 0.808$ |

**Table 1. *Demographic information.*** Mean and standard deviations are shown for each measure. For each comparison, we ran a two-sampled t test, except for gender comparison where chi-squared test was used. The two groups differ only in terms of gambling severity (with no gambling problems reported in the control group) and years of education as often reported in the literature [17]. Note: WAIS IV-Wechsler Adult Intelligence Scale (the block-design component of the WAIS is the subset that best predicts performance IQ [18]); CPGI- Canadian Problem Gambling Index ; AUDIT - Alcohol Use Disorders Identification Test; DAST - Drug Abuse Screening Test; FTND - Fagerström Test for Nicotine Dependence; ACS - Attentional Control Scale; BDI- Beck Depression Inventory; STAI-S - State version of the State-Trait Anxiety Inventory; STAI-T - Trait version of the State-Trait

## Behavioral Task

Participants performed 162 games of a decision-making task [5], which permits the dissociation of reward and information on sequential choices [4] (Fig. 1a, Supplement). Each game consists of two phases (or tasks): participants were initially instructed about which option (deck of cards) to choose from on each trial (*forced-choice task*; Fig. 1b) for six consecutive trials, after which they were free to choose from any of the options (*free-choice task*; Fig. 1c) so as to maximize their total gain. The number of free-choice trials varied from 1 to 6 trials, and was exponentially inversely distributed, such that subjects were most frequently allowed to make 6 free choices. The total gain was shown to the subject at the end of the experiment and converted to a monetary payoff (0.01 euros for every 60 points). We adopted the same conversion procedure for both groups. However, because PGs play regularly with higher amounts of money than those offered in our study, their compensation in the study was 2.5 times more than for HCs. This modification was introduced in order to minimize the differences in motivation between the two groups during the experiment.

When selected, each deck provided a reward (from 1 to 100 points) generated from a truncated Gaussian distribution with a fixed standard deviation of 8 points, and then rounded to the nearest integer. The generative mean for each deck was set to a base value of either 30 or 50 points and adjusted independently by +/- 0, 4, 12, or 20 points (i.e., the generative means ranged from 10 to 70 points) with equal probability, to avoid the possibility that participants might be able to discern the generative mean for a deck after a single observation. The generative mean for each option was stable within a game, but varied across games. The generative mean reward value of the three decks were the same in 50% of the games (*Equal Reward)* and with different values (*Unequal Reward*) in the other 50% of the games. In the Unequal Reward condition, the generative means differed so that two options had the same *higher* reward values compared to the third one in 25% of the games (*High Reward*), and in 75% of the games two options had the same *lower* reward values compared to the third one (*Low Reward*). The appearance of the reward conditions was randomized, as were the assignments of which two arms have the same generative mean within each game (in the Unequal Reward games).

On trials when participants do not choose the option with the highest reward expectation (or reward expectation is equalized across the choices), they can either direct their exploration toward the unexplored or novel alternative (novelty-seeking exploration) or choose at random (undirected or random exploration) [4]. In order to dissociate between these two behavioral patterns, we implemented two conditions in the forced-choice task [4]. Participants were either forced to choose each of the three decks 2 times (*Equal Information*), or to choose one deck 4 times, a second deck 2 times, and the third 0 time (*Unequal Information*). 50% of the games were assigned to the Unequal information condition. The order of card selection was randomized in both information conditions, as was the occurrence of the equal and unequal information conditions. Prior to beginning the main experiment, participants were told that during the forced-choice task, they may sample options at different rates, and that the decks of cards did not change during each game, but were replaced by new decks at the beginning of each new game. However, they were not informed of the details of the reward manipulation or the underlying generative distribution adopted during the experiment.

Considering only the first free-choice trial (the trial where reward and information are least correlated [4]), we then define three types of behaviors, corresponding to three distinct motivational factors: (1) *Novelty-seeking exploration* refers to choosing the novel, never-seen option in the Unequal Information condition; (2) *General information-seeking* refers to choosing partially informative options sampled twice in the Unequal Information condition - these options are still informative when explored but not completely novel; (3) *Reward-seeking* refers to choosing options associated with the highest gain. Additionally, we define a fourth behavior - *undirected exploration*- which refers to choosing options associated with the lowest gain in the Equal Information condition, as this type of choice is neither driven by reward nor by information-seeking.

## Computational modelling

We assume that humans behave according to both reward- and information-related internal beliefs/motivation when preforming the above decision-making task [5]. We formalize this using a reinforcement-learning (RL) type computational model (Fig. 1c). In order to investigate the nature of information valuation in HCs and PGs, we implement a novel computational model that we term the "novelty-knowledge RL" (nkRL) model. nkRL combines reward and information evaluation using a delta learning rule[19] (Eq. S1; Fig. 1d), as in a previously proposed variant (Eq. S3) [5], but nkRL specifically dissociates the values associated with novelty and general information:

$$V_{t,j}(c) = Q_{t+1,j}(c) + \sum_1^t i_{t,j}(c) * k + 1_{\text{novel}} * \nu \quad (1)$$

where $Q_{t,j}(c)$ is the expected reward value on trial $t$ in game $j$ for choice $c$ (computed using Eq. S1), $\sum_1^t i_{t,j}(c)$ is the cumulative information about $c$ acquired through trial t ($i_{t,j}$ is 1 if selected on trial $t$, or 0 otherwise), and $k$ is the *knowledge* (or general information) parameter which defines the weight toward previously acquired information ($k$ being negative means there is a bonus toward lesser known options, while being positive means there is a bonus toward more familiar options). $1_{\text{novel}} * \nu$ captures the value associated with *novelty*, where $1_{\text{novel}}$ is a Kronecker delta function that evaluates to 1 when $c$ has never been seen in the current game and 0 otherwise, and the parameter $\nu$ quantifies the value associated with novelty. Lastly, we assume a choice is made via a softmax function of $V_{t,j}(c)$ [20] (Eq. S2), where the decision policy is controlled by the inverse temperature β (Fig. 1d). nkRL can shed light on the processes that underpin information valuation in both HCs and PGs by distinguishing the effects of reward-seeking and information-seeking on choices (β vs. $k, \nu$), and of novelty and general information on information-seeking ( $\nu$ vs. $k$). The model's parameters are estimated by fitting nkRL to trial-by-trial participants' free choices (see Supplement).

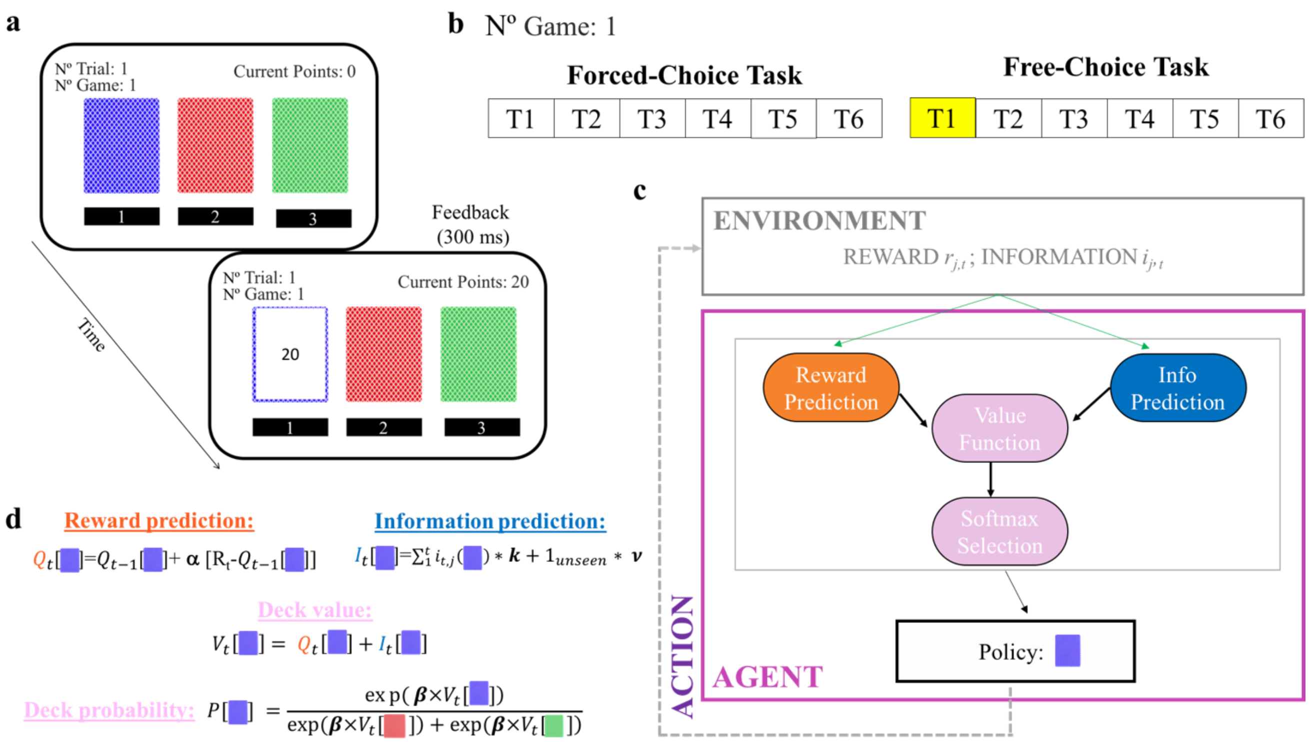



**Figure 1.** ***Behavioral task and RL model.*** **a)** On each trial, participants made choices among three decks of cards. After selecting a deck, the card flipped and revealed the points earned, between 1 and 100 points. Participants were instructed to attempt to maximize the total points earned at the end of the experiment. **b**) On each game, participants play a forced-choice task (6 consecutive trials) followed by a free-choice task (variable between 1 and 6 trials) on the same three decks. Subjects were earning points only on the free-choice task. **c**) On each trial, the novelty-knowledge RL (nkRL) model computes an option value function according to both experienced reward and information associated with each option, then the model generates a choice by passing the option values through a softmax function. **d**) For each chosen option, nkRL uses a delta rule to update the reward prediction ($\alpha$ parameterizes the learning rate), and updates information prediction as sum of general information (total number of times an option has been chosen) and a novelty term. The general information term describes the level of general information participants have about the selected option, while the novelty bonus is assigned to options the outcome of which has never been experienced in previous trials. Reward and information predictions are then combined into an overall action value, which are combined across options to through the softmax function (whose randomness is parameterized by the inverse-temperature parameter β ). Model parameters are shown in bold.

# RESULTS

## Model-free results

### Novelty-seeking in HCs and novelty-failure in PGs

We first examined, in a model-free manner, how HCs and PGs compare in the influence of reward and information on choice behavior. For this, we focus on the Unequal Information condition (equal information games have no informative options) and the first free-choice trial, the one trial where we can be sure that information and experienced reward are uncorrelated [4]. We consider a trial to be *novelty-seeking* if the participant selects the novel option, and *reward-seeking* if the participant selects a previously experienced option with the higher empirical mean (regardless of whether it was seen twice or four times). There may also be trials where the subject chose a previously seen option that had the lower empirical mean reward – those trials were not included in the analysis here. For each subject, we computed the relative frequency of novelty-seeking trials and of reward-seeking trials over the total number of novelty-seeking and reward-seeking trials. We then entered these values into a mixed effects logistic regression predicting choice type (novelty-seeking, reward-seeking) with group (PGs, HCs) and reward condition (Low Reward, High reward) and their interaction as fixed effects and subjects as random intercept term (1|Subject). First, consistent with previous studies using the same experimental design on healthy subjects [5 6], we found a main effect of reward (beta coefficient = -0.824 ± 0.104 (SE), z = -7.90, $p < 10^{-3}$), with novelty-seeking generally more common in the Low Reward condition. More interestingly, we found a significant fixed effect of group (beta coefficient = 0.643 ± 0.268 (SE), z = 2.4, $p$ = 0.016), with PGs engaging in less novelty-seeking and more in reward-seeking behavior (Figure 2a). The interaction between group and reward condition was not significant (beta coefficient = -0.144 ± 0.132 (SE), z = -1.093, $p$ = 0.274), suggesting that the two groups did not differ in the way the reward conditions affected choice behavior.

### PGs and HCs show comparable choice behavior when choices are equally informative

The reduced novelty-seeking behavior in PGs found above could either be due to a specific decrease in the valuation of novelty, or a relative and general increase in the valuation of reward. To examine this, we compare the two groups' behaviors in the Equal Information condition, in which the options have been sampled equal number of times and thus equally informative – any systematic difference in reward-seeking behavior here would be attributable specifically to reward and not influenced by general information or novelty. Again, we focus on the first free-choice trial, where there is no confound between reward and information. We classified choices as *reward-seeking* when choosing the option associated with the highest amount of points and *undirected exploration* otherwise. We then entered these values into a mixed effects logistic regression predicting choice type (reward-seeking, undirected exploration) with group (PGs, HCs) and reward condition (Low Reward, High Reward) and their interaction as fixed effects and subjects as random intercept term (1|Subject). Replicating previous studies using the same experimental design on healthy participants [5 6], we found a fixed effect of reward (beta coefficient = -0.351 ± 0.109 (SE), z = -3.23, $p < 10^{-3}$), with undirected exploration lower in the Low Reward condition. The fact that low reward enhances novelty-seeking but reduces undirected exploration it suggests that these are dissociable exploratory drives in the brain with dissociable neural substrate [6,21]. Most importantly, the effect of group (beta coefficient = 0.113 ± 0.191 (SE), z = 0.589, $p = 0.556$) and the interaction between group and reward (beta coefficient = -0.016 ± 0.135 (SE), z = 0.116, $p = 0.908$) were not significant. The results from the current analysis, along with those from the previous analysis, suggest that the reduced novelty-seeking behavior in PGs is specific to novelty and not an indirect consequence of greater valuation of immediate reward in general (Figure 2b).

**PGs have reduced preference specifically for novelty and not for general information**

Above, we focused our analyses on the first free-choice trial. Here, we examine choices made by participants across the entire set of free choice trials, including both Equal Information and Unequal Information conditions. We classified a trial as an *informative choice* when subjects chose the option sampled the least number of times thus far, and *familiar choice* when they chose the option sampled the most number of times so far. We calculated the number of trials in which each choice was made and divided them by the total number of informative and familiar trials to obtain their relative frequencies (i.e. we exclude trials in which the subject chose the option that was neither most familiar nor most informative). We then entered those values into a mixed effects logistic regression predicting choices (informative, familiar) from group (PGs, HCs) and trial (1,2,3,4,5,6), and their interaction as fixed effects and subjects as random intercepts (1|Subject). We observed a fixed effect of trials (beta coefficient = 0.361 ± 0.016 (SE), z = 19.95, $p < 10^{-3}$), with choices toward familiar options higher at the end of the free choice task – as would be expected if subjects were able to identify the more rewarding options and take advantage of those later on in the game. We also found a fixed effect of group (beta coefficient = 0.418 ± 0.178 (SE), z = 2.36, $p = 0.018$), with choices toward familiar options higher in PGs, consistent with the previous finding that they shy away from novel options. However, we did not observe an interaction effect between group and trial (beta coefficient = -0.012 ± 0.02 (SE), z = -0.599, $p = 0.549$).

We then ran the same analysis considering only trials from the Unequal Information condition. This revealed a fixed effect of group (beta coefficient = 0.546 ± 0.203 (SE), z = 2.69, $p = 0.007$) and fixed effect of trials (beta coefficient = 0.419 ± 0.021 (SE), z = 19.58, $p < 10^{-3}$) as when both equal information and unequal information games were included (Figure 2c,d). However, narrowing the analysis to the Unequal Information condition also revealed an interaction effect between group and trial (beta coefficient = -0.06 ± 0.027 (SE), z =-2.23, $p$ = 0.026), such that the shift in preference from more informative options early on in the free-choice task to more familiar options later on was smaller in PGs than HCs. To better understand this interaction, we compared subjects' tendency to choose the most informative versus most familiar option on the first and sixth trial of the free choice task. We found that control subjects preferred novel options (M= 0.641, SD= 0.257) over familiar options (M= 0.359, SD= 0.257; $p$ = 0.002; Figure 2f) on trial 1, but reversed preferences to prefer familiar options over informative options on trial 6 (M = 0.705, SD = 0.121, $p < 10^{-3}$). In contrast, PGs preferred novel options (M= 0.51, SD = 0.222) and familiar options (M= 0.49, SD = 0.222) equally on trial 1, but strongly preferred familiar options (M= 0.807, SD = 0.149, $p < 10^{-3}$) over informative options (M= 0.193, SD = 0.149) on trial 6. Thus, the "novelty-familiarity" shift was apparent in HCs but absent in PGs.

The above analyses yielded hints that PGs have reduced preference specifically for novelty, indeed the interaction effect between group and trial was only observed when narrowing the analysis to the Unequal Information condition, and in particular to the first free choice trial where novel options are encountered. To test this suggestion, we calculated the number of trials in which participants engaged in novelty-seeking and in general information-seeking (partially informative options sampled twice during the forced-choice task) and divided them by the total number of novel and general information trials to obtain their relative frequencies (i.e. we exclude trials in which the subject chose the option that was selected 4 times during the forced choice task). If alterations in PGs' behavior are not specific to novelty, we should also expect to find lower selection of options experienced twice during the forced-choice task. Results showed that while PGs chose the novel option less often than HCs ($p$ = 0.015, Figure 2e) on the first free-choice trial in the Unequal Information condition, PGs chose the partially informative option (seen twice) *more often* (M =0.446, SD = 0.21) compared to HCs (M =0.32, SD = 0.239; Wilcoxon Signed Rank test, $p$ = 0.015, Figure 2e), suggesting that PGs specifically shy away from novelty-seeking and not from general information-seeking. As an additional check, we constructed a logistic regression to predict choice type (partially informative option, familiar option, i.e. excluding novel option trials) from group (PGs, HCs) as fixed effect and subjects as random intercept term (1|Subject), and found no effect of group (beta coefficient = 0.011 ± 0.088 (SE), z = 0.12, $p$ = 0.905), additionally suggesting no decrease in general information-seeking in PGs compared to HCs. We further examine this point in the next section.

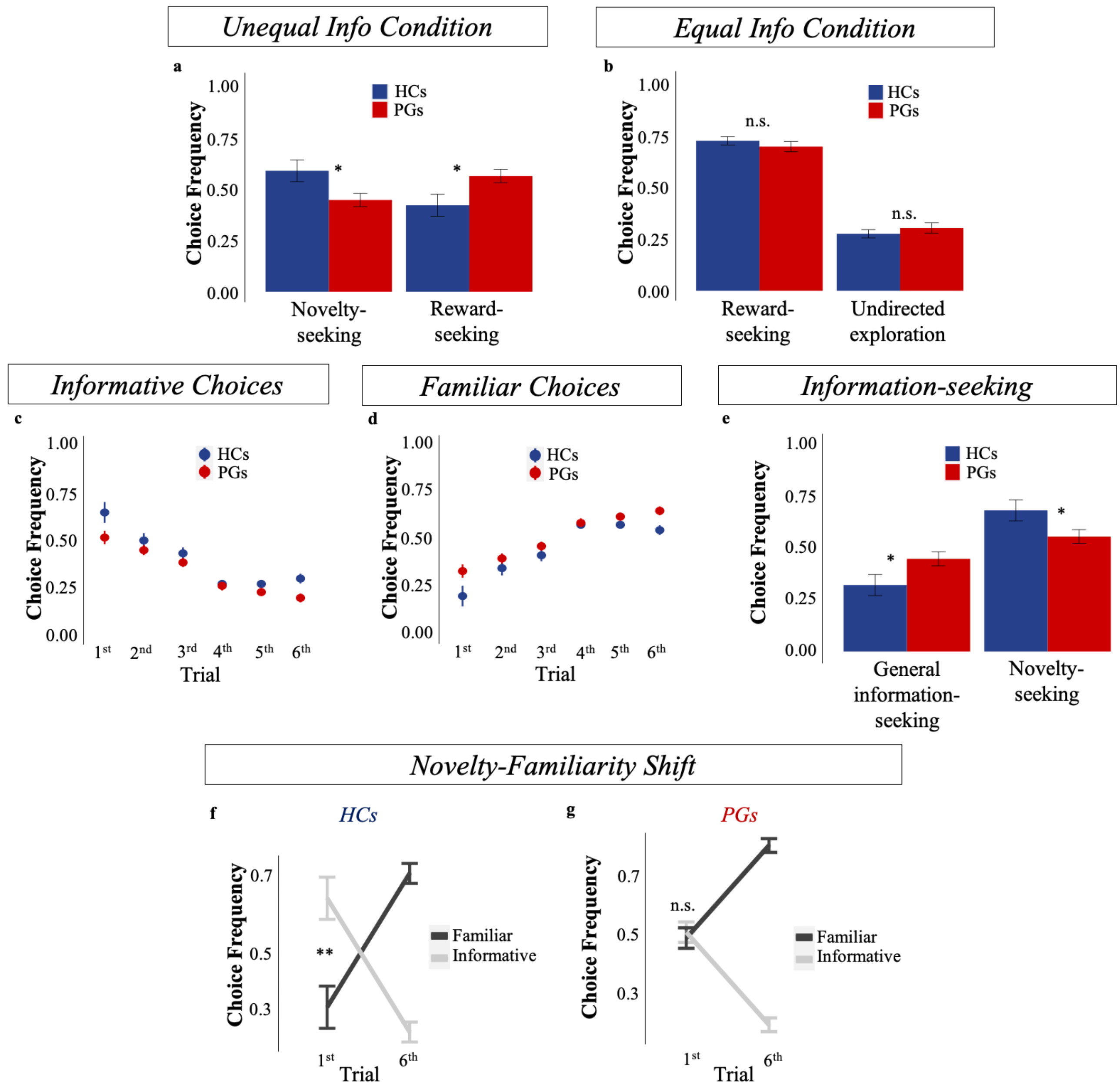



**Figure 2.** ***Model-Free analysis.*** **a**) Frequency of making *novelty-seeking* and *reward-seeking* choices in the first free-choice trial of the Unequal Information condition (i.e., when options are sampled unequally during the forced-choice task; Unequal Info Condition in the figure). Novelty-seeking choices decreased and reward-seeking choices increased in PGs compared to HCs. **b**) Frequency of engaging in *reward-seeking* and *undirected exploration* in the first free-choice trial of the Equal Information condition (i.e., when options are sampled equally during the forced-choice task; Equal Info Condition in the figure). No difference was observed between the two groups. **c**) Frequency of selecting the option seen the *least* number of times in previous trial history (*informative choices*) in the Unequal Information condition. **d**) Frequency of selecting the option seen the *most* number of times in previous trial history (*familiar choices*) in the Unequal Information condition. In **c**, **d,** the frequencies were averaged across games in which participants were choosing informative and familiar options, thus the frequencies add to 1. **e**) Frequency of engaging in information-seeking in the first free-choice trial of the Unequal Information condition: PGs have reduced information-seeking toward novel options (*novelty-seeking*), but increased information-seeking toward options selected twice in the forced-choice task

(*general information-seeking*). **f**) HCs showed a novelty-familiarity shift: increased preference toward informative options in the first free-choice trial and an increased preference for familiar alternatives in the last free-choice trial. **g**) PGs showed no preference between informative and familiar options in the first free-choice trial, but a significant preference toward familiar options on the last free-choice. In all the figures, error bars represent standard error of the mean (s.e.m).

# Model-based results

## HCs have increased novelty bonus, while PGs have increased knowledge parameter

In order to elucidate the mechanisms underlying information-seeking in HCs and PGs, we turn to model-based analyses. Here, we propose a novel reinforcement learning-type model that we call novelty-knowledge RL (nkRL, see Methods). We first ran a model comparison analysis (Supplement) and observed that nkRL was better able to explain participants' behavior compared to the following models: a standard RL (sRL) model [19] - where only reward predictions influence choices; a knowledge RL (kRL) model [5] – which linearly combines reward and information associated with options without explicitly decomposing information into novelty and general information; leaky nkRL where information accumulation across trials proceeds in a leaky fashion; gamma nkRL (gnkRL) where information is measured sub- or super-linearly in the number of observations (Figure 3a, b; Supplement).

We then utilized nkRL to better investigate the process underlying the differences in information-seeking between PGs and HCs. We first simulated nkRL, using the individually fitted parameters, to verify that the model was able to replicate key behavioral patterns observed in the data. As shown in Figure 3, nkRL is able to qualitatively reproduce key behavioral patterns observed in both groups, including reduced novelty-seeking in PGs compared to HCs (Figure 3c), comparable choice behavior when choices are equally informative (Figure 3d), an increase of preference for partially informative options (general information-seeking, Figure 3e), and the absence of novelty-familiarity shift in PGs (Figure 3g).

Next, we performed parameter comparison analyses to examine which component of the decision-making process may be responsible for the behavioral pattern observed in PGs. We first performed a parameter recovery analysis to estimate the degree of accuracy of the fitting procedure (Supplement; Figure S1). We were able to recover all the parameters with high accuracy (all $r > 0.8$). We then compared the parameter estimates between the two groups. A Wilcoxon Signed Rank Test showed smaller novelty parameter $\nu$ in

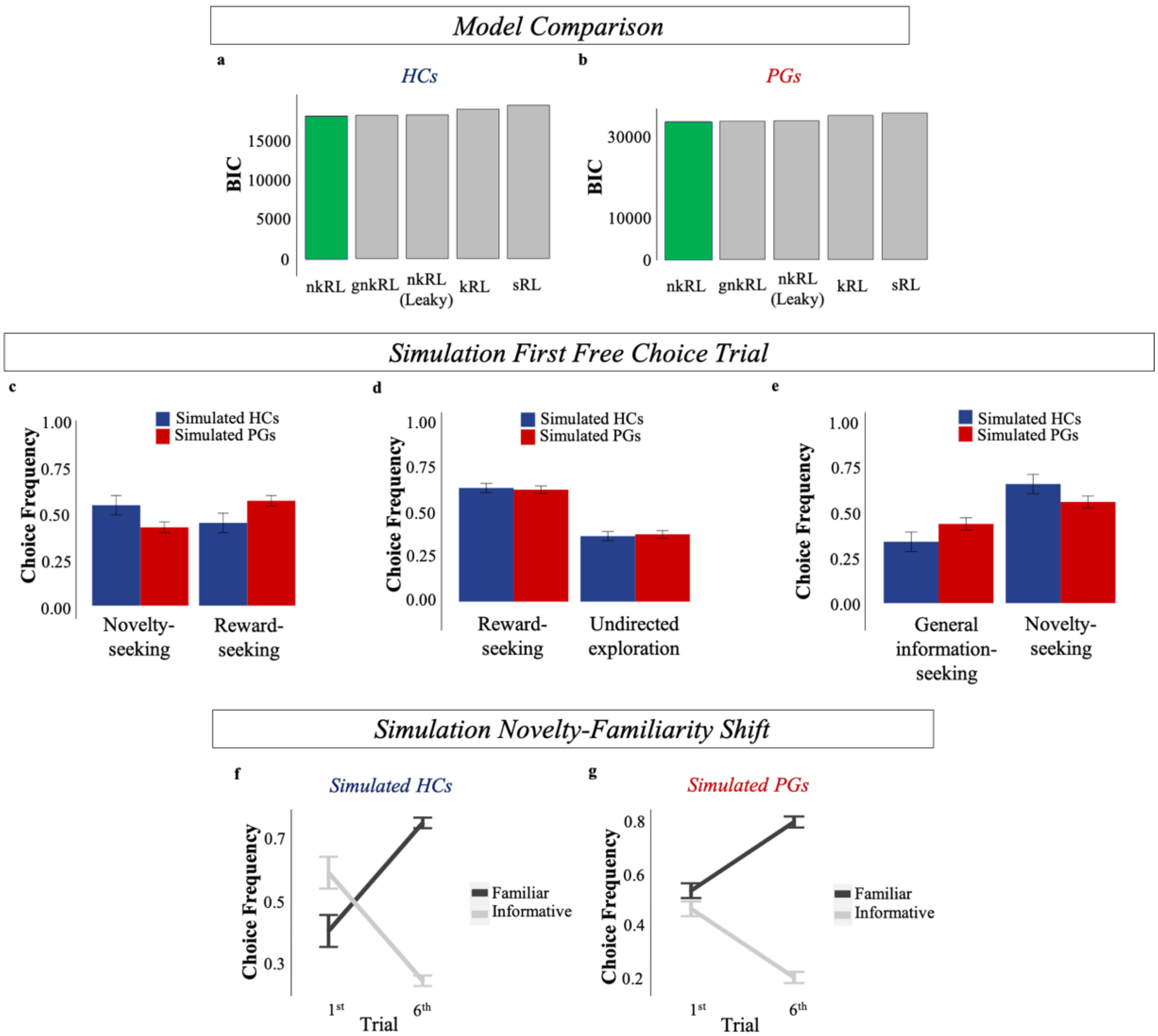



**Figure 3.** ***Model Comparison and nkRL simulations*****.** BIC comparison of the 5 RL models in HCs (**a**) and PGs (**b**). The comparative fit is based on the sum of individual BIC computed by fitting each model to participants' free choices. In both groups, novelty-knowledge RL model (nkRL, in green) better explains participants' behavior compared to a leaky novelty-knowledge RL model (leaky nkRL), a knowledge RL model (kRL), a standard RL model (sRL) and a gamma novelty-knowledge RL model (gnkRL). By using the estimated individual parameters, simulations of nkRL in the first free choice trial reproduced the empirically observed decrease in novelty-seeking in PGs (Unequal Information condition, **c**), comparable choice behavior when choices are equally informative (Equal Information condition, **d**), an increase of preference for partially informative options (general information-seeking, **e**). **f**) nkRL correctly predicts the novelty-familiarity shift in the healthy sample, **g**) and its absence in the PG group. Error bars: s.e.m.

PGs (M = 5.58., SD = 12.11) compared to HCs (M = 12.43, SD = 12.91, $p$ = 0.0416; Figure 4 a), while the knowledge parameter $k$ was higher in PGs (M = 1.38, SD = 2.01) compared to HCs (M = 0.43, SD = 1.04, $p$ = 0.0017; Figure 4 b). In line with our model-free results, these results suggest that PGs have reduced information-seeking for novelty, but not for general information. We further explored this result by entering parameter ($\nu, k$) and group (HCs, PGs) in a two-way repeated measure ANOVA in a non-parametric setting using aligned rank transformation (e.g., ARTool package in R, http://depts.washington.edu/madlab/proj/art/)[22]. This revealed an effect of group ($F(1,58) = 10.06$, $p = 0.002$), an effect of parameters ($F(1,58) = 40.19$, $p<10^{-3}$) and an interaction between group and parameter ($F(1,58) = 18.13$, $p < 10^{-3}$). These results seem to confirm that the decrease in information-seeking in PGs is due to a failure in either computing or utilizing a novelty bonus early on in the free-choice period. As an additional check, by simulating nkRL with a low novelty parameter, the model was able to predict the behavioral pattern observed in PGs (Supplement, Figure S2). Lastly, PGs and HCs did not differ in either learning rate $\alpha$ or softmax parameter $\beta$ ($p < 0.2$; Figure 4c , Figure 4d) suggesting that the behavioral patterns observed in PGs were not related to learning alterations or due to an increase/decrease of random stochasticity in choice distribution. This latter result additionally confirms that exploratory impairments in PGs were specifically driven by novelty-related information valuation without affecting other undirected or unexplained exploratory components (e.g., softmax parameter). Overall, the model-based analyses appear to suggest that HCs are specifically driven by novelty during exploratory behavior, while in gamblers the integration of novelty is reduced and the integration of general information is enhanced.

## HCs and PGs adopt distinct information-seeking modes

Previous analyses showed that PGs exhibit reduced information-seeking for novel options as a consequence of a reduced ability to either computing or utilizing a novelty bonus. However, their preferences for general information was enhanced compared to HCs. These results may suggest that HCs' information-seeking behavior is mostly driven by novelty, while PGs' information behavior by general knowledge. To test this hypothesis, we entered the parameter estimates for novelty and knowledge in a Wilcoxon Signed Rank and we tested their difference against zero. Results showed that novelty was significantly differed from zero in both groups ($p_{HCs} = 0.0003$, $p_{PGs} = 0.002$), while knowledge was significantly different from zero in PGs ($p <10^{-3}$) but there was not substantial evidence in favor of the alternative hypothesis in HCs ($p = 0.065$; $BF_{10} = 1.031$). To better understand whether HCs' information behavior was mostly driven by novelty, we implement an additional model - the novelty RL model (nRL, S8) - which combines both reward and novelty bonus, but eliminates the contribution of general knowledge in the value function. We then fit this model to participants' data (Supplement) and we computed an approximation of model evidence as –BIC/2. We then adopted Bayesian Model Selection [23] to compare nRL to nkRL. nnRL model was better able to explain choice behavior in PGs ($xp_{nkRL}$=0.9999, $BIC_{nkRL}$=33577.2; $xp_{nRL}$=0.0001, $BIC_{nRL}$= 34525.4). However, most HCs were better explained by the novel RL model ($xp_{nkRL}$=0.134, $BIC_{nkRL}$= 18065.6;

$xp_{nRL}$=0.866, $BIC_{nRL}$= 18097.7). The model comparison suggests that PGs and HCs differ in the way information is integrated in the value function: HCs appear to be driven primarily or solely by novelty, while PGs are driven by both novelty and general information. In particular, as suggested by our previous analyses PGs have decreased novelty-seeking but increased general information-seeking compared to HCs.

Next, we analyzed how this particular pattern of altered information-seeking, decreased novelty seeking and increased general information-seeking, might affect PGs' reward accumulation performance in the task. We define task performance as average points earned on free-choice trials, averaged across games. Our results showed no differences in task performance (Π) between PGs and HCs throughout the task (all $p > 0.05$). We then correlated participants' Π with the estimated model parameters for each subject in both groups. We entered Π and model parameters into a correlation matrix where p-values were corrected for multiple comparisons using False Discovery Rate correction (FDR [24]). Results showed that both high novelty parameter and high knowledge parameter relate to higher performance in the task (points earned; $p < 0.05$). This seems to suggest that having either high novelty or high knowledge parameters enables high performance. We further simulated the nkRL model with different settings of knowledge and novelty parameters, while keeping constant both alpha and beta parameters, to understand whether there were indeed two different modes that yield good performance in the task. We computed Π for each simulation and we plotted it in the parameter space. Results showed that two modes gave high performance (Figure 4): one mode with high novelty and low knowledge parameters ($\nu = 19.02$; $\kappa = 5.37$, $\Pi = 48835$ points) and a second mode with similar values for knowledge and novelty parameters ($\nu = 2.55$; $\kappa = 2.97$, $\Pi = 49251$ points). Interesting, average estimated values of $\nu$ and $\kappa$ for the two groups are close to the two locally optimal modes. These results not only suggest that differences between HCs and PGs' information-seeking behavior correspond to adopting two alternative modes of adaptive behavior for the task, but that reward feedback from the task would not be effective for shifting either group's behavior to the alternative local optimum.

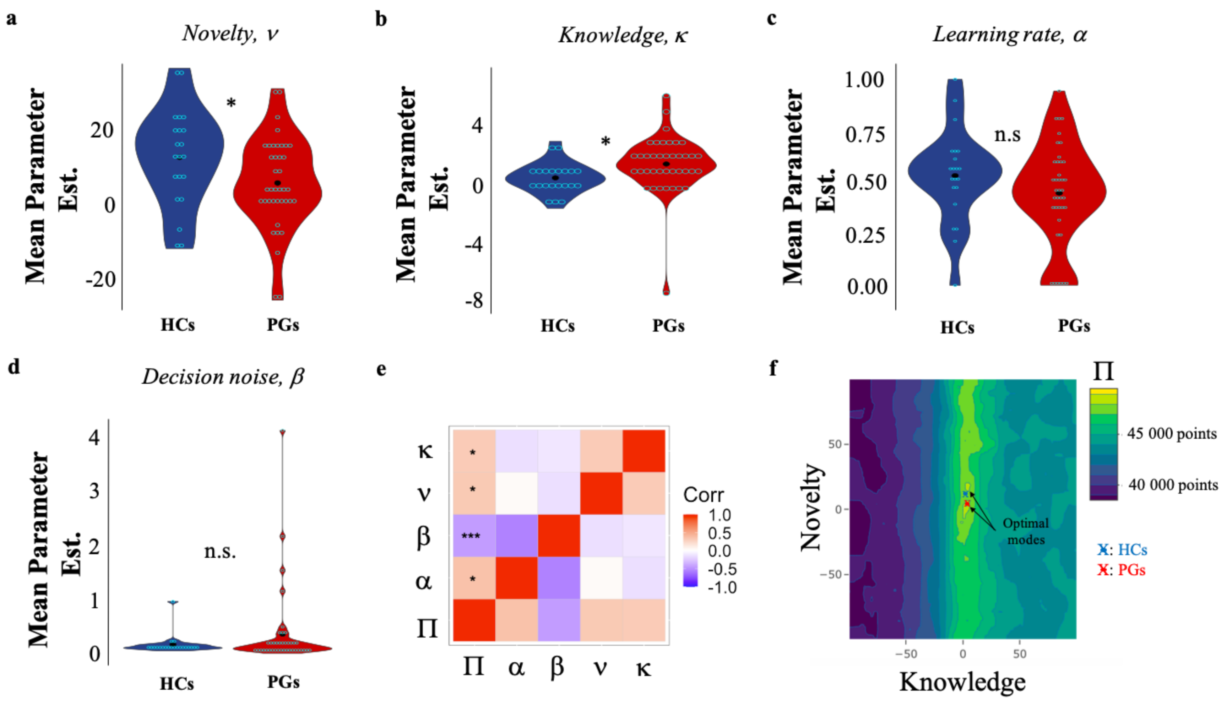



**Figure 4. *nkRL* parameters and information-seeking modes.** Model fit on all free-choice trials revealed a decrease in the novelty parameter ν (**a**) in PGs compared to HCs, while the knowledge parameter κ was higher in PGs compared to HCs (**b**). Learning rate α (**c**) and, decision noise β (**d**) did not differ between the two groups. **e**) Correlation matrix between nkRL model parameters and task performance Π . P-values are corrected for multiple comparison (FDR). Both ν and κ positively correlated with Π. **f**) Performance Π across ν and κ parameter space. Averaged value of ν and κ for HCs is shown in blue, while in red for PGs. The two averaged values are expressed closer to the two optimal modes (in yellow).

# DISCUSSION

In this study, we adopted behavioral, self-reported, and computational measures to investigate the processes underlying healthy and pathological information-seeking. Our results showed that in contrast to previous bandit studies, which found HCs to accord value to general information[4 5], our careful analyses indicate that HCs have a specific novelty bonus, and little to no effect of general information-seeking. Moreover, we found that HCs and PGs adopt distinct information-seeking modes. In particular, HCs' information-seeking behavior was driven by novelty, while PGs' information-seeking behavior was driven by both novelty and general information with higher weights given to the later than to the former. Enhanced novelty-seeking behaviors might therefore be a predictor of wellbeing. We additionally observed that reduced novelty-seeking behavior in PGs did not relate to either greater valuation of reward or decreased desire to reduce uncertainty about the environment. Instead it was due to a reduced ability to either computing or utilizing a novelty bonus and to increased weights to partially informative experiences. One interesting implication of our findings is that the altered information-seeking pattern in addicted individuals may be a compensatory strategy that help them to cope with decision making under uncertainty. More generally, by showing HCs and PGs adopt distinct information-seeking modes, this study not only sheds light on reduced novelty-seeking behaviors in addiction, but it also highlights the likely functional and biological dissociation of novelty-seeking and general information-seeking in the human brain.

Information-seeking is an important aspect of human cognition observed both in healthy humans [1] and animals [25]. Defective information-seeking can indeed evolve in or contribute to certain psychopathologies [26] [27-29] [30]. When humans decide what they want to know, different motives drive their choices [13], including a general desire to reduce *uncertainty* (general information-seeking;[4] [5]) and a desire for *novelty* (novelty-seeking;[14] [1]). Here, we show that under repeated choices the search for knowledge is mostly driven by a desire for novelty. In fact, HCs' behavior was best explained by a model which considered novelty as the unique motive for information-seeking. This novelty bias is essential for learning, exploration [1] and for adapting to the surrounding environment [6]. By showing reduction in novelty-seeking behavior in PGs compared to HCs, our results suggest that novelty-seeking behaviors might be a predictor of wellbeing. Further work is needed to better understand the link between novelty-seeking behaviors and human wellbeing.

In previous RL models, information-seeking under repeated choices (or directed exploration) was modelled as general information or uncertainty parameter added to the value function [4,31] [5] [32]. Here, by using a behavioral task and a model which were able to dissociate novelty-seeking and general information-seeking, we show that HCs mostly rely on novelty bonus when searching for knowledge in partially known environments. Our results, therefore, show a nuanced view over directed exploration and its underlying mechanisms. Moreover, our results replicate previous findings that assign different behavioral roles and neurocognitive mechanisms to informative and undirected component of exploration [4-6,21,33,34]. Indeed, we found PGs reduced directed exploration (defined here as choosing the most informative option- the novel option) and, not undirected (or random) exploration. This emerges both in the model-free analysis and in the model-based analysis where we found that there was no difference between HCs and PGs in terms of the softmax decision policy's temperature parameter.

While HCs' information-seeking behavior was driven by novelty-seeking, PGs' information-seeking behavior was driven by both novelty and general information. Our results therefore suggest that the reduced information-seeking previously observed in this population [12] might be the result of this particular alteration in information-seeking pattern: the novelty bonus is reduced but the weights to partially informative options are enhanced. We further show that this reduction was not due to a greater valuation of reward as usually observed in addicted individuals [35] [36]. This reduced novelty-seeking in PG's may be related to a tendency to quickly jump to conclusions, related to previously suggested abnormalities in confidence judgements and other metacognitive capacities in problem gambling [37] and addiction in general [38]. After seeing the outcome of 2 out of 3 options, they might have been highly confident in their representation of the environment and the search for novel information resulted "unnecessary." However, it may also be possible that the reduced novelty bonus is due a poor ability to dynamically represent the surrounding environment. PGs might be unable to represent changes in the environment, as when new options are available for selection. Model-based impairments have also been found to be associated with addictive disorders [39] [40], and in particular with problem gambling [41]. Future experiments should explicitly test these alternative hypotheses and their relation to reduced novelty-seeking behaviors in PGs.

By focusing on problem gambling, the results of this study clarify that exploratory impairments in addiction [11] are the results of modifications in decision-making processes related to addictive behaviors *per se*, and not by a long-term intake of chemical compounds – although our study does not rule out the possibility that neurophysiological alterations in the brain could pre-date or even induce problem gambling. In particular, it might be possible that individuals who show distinct information-seeking modes may be more predisposed for developing addiction. When addictive behaviors arise, the reduced ability to represent novel behavioral patterns may freeze their decision processes and trap them into the same behavioral routines. The emergence of enhanced general information might then arise as a compensatory mechanism which guarantees the maintenance of their performance. Indeed, aberrant decisions and loss of will power emerge only in certain conditions [42]. For example, addicted individuals can come up with creative solutions, engage in complex decision plans or in goal-directed behaviors in order to obtain the dose they are looking for. Further work is needed to test whether novelty-seeking and general information-seeking may be a potential marker for addiction, and whether these behaviors should be targeted during clinical intervention to reduce the impact of perseveration in addicted individuals.

By showing that HCs and PGs adopt different information-seeking modes, our results appear to suggest that both motives are not only functionally but also biologically dissociable. Information-seeking behaviors are controlled by an interconnected cortico-basal ganglia network [43] and novelty-seeking is believed to be motivated by the dopaminergic system [2 14 44 45]. However, the biological markers of both novelty and general information within the information-seeking network are still unknown. Further work is needed to individuate the neural markers for novelty and general information and their reciprocal expression in addictive individuals.

Although our study adds additional insight on healthy and pathological information-seeking, some limitations may influence the scope of our results. First, in order to have a HC group as similar as possible to the PG group (Table 1), the number of HCs we were able to include in the study after pre-screening was 22 (Supplement). The behavioral pattern observed in the HC group (Figure 2a, S3a), however, replicates our previous findings on healthy humans playing with the behavioral task adopted in the current study [5,6]. Furthermore, although testing PGs appears relevant for minimizing the confounding effects of chemical compounds, most of gambling games involve exploration/exploitation problems. Therefore, the observed behavioral alterations might have been affected by excessive gambling experience. However, we observed no differences between strategic and non-strategic gamblers (who usually play with games that employ different decision strategies, Supplement[46]), and also some HCs had previous gambling experience. Moreover, our findings on alterations in information-seeking behaviors are consistent with previous work on substance addiction where gambling experience was absent. Therefore, it is unlikely that our findings are an artifact resulting from more gambling experience. Lastly, while we showed the PGs and HCs did not differ in terms of decision stochasticity, we cannot rule out that alterations in learning noise [47] may play a role in problem gambling. However, our behavioral task and computational models were not suited to further investigate this question.

Our findings extend the scientific understanding of human information-seeking behavior in healthy individuals and behavioral addiction. HCs and PGs showed distinct information-seeking modes. Healthy information-seeking behavior was motivated by novelty, while PGs' information-seeking behavior by novelty and general information. Our results suggests that the expression of novelty-seeking behaviors might be a potential predictor of human wellbeing, and the expression of altered information-seeking pattern a potential marker of addiction. Methodologically, this work offers promising novel experimental and computational approaches for studying the mechanisms underlying information-seeking under repeated choices in both healthy and pathological populations.

# ACKNOWLEDGEMENT AND DISCLOSURE

ICD was supported by F.R.S.-FNRS grant (Belgium). XN is a research associate the F.R.S.-FNRS (Belgium). AC is a research director at the F.R.S.-FNRS (Belgium). AJY is funded in part by an NIH/NIDA CRCNS grant (R01 DA050373-01). This work was in part funded by ERC Advanced Grant RADICAL to AC. We would like to thank Pauline Deroubaix for helping I.C.D in recruiting participants and collecting their data and Tali Sharot for comments on previous versions of this manuscript. The authors report no competing financial interests.

# REFERENCES

1 Gottlieb, J., Oudeyer, P. Y., Lopes, M. & Baranes, A. Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends Cogn Sci* **17**, 585-593, doi:10.1016/j.tics.2013.09.001 (2013).

2 Kidd, C. & Hayden, B. Y. The Psychology and Neuroscience of Curiosity. *Neuron* **88**, 449-460, doi:10.1016/j.neuron.2015.09.010 (2015).

3 Sakaki, M., Yagi, A. & Murayama, K. Curiosity in old age: A possible key to achieving adaptive aging. *Neurosci Biobehav Rev* **88**, 106-116, doi:10.1016/j.neubiorev.2018.03.007 (2018).

4 Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A. & Cohen, J. D. Humans use directed and random exploration to solve the explore-exploit dilemma. *Journal of experimental psychology. General* **143**, 2074-2081, doi:10.1037/a0038199 (2014).

5 Cogliati Dezza, I., Yu, A. J., Cleeremans, A. & Alexander, W. Learning the value of information and reward over time when solving exploration-exploitation problems. *Sci Rep* **7**, 16919, doi:10.1038/s41598-017-17237-w (2017).

6 Cogliati Dezza, I., Cleeremans, A. & Alexander, W. Should we control? The interplay between cognitive control and information integration in the resolution of the exploration-exploitation dilemma. *Journal of experimental psychology. General*, doi:10.1037/xge0000546 (2019).

7 Stephens, D. W. & Krebs, J. R. *Foraging Theory*. (Princeton University Press, 1986).

8 Schulz, E. *et al.* Structured, uncertainty-driven exploration in real-world consumer choice. *Proceedings of the National Academy of Sciences of the United States of America* (2019).

9 Smirnov, Y., Koening, S., Veloso, M. M. & Simmons, R. G. Efficient Goal-Directed Exploration. *AAAI-96 Proceedings* (1996).

10 Chung, J. J., Lawrance, N. R. J. & Sukkarieh, S. Learning to soar: Resource-constrained exploration in reinforcement learning. *The international journal of robotics research* **34**, 158-172 (2015).

11 Morris, L. S. *et al.* Biases in the Explore-Exploit Tradeoff in Addictions: The Role of Avoidance of Uncertainty. *Neuropsychopharmacology* **41**, 940-948, doi:10.1038/npp.2015.208 (2016).

12 Wiehler, A., Chakroun, K. & Peters, J. Attenuated directed exploration during reinforcement learning in gambling disorder. *bioRxiv preprint* (2019).

13 Sharot, T. & Sunstein, C. R. How people decide what they want to know. *Nat Hum Behav* **4**, 14-19, doi:10.1038/s41562-019-0793-1 (2020).

14 Kakade, S. & Dayan, P. Dopamine: generalization and bonuses. *Neural Netw* **15**, 549-559 (2002).

15 APA. *Diagnostic and Statistical Manual of Mental Disoders*. (American Psychiatry Publishing, 2013).

16 Leeman, R. F. & Potenza, M. N. Similarities and differences between pathological gambling and substance use disorders: a focus on impulsivity and compulsivity. *Psychopharmacology* **219**, 469-490 (2012).

17 Lorains, F. K. *et al.* Strategic and non-strategic problem gamblers differ on decision-making under risk and ambiguity. *Addiction* **109**, 1128-1137, doi:10.1111/add.12494 (2014).

18 Joy, S., Fein, D., Kaplan, E. & Freedman, M. Quantifying qualitative features of Block Design performance among healthy older adults. *Arch Clin Neuropsychol* **16**, 157-170 (2001).

19 Rescorla, R. A. & Wagner, A. R. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning: Current research and theory*, 64-99 (1972).

20 Bishop, C. M. *Pattern recognition and machine learning*. (2006).

21 Zajkowski, W. K., Kossut, M. & Wilson, R. C. A causal role for right frontopolar cortex in directed, but not random, exploration. *Elife* **6**, doi:10.7554/eLife.27430 (2017).

22 Conover, W. J. & Iman, R. L. Rank transformations as a bridge between parametric and nonparametric statistics. *American Statistician* **35**, 124-129 (1981).

23 Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J. & Friston, K. J. Bayesian model selection for group studies. *Neuroimage* **46**, 1004-1017, doi:10.1016/j.neuroimage.2009.03.025 (2009).

24 Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. doi:https://doi.org/10.1111/j.2517-6161.1995.tb02031.x (1995).

25 Bromberg-Martin, E. S. & Hikosaka, O. Midbrain dopamine neurons signal preference for advance information about upcoming rewards. *Neuron* **63**, 119-126, doi:10.1016/j.neuron.2009.06.009 (2009).

26 Camp, C. J. I am curious - grey: Information seeking and depression across the adult lifespan. *Educational Gerontology* **12**, 375-384 (1986).

27 Gray, K. M. & Tonge, B. J. Are there early features of autism in infants and preschool children? *J Paediatr Child Health* **37**, 221-226 (2001).

28 Aderka, I. M., Haker, A., Marom, S., Hermesh, H. & Gilboa-Schechtman, E. Information-seeking bias in social anxiety disorder. *J Abnorm Psychol* **122**, 7-12, doi:10.1037/a0029555 (2013).

29 Van den Driessche, C., Chevrier, F., Cleeremans, A. & Sackur, J. Lower Attentional Skills predict increased exploratory foraging patterns. *Sci Rep* **9**, 10948, doi:10.1038/s41598-019-46761-0 (2019).

30 Hauser, T. U., Moutoussis, M., Consortium, N., Dayan, P. & Dolan, R. J. Increased decision thresholds trigger extended information gathering across the compulsivity spectrum. *Transl Psychiatry* **7**, 1296, doi:10.1038/s41398-017-0040-3 (2017).

31 Auer, P., Cesa-Bianchi, N. & Fischer, P. Finite-time analysis of the multiarmed bandit problem. *Machine Learning* **47**, 235-256 (2002).

32 Gershman, S. J. Deconstructing the human algorithms for exploration. *Cognition* **173**, 34-42, doi:10.1016/j.cognition.2017.12.014 (2018).

33 Somerville, L. H. *et al.* Charting the expansion of strategic exploratory behavior during adolescence. *Journal of experimental psychology. General* **146**, 155-164, doi:10.1037/xge0000250 (2017).

34 Krueger, P. M. Strategies for exploration in the domain of losses. *Judgement and Decision Making* **12**, 104-117 (2017).

35 Bechara, A. Decision making, impulse control and loss of willpower to resist drugs: a neurocognitive perspective. *Nat Neurosci* **8**, 1458-1463, doi:10.1038/nn1584 (2005).
36 Brevers, D., Koritzky, G., Bechara, A. & Noel, X. Cognitive processes underlying impaired decision-making under uncertainty in gambling disorder. *Addict Behav* **39**, 1533-1536, doi:10.1016/j.addbeh.2014.06.004 (2014).
37 Brevers, D. *et al.* Impaired Metacognitive Capacities in Individuals with Problem Gambling. *J Gambl Stud* (2012).
38 Hoven, M. *et al.* Abnormalities of confidence in psychiatry: an overview and future perspectives. *Transl Psychiatry* **9**, 268, doi:10.1038/s41398-019-0602-7 (2019).
39 Simon, D. A. & Daw, N. D. in *Computational Neuroscience of Drug Addiction* Vol. 10 (eds B. Gutkin & S.H. Ahmed) (Springer Series in Computational Neuroscien, 2012).
40 Sebold, M. *et al.* Model-based and model-free decisions in alcohol dependence. *Neuropsychobiology* **70**, 122-131, doi:10.1159/000362840 (2014).
41 Wyckmans, F. *et al.* Reduced model-based decision-making in gambling disorder . *Sci Rep* (2019).
42 Gutkin, B. & Ahmed, S. H. *Computational Neuroscience of Drug Addiction*. (Springer, 2012).
43 White, J. K. *et al.* A neural network for information seeking. *Nat Commun* **10**, 5168, doi:10.1038/s41467-019-13135-z (2019).
44 Lusher, J. M., Chandler, C. & Ball, D. Dopamine D4 receptor gene (DRD4) is associated with Novelty Seeking (NS) and substance abuse: the saga continues. *Mol Psychiatry* **6**, 497-499, doi:10.1038/sj.mp.4000918 (2001).
45 Costa, V. D., Tran, V. L., Turchi, J. & Averbeck, B. B. Dopamine modulates novelty seeking behavior during decision making. *Behav Neurosci* **128**, 556-566, doi:10.1037/a0037128 (2014).
46 Grant, J. E., Odlaug, B. L., Chamberlain, S. R. & Schreiber, L. R. Neurocognitive dysfunction in strategic and non-strategic gamblers. *Prog Neuropsychopharmacol Biol Psychiatry* **38**, 336-340, doi:10.1016/j.pnpbp.2012.05.006 (2012).
47 Findling, C., Skvortsova, V., Dromnelle, R., Palminteri, S. & Wyart, V. Computational noise in reward-guided learning drives behavioral variability in volatile environments. *Nat Neurosci* **22**, 2066-2077, doi:10.1038/s41593-019-0518-9 (2019).

# SUPPLEMENTARY MATERIAL

## What drive information-seeking in healthy and addicted behaviors

**Irene Cogliati Dezza[1,2,*], Xavier Noel[3], Axel Cleeremans[1], Angela J. Yu[4]**

[1] Centre for Research in Cognition & Neurosciences, ULB Neuroscience Institute, Université Libre de Bruxelles, Belgium

[2] Department of Experimental Psychology, Faculty of Brain Sciences, University College London, London, UK

[3] Faculty of Medicine, Université Libre de Bruxelles, Belgium

[4] Department of Cognitive Science, University of California San Diego, United States

# SUPPLEMENTARY METHODS

## Clinical and demographic characteristics

Inclusion/exclusion criteria were examined the day before the experiment by conducting a short telephone interview as well as on the day of the experiment by filling self-reported questionnaires presented in a random order during the last part of the experimental session. The telephone interview was adopted as pre-screening for both PGs and HCs. We specifically asked for information concerning age, gender, frequency of gambling per week (for PGs) or last gambling experience (for HCs), consumption of alcohol per week or substance (including legal and illegal drugs), inability to stop drinking alcohol, undergoing psychological treatments, and possible brain surgeries underwent in the past. We interviewed about N=60 gamblers. Gamblers who met the criteria were then invited to take part to the experiment (N=40). We then took the demographics of the gambling group (gender and age) and we set them as criteria for selecting the control group (alongside with no gambling experience in the year before the study, no sign of excessive use of alcohol or use of substances, psychological treatments, possible brain surgeries etc.). We interviewed about the same number of participants as for the gambling group. More than half of the sample was rejected because of gender (as the gambling group was mostly composed of males) and age (gamblers were quite old compared to usual undergraduates or master students who take part to psychological experiments at the University). In the following two sections, we describe the clinical and demographic characteristics of PGs and HCs.

### *Problem gamblers*

Gambling severity was evaluated using the Canadian Problem Gambling Index (CPGI [1]). Eight gamblers were classified as low level of problem gambling with $1 \leq$ GPCI $\leq 3$, thirteen gamblers with moderate level of problem gambling (leading to some negative consequences; $4 \leq$ GPCI $\leq 7$), and nineteen as exhibiting pathological problem gambling (with negative consequences and possible loss of control; GPCI$\geq$8). We also interviewed participants using DSM-V (French translation) and we observed that 52.4% of PGs met the DSM-V criteria for gambling disorder [2]. The relatively low level of gambling addiction presented in this population is the result of including only participants who showed no co-morbidities with substance abuse or alcohol use disorder. Specifically, to be able to tell apart effects of addictive behaviors *per se* on decision-making from effects of long-term intake of chemical compound, we tested PGs with no use (N= 31, Drug Abuse Screening Test [3]- DAST =0) or non-problematic use (N=9, DAST =1) of legal and illegal substances and with absence of alcohol addiction (Alcohol Use Disorders Identification Test [4]- AUDIT- <12 in men and AUDIT < 11 in women, M = 4.625, SD = 3.868; N=30 did not show any misuse of alcohol

AUDIT< 8). We also controlled for smoking addiction using the Fagerström Test for Nicotine Dependence-FTND [5]. Seven participants reported to smoke, but only 2 were classified with a mid-dependence and 2 with a weak-dependence, the other 3 were not dependent. Given that the main statistical results remained unchanged after removing those participants, we decided to include them in all the analyses. Additionally, to avoid the scenario that participants under psychological treatment may have developed a certain type of cognitive strategy over their decision processes, we included only participants who were not undergoing or seeking for psychological treatment. Moreover, we only included regular gamblers that were gambling at least once per week. Finally, we recruited both strategic PGs (sport betting, poker, black jack; N=22) and non-strategic PGs (bingo, lotto, slot machine, roulette; N=18) [6]. Given that no behavioral difference was found between the two sub-types (in line with [7]), we combined strategic and non-strategic gamblers in the same gambling group in all analyses reported in this manuscript.

*Healthy controls*

The inclusion criteria for the HC group were as follow: CPGI=0 and no gambling experience in the past 12 months. 40% of control participants reported to have gambled in the past years, whereas the rest of the group reported to have never gambled in their life. As for the problem gambling group, we only included participants who scored DAST < 2 (with 17 subjects DAST = 0) and AUDIT < 12 (for the men), 11 (for the women) (with 17 subjects scored AUDIT< 8; M = 5.3, SD = 3.1). Three participants reported to smoke, two of them showed no sign of addiction (FTND = 0 ; 2) and one showed mid-level of addiction (FTND = 7). Removing this participant did not change the main statistical results, therefore the participant was included in all the analyses.

## Behavioral Task

To study information-seeking behavior under repeated choices, we adopt a modified version of a popular task (i.e., the multi-armed bandit) often used to study sequential learning and decision-making behavior. In the bandit task, the decision-maker must make repeated choices among options characterized by initially unknown reward distributions. Each choice can be driven either by a more myopic desire to maximize immediate gain (based on knowledge gained from previous choices and outcomes) or by a more long-term goal of being more informed about all the options. In these repeated scenarios, however, the more the decision-maker tends to choose the most rewarding options, the more those rewarding options tend to be (anti-) correlated with the amount of (remaining) information that can be obtained [8,9]. Accordingly, these classical decision-making tasks make it difficult to quantify exactly how much reward and information each contribute independently to choices [9]. Here, we therefore adopt a novel variant of the bandit task [10], inspired by [9], which has an initial phase of forced choices that carefully controls for reward and information associated with each option. In particular, the influence of reward and information on choices is orthogonalized in the first free-choice trial (since after receiving the feedback on the first free-choice trial, subjects tend to choose the more rewarding options more often, thus reward and information become anti-correlated). Adding a forced-choice task before the actual decision task allows to control for available information and the reward magnitude associated with each option (i.e., options associated with the lowest amount of information were least associated with experienced reward values)[9]. This procedure allows to dissociate between information-driven exploration and undirected exploration. For instance, in the unequal sampling condition, the deck never selected during the forced choice task has highest informative value (it is completely unknown to participants) but it has no reward value associated with. By choosing that deck, participants are engaging in information-driven exploration. On the contrary, in the equal information condition, no differences are observed in terms of information. Therefore, whenever participants choose to explore, this strategy is not driven by an informative drive but only by decision noise[9].

Contrary to our previous versions of this task [10] [11], in half of the games of the equal reward-equal information condition, we introduced an unusually high reward outcome (with respect of the deck mean in that game) for a specific option (e.g., 90 points) the first time that this option was selected in the forced-choice task (subsequently the mean of the deck was set to its original value). This manipulation was introduced as a control condition in order to test whether gamblers' perseverate in choosing a generally poor option that they initially have a good experience with (the 'big win' hypothesis for gambling addiction [12]).

## Computational Modelling

In this section, we provide details on the RL models adopted in this study.

*Standard RL model*

The standard RL (sRL) model learns reward values on each trial using the delta learning rule[13]:

$$Q_{t+1,j}(c) = Q_{t,j}(c) + \alpha \times \delta_{t,j}$$

$$\text{where, } \delta_{t,j} = R_{t,j}(c) - Q_{t,j}(c) \quad \text{(S1)}$$

where $Q_{t,j}(c)$ is the expected reward value for trial *t* and game *j* and $\delta_{t,j}$ is the *prediction error*, which quantifies the discrepancy between the previous predicted outcome $Q_{t,j}(c)$ and the actual outcome $R_{t,j}$ obtained at trial *t* and game *j*. Since participants were told that games were independent from one another, $Q_0$ is initialized at the beginning of each game to the global estimate of the expected reward values for each deck. We previously showed that this initialization was better able to capture healthy participants' behaviour than learning $Q_0$ on a trial-by-trial basis [10]. Next, a choice is made by entering expected reward values into the softmax function [14], as follows:

$$P(c/Q_{t,j}(c_i)) = \frac{\exp(\beta \times Q_{t,j}(c))}{\sum_i \exp(\exp \beta \times Q_{t,j}(c_i))} \quad \text{(S2)}$$

where $\beta$ is the inverse temperature that determines the degree to which choices are randomized by decision stochasticity (or choice variability).

*Knowledge RL model*

As sRL, the knowledge RL (kRL) model learns reward values using Eq. S1 but it additionally integrates information obtained from each deck into the value function:

$$V_{t,j}(c) = Q_{t+1,j}(c) + I_{t,j}(c) * k \quad \text{(S3)}$$

$$\text{where, } i_{t,j}(c) = \begin{cases} 0, & choice \neq c \\ 1, & choice = c \end{cases}$$

κ modulates the importance of information relative to experienced reward. With large κ the model favors already experience decks, while with negative values of κ the model explores new information more frequently. A choice is made by entering choice values $V_{t,j}(c)$ into Eq. S2.

*Novelty-knowledge RL model*

As the above models, the novelty-knowledge RL (*nkRL*) model learns reward values using Eq. S1. And, it additionally integrates information into the value function as kRL. However, as described in the main text, nkRL computes information as a sum of knowledge term and novelty term resulting in the following value function:

$$V_{t,j}(c) = Q_{t+1,j}(c) + \sum_{1}^{t} i_{t,j}(c) * k + 1_{\text{novel}} * \nu \quad \text{(S4)}$$

A choice is made by entering choice values $V_{t,j}(c)$ into Eq. S2.

*Leaky nkRL model*

The leaky nkRL model learns reward values using Eq. S1 and it integrates both knowledge and novelty term into the value function as nkRL. However, in leaky nkRL each bit of new information is integrated in a leaky fashion as follow:

$$V_{t,j}(c) = Q_{t+1,j}(c) + \sum_{1}^{t} i_{t,j}(c) * k + 1_{unseen} * \nu \quad \text{(S5)}$$

$$\text{where, } i_{t,j}(c) = \begin{cases} 0, & choice \neq c \\ 1 * \lambda, & choice = c \end{cases} \quad \text{(S6)}$$

*Gamma nkRL model*

The gamma nkRL (gnkRL) model learns reward values using Eq. S1, and it integrates both knowledge and novelty term into the value function as nkRL. However, gnkRL allows a non-linear integration of information:

$$V_{t,j}(c) = Q_{t+1,j}(c) + \left(\sum_{1}^{t} i_{t,j}(c)\right)^{\gamma} * k + 1_{unseen} * \nu \quad \text{(S7)}$$

$\gamma$ defines both the degree of non-linearity in the amount of observations obtained from options after each observation and its related importance. Under high $\gamma$ the information already gained is highly relevant, whereas the information to be acquired is less relevant or penalized. $\gamma$ is constrained to be > 0.

*Novel RL model*

The novel RL (nRL) model learns reward values using Eq. S1, and it integrates novelty, but not knowledge, into the value function:

$$V_{t,j}(c) = Q_{t+1,j}(c) + 1_{unseen} * \nu \quad \text{(S8)}$$

*Model fitting and Model selection*

The models' parameters were estimated by fitting the model to trial-by-trial participants' free choices (~600 choices for each subject). The fitting procedure was performed using MATLAB function *fminsearchbnd* and iterated for 15 randomly chosen multiple starting points in order to minimize the chance of finding a local optimum instead of a global one. The fitting procedure was validated by running a recovery analysis: the model was simulated on the task using the retrieved parameter estimates to generate synthetic behavioral data and then the fitting procedure was applied to the synthetic data in order to check whether previously estimated parameters were indeed recovered [15] (Figure S1). For model comparisons, negative log likelihoods obtained during the fitting procedure were used to compute model evidence (the probability of obtaining the observed data given a particular model). We adopted an approximation to the (log) model evidence, namely the Bayesian Information Criterion (BIC) [16] and we compared its estimate across different models (fixed-effect comparison). Additionally, we used random-effects procedure to perform Bayesian model selectin at group level [17]. In order to inspect the fitting procedure for overfitting we adopted cross validation procedure [18]. We fitted the model to 70% of the trials and we tested its ability to predict choices on future data (30% of the trials) compared to a simpler nested model. We then adopted the likelihood ratio test to determine if the better fit of complex model was due to noise captured in the data.

## Statistical analysis

Statistical analysis was performed using RStudio (https://www.rstudio.com/). When violations of parametric tests were indicated, non-parametric tests were performed. *P*-values < .05 were considered significant.

# SUPPLEMENTARY RESULTS

## Model comparison

We first examine whether our nkRL model was better able to explain participants' behavior compared to a standard RL (sRL) model [13] -where only reward predictions influence choices- and, to a knowledge RL (kRL) model [10] –which combines both reward and knowledge associated with options without explicitly decomposing information into novelty and general information. We chose kRL as example of unitary models (i.e., information is not decomposed in different drives) because previous researches showed that kRL was better able to explain human behavior in our behavioral task compared to models which update learning rate as number of observations (e.g., Kalman filter, [10]). We fit the 4 models to participants' data and we computed model evidence as approximation of –BIC/2. We removed two subjects (one from each group) for bad fitting. These subjects were removed from all model-based analyses reported in the main text. We then utilized Bayesian Model Selection [17] to compare the 3 models. We found nkRL model was the best model for predicting choice behavior in both HCs ($xp_{nkRL}$=1, $BIC_{nkRL}$=18065.6; $xp_{kRL}$=0, $BIC_{kRL}$= 18918; $xp_{sRL}$=0, $BIC_{sRL}$= 19407; Figure 3a) and PGs ($xp_{nkRL}$=0.877, $BIC_{nkRL}$= 33577.2; $xp_{kRL}$=0.058, $BIC_{kRL}$= 35080.1; $xp_{sRL}$=0.065, $BIC_{sRL}$= 35683.9; Figure 3b). Next, we asked whether participants were integrating complete information into the value function, as predicted by nkRL, or instead information was integrated in a leaky fashion. We implemented a new model (leaky nkRL) where each sample of information integrates as 1*λ, where λ is the leaky integration parameter. Model comparison showed that nkRL model was better able to explain both PGs ($xp_{nkRL}$= 0.9999, $BIC_{nkRL}$= 33577.2; $xp_{leaky\_nkRL}$=0.0001, $BIC_{leaky\_nkRL}$=33795.2; Figure 3b) and HCs' choices ($xp_{nkRL}$= 1, $BIC_{nkRL}$=18065.6; $xp_{leaky\_nkRL}$=0, $BIC_{leaky\_nkRL}$=18188.7; Figure 3a). Lastly, we examined how information affects choice values. It may be the case that at least for certain situations (as in the present task) in which only a few samples of each option are available, additional observations may provide a non-constant amount of information and therefore they may scale choice value in a sub or super-linearly fashion. We compared nkRL, where information is measured linearly in the number of observations, with a model that permits the integration of information sub- or super-linearly (gnkRL). Model comparison showed that nkRL model was better able to explain both PGs ($xp_{nkRL}$=1, $BIC_{nkRL}$= 33577.2; $xp_{gnkRL}$=0, $BIC_{gnkRL}$= 33703.9; Figure 3b) and HCs' choices ($xp_{nkRL}$=1, $BIC_{nkRL}$= 18065.6; $xp_{gnkRL}$=0, $BIC_{gnkRL}$= 18137.4; Figure 3a). Thus, we found nkRL to be the best-fitting model among all those that we considered.

## Parameter recovery

We performed a parameter recovery analysis to estimate the degree of accuracy of the fitting procedure . To do so, we simulated data from nkRL using the parameters obtained from the fitting procedure (*true* *parameters*), and we fit the model to those simulated data to obtain the estimated parameters (*fit parameters*). We then ran a correlation for each pair of parameters [15] (Figure S1). This revealed high correlation coefficients for alpha ($r_{HCs} = 0.8$, $p_{HCs} < 10^{-3}$; $r_{PGs} = 0.9$, $p_{PGs} < 10^{-3}$), knowledge ($r_{HCs} = 0.9$, $p_{HCs} < 10^{-3}$; $r_{PGs} = 0.6$, $p_{PGs} < 10^{-3}$) and novelty ($r_{HCs} = 0.98$, $p_{HCs} < 10^{-3}$; $r_{PGs} = 0.8$, $p_{PGs} < 10^{-3}$). The beta parameter showed high correlation coefficient in PGs ($r = 0.9$, $p < 10^{-3}$). In HCs one participant showed bad fitting while the rest of the group showed high correlation coefficient ($r = 0.97$, $p < 10^{-3}$). We removed this participant during the comparison of the beta parameter.

**Simulations nkRL with random parameters**

In this section, we report the result of the simulation of the nkRL model with random parameters to better understand the effect of novelty on choice behavior. We simulated nkRL with High Novelty and Low Novelty parameter. In each set of simulations, nkRL was simulated 100 times. In High Novelty, the averaged values of the parameters were as follow: alpha (M = 0.513, SD = 0.315), beta (M= 0.52, SD = 0.283), knowledge (M = 0.493, SD = 0.288), novelty (M = 41.38, SD = 11.31). In Low Novelty, we used the following averaged values: alpha (M = 0.519, SD = 0.304), beta (M = 0.51, SD = 0.293), knowledge (M = 0.479, SD = 0.282), novelty (M = -0.839 SD = 0.584). We then classified model choices in reward-seeking (when the model chooses the experienced decks with the highest average of points regardless of the number of times that deck had been selected during the forced-choice task) and novelty-seeking (when the model selects the option never sampled during the forced-choice task) in the first free-choice trial of the unequal information condition. As shown in Figure S2a, under Low Novelty the model increases reward-seeking at the expense of novelty-seeking as observed in PGs (Figure 2a). Next, we calculated the number of trials in which the model was choosing the partially informative option (seen twice) in the first free-choice trials of the unequal information condition and we averaged those estimates across the trials in which the model engages in information-seeking (novelty-seeking + general information-seeking). As shown in Figure S2b, under Low Novelty the model increases the selection of options selected twice during the forced-choice task (general information-seeking) at the expense of novel options as observed in PGs (Figure 2e).

**Personality traits**

In this section, we explore the individual differences between PGs and HCs to investigate whether personal traits could explain the behavioral differences observed throughout our analyses. We focus on intolerance of uncertainty (EII [19]), impulsivity (UPPS-P [20]), sensation-seeking (SSS [21]), and sensitivity to punishment and reward (SPSRQ [22]). Comparisons between HCs and PGs revealed no differences in the scores obtained from

EII ($p$ = .785, $BF_{01}$ = 3.61), UPPS-P ($p$ = .217, $BF_{01}$ = 1.89), SSS ($p$ = .483, $BF_{01}$ = 3.02), and SPSRQ (sensitivity to reward $p$ = .399, $BF_{01}$ = 2.81; sensitivity to punishment $p$ = .266, $BF_{01}$ = 2.4), suggesting that the behavioral alterations observed in PGs are unlikely to be explained as differences in terms of personality traits (or in some cases there was not substantial evidence in favor of the alternative hypothesis). These results appear to suggest that reduced novelty-seeking in PGs may relate to a process or mechanism that is independent from individual subjective preferences toward uncertainty, sensation-seeking, or punishment and reward sensitivity.

## The 'big win' hypothesis

The results reported in this study showed that PGs reduced novelty-seeking behaviors as a consequence of a failure to represent or incorporate a novelty bonus. However, these parametric alterations might have been confounded by the inability of PGs of moving away from an option after experiencing fairly positive outcomes in the past, i.e., the 'big win' hypothesis. To better investigate this point, we computed the empirical probability of choosing an option associated with an unusually high score ("big win" options) when first selected in the forced-choice task. A two-sample t test showed no differences in the probability of choosing the "big win" option in PGs (M = 0.607 SD = 0.187) compared to HCs (M = 0.596 SD = 0.144), $p$ = .798 suggesting that PGs' choice behavior was not driven by the persistence in choosing options associated with unusually good outcomes in the past.

# SUPPLEMENTARY REFERENCES

1 Ferris, J. & Wynne, H. The Canadian problem gambling index: Final report. (Canadian Centre on Substance Abuse., Ottawa (ON)
, 2001).
2 APA. *Diagnostic and Statistical Manual of Mental Disoders*. (American Psychiatry Publishing, 2013).
3 Skinner, H. A. The Drug Abuse Screening Test. *Addict Behav* **7**, 363-371 (1982).
4 Babor, T. F., de la Fuente, J. R., Saunders, J. & Grant, M. (World Health Organization).
5 Fagerström, K. O. Measuring degree of physical dependence to tobacco smoking with reference to individualization of treatment. *Addictive Behaviors*, 235–241 (1978).
6 Lorains, F. K. *et al.* Strategic and non-strategic problem gamblers differ on decision-making under risk and ambiguity. *Addiction* **109**, 1128-1137, doi:10.1111/add.12494 (2014).
7 Grant, J. E., Odlaug, B. L., Chamberlain, S. R. & Schreiber, L. R. Neurocognitive dysfunction in strategic and non-strategic gamblers. *Prog Neuropsychopharmacol Biol Psychiatry* **38**, 336-340, doi:10.1016/j.pnpbp.2012.05.006 (2012).
8 Hertwig, R. & Erev, I. The description-experience gap in risky choice. *Trends Cogn Sci* **13**, 517-523, doi:10.1016/j.tics.2009.09.004 (2009).
9 Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A. & Cohen, J. D. Humans use directed and random exploration to solve the explore-exploit dilemma. *Journal of experimental psychology. General* **143**, 2074-2081, doi:10.1037/a0038199 (2014).
10 Cogliati Dezza, I., Yu, A. J., Cleeremans, A. & Alexander, W. Learning the value of information and reward over time when solving exploration-exploitation problems. *Sci Rep* **7**, 16919, doi:10.1038/s41598-017-17237-w (2017).
11 Cogliati Dezza, I., Cleeremans, A. & Alexander, W. Should we control? The interplay between cognitive control and information integration in the resolution of the exploration-exploitation dilemma. *Journal of experimental psychology. General*, doi:10.1037/xge0000546 (2019).
12 Kassinove, J. I. & Schare, M. L. Effects of the "near miss" and the "big win" on persistence at slot machine gambling. *Psychol Addict Behav* **15**, 155-158 (2001).
13 Rescorla, R. A. & Wagner, A. R. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning: Current research and theory*, 64-99 (1972).
14 Bishop, C. M. *Pattern recognition and machine learning*. (2006).
15 Wilson, R. C. & Collins, A. G. Ten simple rules for the computational modeling of behavioral data. *Elife* **8**, doi:10.7554/eLife.49547 (2019).
16 Schwarz, G. Estimating the dimension of a model. *Ann. Stat* **6**, 461-464 (1978).
17 Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J. & Friston, K. J. Bayesian model selection for group studies. *Neuroimage* **46**, 1004-1017, doi:10.1016/j.neuroimage.2009.03.025 (2009).
18 Daw, N. D. in *Affect, Learning and Decision Making, Attention and Performance XXIII* (eds E.A. Phelps, T.W. Robbins, & M. Delgado) (Oxford University
Press, 2009).

19 Buhr, K. & Dugas, M. J. The Intolerance of Uncertainty Scale: psychometric properties of the English version. *Behav Res Ther* **40**, 931-945 (2002).

20 Billieux, J. *et al.* Validation of a short French version of the UPPS-P Impulsive Behavior Scale. *Compr Psychiatry* **53**, 609-615, doi:10.1016/j.comppsych.2011.09.001 (2012).

21 Zuckerman, M. & al., e. Development of a sensation-seeking scale. *Journal of Consulting Psychology* (1964).

22 Torrubia, R., Avila, C., Moltó, J. & Caseras, X. The sensitivity to punishment and sensitivity reward questionnaire (SPSRQ) as a measure of Gra's anxietyand impulsivity dimensions. *Personality and Individual Differences* **31**, 837-862 (2001).

# SUPPLEMENTARY FIGURES

## Figure Captions

**Figure S1. *Parameter Recovery*.** Correlation between true and fit parameters for nkRL model. True parameters are those recovered during the fitting procedure, while fit parameters are those recovered after fitting the model to synthetic data (obtained by simulating nkRL with parameters estimated in the two groups).

**Figure S1. *nkRL simulations with random parameters*.** Under Low Novelty the model frequently engages in reward-seeking (**a**) and in general information-seeking (**b**).

**Figure S1.**

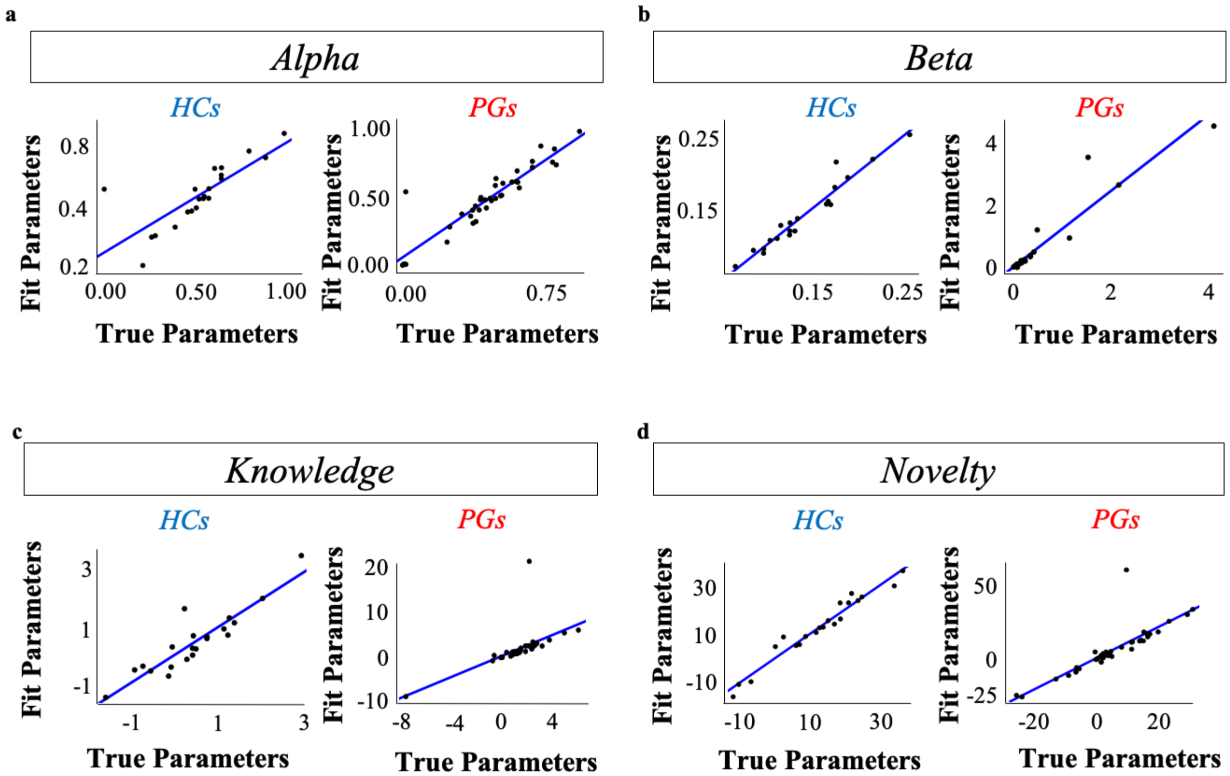
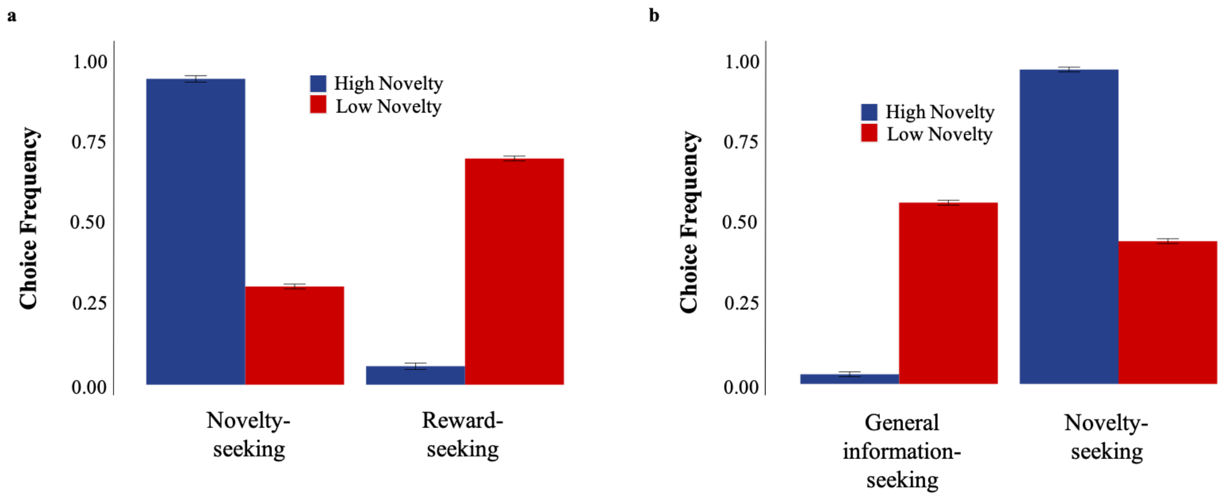

a
Alpha
HCs
PGs
b
Beta
HCs
PGs
c
Knowledge
HCs
PGs
d
Novelty
HCs
PGs
Fit Parameters
True Parameters

**Figure S2.**


a
1.00
0.75
0.50
0.25
0.00
Choice Frequency
High Novelty
Low Novelty
Novelty-seeking
Reward-seeking
b
1.00
0.75
0.50
0.25
0.00
Choice Frequency
High Novelty
Low Novelty
General information-seeking
Novelty-seeking