The Exploration-Exploitation Dilemma as a Tool for Studying Addiction

I Cogliati Dezza* (irene.cogliatidezza@gmail.com)
Neuroscience Institute, Université Libre de Bruxelles, Av. F.-D. Roosevelt 50
Brussels 1050, Belgium

X Noel (xnoel@ulb.ac.be)
Faculty of Medicine, Université Libre de Bruxelles, place Van Gehuchten 4
Brussels 1020, Belgium

A Cleeremans (axcleer@ulb.ac.be)
Neuroscience Institute, Université Libre de Bruxelles, Av. F.-D. Roosevelt, 50
Brussels 1050, Belgium

A Yu (ajyu@ucsd.edu)
University of California San Diego, 9500 Gilman Dr
La Jolla, CA 92093 United States

Abstract:

Addiction is a complex psychiatry condition manifested by the loose of control over drugs or nondrug behaviors despite harmful consequences. In this study, we argue that the exploration-exploitation dilemma and its computational mechanisms can help us understand this disorder and its underlying mechanisms. We tested problem gamblers – for whom the confounding effects of substance abuse typically observed in drug addiction are nullified - and controls in a sequential decision-making task and adopted a reinforcement learning model so as to explore the mechanisms involved. The results show an unbalance in how problem gamblers solved decision problems, where reward learning and information control mechanisms both appear to play key roles. By studying addiction under the exploration-exploitation framework, this study opens up a new way of investigating this disorder and of reinterpreting its main symptoms as impairments of both learning and control mechanisms.

Keywords: gambling addiction; decision-making; RL

Introduction

Addiction is defined as the loss of control over drugs or nondrug behaviors despite negative consequences (e.g, financial problems, health issues). Addiction emerges as a consequence of an unbalance between responses toward rewarding or salient stimuli (eg., money, drug) and inhibitory control over those responses (Noel, Brevers, & Bechara, 2013). Although it has been studied for decades, both in the clinical and in the neuroscientific literatures, a comprehensive understanding of the mechanisms underlying this unbalance is still pending.

In this paper, we argue that the exploration-exploitation dilemma and its related computational mechanisms might bring light upon this issue. The exploration-exploitation dilemma is a decision conflict faced by humans (and animals) in order to flexibly adapt to the surrounding environment: sticking with what they know (familiar rewarding options) vs. selecting unknown but potentially more rewarding alternatives. In order to solve this dilemma, humans adopt different exploratory strategies leading to either choose at random among alternative options (random exploration) or selecting choices directed towards the most informative option among those available (directed exploration; Wilson, Geana, White, Ludvig, & Cohen, 2014). Interestingly, this resolution is guided by the integration of reward and information during the learning process (I. Cogliati Dezza, Yu, Cleeremans, & Alexander, 2017) and it is sustained by behavioral control mechanisms (I. Cogliati Dezza, Cleeremans, & Alexander, under review) (Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006).

To investigate decision-making alterations in addiction under the exploration-exploitation framework, we focused on problem gamblers (PGs) for who the confounding effects of chemical substances are absent, even though such patients present the typical symptoms of addiction as defined in the DSM-V (APA, 2013).

Methods

Participants

Forty PGs and nineteen controls participated to this study. Demographic information can be found in Table 1. To be able to tell a part the decision-making alterations caused by addictive processes from the effects of chemical substances, we recruited PGs with no problematic use of either alcohol (AUDIT < 12) or legal/illegal substances (DAST < 2).

Task

Participants performed 162 games of a modified version of the multi-armed bandit task where, on each game, participants were initially instructed as to which options to choose (*forced-choice task*), after which they were free to choose between options (*free-choice task*) so as to maximize their final gain (Figure 1). This task structure decorrelates reward and information, as shown elsewhere (i.e., the options associated with the lowest amount of information were not associated with experienced reward values), thus enabling the identification of both random and directed exploration in participants' choices (I. Cogliati Dezza et al., 2017; Wilson et al., 2014). When selected, each option provides a reward (from 1 to 100 points) generated from a Gaussian distribution with constant standard deviation and variable mean over games.

Computational model

We adopted a previously implemented version of a reinforcement learning model that learns reward values on each trial and incorporates a mechanism reflecting the knowledge gained about each deck during previous experience - the gamma-knowledge Reinforcement Learning model (gkRL, see I. Cogliati Dezza et al., 2017).

The gkRL model learns reward values using the δ learning rule (Rescorla & Wagner, 1972):

$$Q_{t+1,j}(c) = Q_{t,j}(c) + \alpha \times \delta_{t,j}$$
$$\text{where, } \delta_{t,j} = R_{t,j}(c) - Q_{t,j}(c) \quad (1)$$

Moreover, it accumulates information over time by:

$$I_{t,j}(c) = \left(\sum_{1}^{t} i_{t,j}(c)\right)^{\gamma}$$
$$\text{where, } i_{t,j}(c) = \begin{cases} 0, & choice \neq c \\ 1, & choice = c \end{cases} \quad (3)$$

$I_{t,j}(c)$ is computed by including an exponential term γ that defines the degree of non-linearity in the amount of observations obtained from options after each observation. γ is constrained to be > 0. Both reward and information are integrated in the choice value:

$$V_{t,j}(c) = Q_{t+1,j}(c) - I_{t,j}(c) * \omega \quad (4)$$

where ω determines the degree of which information integrates with reward value. Finally, a choice is made by entering choice values into the softmax function (Bishop, 2006), as follows:

$$P(c/V_{t,j}(c_i)) = \frac{\exp(\beta \times V_{t,j}(c))}{\sum_i \exp(\exp \beta \times V_{t,j}(c_i))} \quad (5)$$

The model's parameters α, β, and ω, γ were free-parameters and were estimated by fitting the model to trial-by-trial participants' choices. The fitting procedure was performed using MATLAB function *fminsearchbnd* and iterated for 15 randomly chosen multiple starting points in order to increase the likelihood of finding the global optimum, instead of a local optimum.

Results

Model-free analysis

We investigated whether problem gamblers resolved the exploration-exploitation dilemma in a different fashion compared to controls. To do so, we computed exploitation, random exploration and directed exploration on the first free choice trial (being the only trial where both strategies can be discerned - Wilson et al., 2014). Trials were classified as "directed exploratory" when participants chose the option that had never been sampled during forced-choice trials, as "exploitative" when participants chose the experienced deck with the highest average of points (regardless of the number of times that deck had been selected during the forced-choice task) and as "random exploratory" when the classification did not meet the previous criteria. A 2 (Group: PGs, Controls) X 2 (Strategy: Directed, Random, Exploit) between-subject ANOVA revealed an effect of strategy $p < 10^{-15}$ and a strategy X group effect $p < 10^{-3}$ (Figure 2). *Post-hoc* comparisons revealed a decrease in directed exploration in PGs (M = 0.377 SD = 0.19) compared to controls (M = 0.526, SD = 0.27) p = 0.04, and an increase in random exploration in PGs (M = 0.146, SD = 0.098) compared to controls (M = 0.098, SD = 0.095), p =0.046. A marginal increase in exploitation was also observed in PGs (M = 0.476, SD = 0.151) compared to controls (M = 0.376, SD = 0.199), p = 0.06.

Model-based analysis

Our previous analysis showed that problem gambling affected how participants balanced the exploration-exploitation dilemma, exploring more randomly overall. In this section, we asked whether this effect was due to an increase in the randomness in participants' choices, or

whether this effect was due to alterations in reward and information processing that subtend the resolution of the dilemma through directed exploration (I. Cogliati Dezza et al., 2017). To understand the underlying mechanisms behind the observed unbalance in solving the exploration-exploitation dilemma in PGs, we adopted the gkRL model and compared the related parameters obtained during the fitting procedure. The analysis showed a decrease in information integration ω in PGs (M = -11.7, SD = 79.5) compared to controls (M = 2, SD = 39.2), $p = 0.007$ (Figure 2 a) whereas the softmax parameter β did not differ between groups, $p = 0.354$ ($M_G = 0.599$, $SD_G = 1.199$; $M_C = 0.141$, $SD_C = 0.069$). Moreover, results showed a decrease in learning rate α in PGs (M =0.39, SD = 0.26) compared to controls (M = 0.55, SD = 0.22), $p = 0.022$ (Figure 2b).

Discussion and conclusion

In this study, we investigated decision-making alterations in problem gambling under the explore-exploit framework. Results showed an unbalance in how PGs solve the dilemma compared to controls. In particular, PGs showed reduced ability to perform directed exploratory strategy, increasing random exploration. To better investigate the hidden mechanisms behind this unbalance, we fitted a reinforcement learning model to trial-by-trial participants' choices. PGs' showed lower and negative values for the information parameter ω compared to controls, whereas the softmax parameter β did not differ between groups. These results appear to suggest that PGs' increased random exploration is not as an effect of the increase in randomness in their choices (as it would have been expected if β differ between groups), but rather as a consequence of an impairment in integrating information into choice values. Integrating information during learning appears to be a control signal that 'neurotypical' humans use to adapt to the surrounding environment by favoring the exploration of the unknown and the disengagement from highly rewarded options (I. Cogliati Dezza et al., under review). By showing a decrease in this ability, this study suggests that inhibitory control processes impaired in substance and behavioral addiction (Noel et al., 2013) might relate to this information control signal.

Moreover, PGs showed a reduced efficiency in updating new reward inputs from the environment as expressed by the decrease in learning rate α compared to controls. An alteration in learning rate might thus play a key role in PGs' inability to stop gambling despite losses (e.g., financial problems) or the 'loss chasing' phenomenon (i.e., gamblers trying to win back money they have already lost by gambling more; Linnet, Peterson, Doudet, Gjedde, & Moller, 2010). The results from this study can help in reinterpreting those phenomena as learning impairments.

In conclusion, through the study of decision-making processes under the exploration-exploitation framework we opened up a new way of understanding problem gambling, and addiction more generally. The results of our study seem to suggest that the core mechanism behind the unbalance between impulsive responses towards rewarding stimuli and inhibitory control processes (Noel et al., 2013) is driven both by poor learning, that is, the inability to update changes in reward contingencies and by weak control processes represented by the information control signal.

Tables

Table 1: Demographic Information. Values shown are the mean and standard deviations on each measure.

	Normal control	Problem Gambling	Test Statistics
Gender	15/4	36/4	
Gambling regularly	No, but they might have experienced gambling but not in the last year	Yes, at least one per week	
Age	28.1 (6.3)	30.1(9.3)	$p = 0.6548$
Alcohol dependence – AUDIT < 12	0	0	
Substance dependence – DAST < 2	0	0	
IQ - WAIS block	9 (1.9)	8.4(2.6)	$p = 0.337$
Memory Capacity – memory subset WAIS	9.4(3.9)	10.3(3.5)	$p = 0.384$
Attentional Control - ACS	37.8(6.4)	35.4(9)	p =0.233
Depression - BDI	4.5(5)	5.6(4.9)	$p = 0.431$
Anxiety – STAI-S, STAI-T	39.6(9.1) 44.9(10.6)	35.1(10.9) 39.6 (12.4)	$p = 0.103$ $p = 0.094$

Note. Values shown are the mean and standard deviations on each measure. AUDIT - Alcohol Use Disorders Identification Test, DAST - Drug Abuse Screening Test, WAIS-Wechsler Adult Intelligence Scale, 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 Anxiety Inventory

Figures

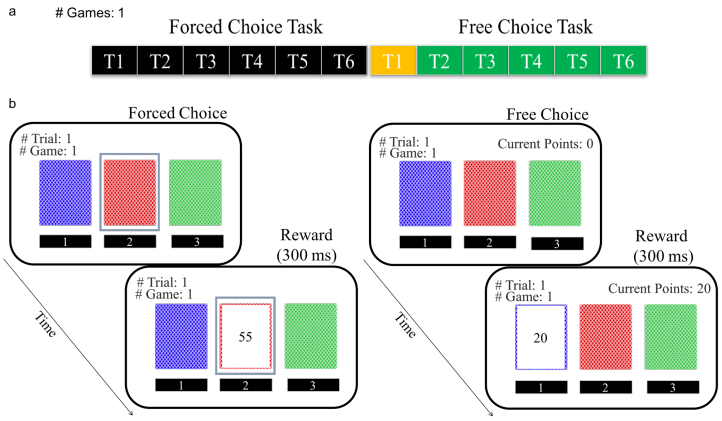

Figure 1: Decision Task. a) Organization of games and trials. In yellow the first free choice trial where reward and information decorrelate (Wilson et al., 2014). b) Explicative trial of forced choice task (left)- participants were forced to choose a pre-selected option; explicative trial of free choice task (right) - participants were free to choose among options.

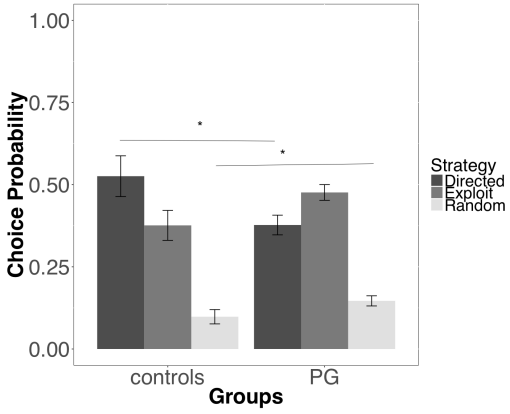

Figure 2: PGs decreased the probability to perform a directed exploratory strategy increasing random exploration compared to controls.

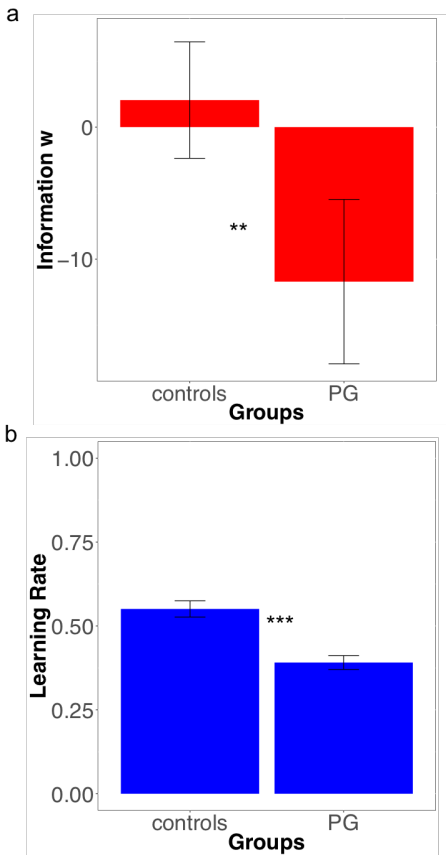

Figure 3: a) Decrease in information parameter ω in PGs compared to controls; b) Decrease in learning rate α in PGs compared to controls.

Acknowledgments

This work was supported by F.R.S.-FNRS grant (I.C.D.) and ERC grant (A.C.). Pauline Deroubaix helped I.C.D in recruiting participants and collecting their data.

References

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

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

Cogliati Dezza, I., Cleeremans, A., & Alexander, W. (under review). Should we control? The interplay between cognitive control and information integration in the resolution of the exploration-exploitation dilemma.

Cogliati Dezza, I., Yu, A. J., Cleeremans, A., & Alexander, W. (2017). Learning the value of information and reward over time when solving exploration-exploitation problems. *Sci Rep, 7*(1), 16919. doi:10.1038/s41598-017-17237-w

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature, 441*(7095), 876-879. doi:10.1038/nature04766

Linnet, J., Peterson, E., Doudet, D. J., Gjedde, A., & Moller, A. (2010). Dopamine release in ventral striatum of pathological gamblers losing money. *Acta Psychiatr Scand, 122*(4), 326-333. doi:10.1111/j.1600-0447.2010.01591.x

Noel, X., Brevers, D., & Bechara, A. (2013). A neurocognitive approach to understanding the neurobiology of addiction. *Curr Opin Neurobiol, 23*(4), 632-638. doi:10.1016/j.conb.2013.01.018

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

Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A., & Cohen, J. D. (2014). Humans use directed and random exploration to solve the explore-exploit dilemma. *J Exp Psychol Gen, 143*(6), 2074-2081. doi:10.1037/a0038199