



**HAL**  
open science

# Motivation intrinsèque pour le choix dans la maladie de Parkinson : implication des ganglions de la base et du système dopaminergique

David Bendetowicz

## ► To cite this version:

David Bendetowicz. Motivation intrinsèque pour le choix dans la maladie de Parkinson : implication des ganglions de la base et du système dopaminergique. Neurosciences [q-bio.NC]. Sorbonne Université, 2022. Français. NNT : 2022SORUS568 . tel-04102413

**HAL Id: tel-04102413**

**<https://theses.hal.science/tel-04102413>**

Submitted on 22 May 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Sorbonne Université

Ecole doctorale Cerveau-Cognition-Comportement

*Institut du Cerveau*

## **Motivation intrinsèque pour le choix dans la maladie de Parkinson**

Implication des ganglions de la base et du système dopaminergique

Par David Bendetowicz

Thèse de doctorat de neurosciences

Dirigée par la Pr. Carine Karachi

Présentée et soutenue publiquement le 22 novembre 2022

Devant un jury composé de :

Dr. Christelle Baunez	Directrice de recherche	Présidente, rapporteure
Pr. Julie Péron	Professeure associée	Rapporteure
Dr. Thomas Boraud	Directeur de recherche	Examineur
Pr. Alexandre Eusebio	Professeur des Universités- Praticien Hospitalier	Examineur
Pr. Carine Karachi	Professeur des Universités- Praticien Hospitalier	Directrice de thèse



## Remerciements

Je tiens tout d'abord à adresser mes plus sincères remerciements au Professeur Carine Karachi pour m'avoir accueilli dans son équipe et accepté de diriger cette thèse. Je tiens particulièrement à te remercier pour ta confiance et la considération que tu as eue pour mon travail. J'ai apprécié ta disponibilité, ta rigueur et ton aptitude à répondre à mes difficultés et mes questionnements tout le long de ce parcours. J'ai été également sensible à la bienveillance et la qualité d'écoute dont tu as fait preuve à mon égard.

Mes remerciements chaleureux vont également au Docteur Jérôme Munuera qui a su lier convivialité et encadrement tout au long de ma thèse. J'ai apprécié ta pédagogie et les nombreux échanges enrichissants qui ont rythmé ces trois dernières années. Merci Jérôme pour les efforts que tu as engagés pour que nous puissions aboutir à ce moment.

Je remercie également le Docteur Brian Lau pour toutes ces discussions passionnantes, pour m'avoir guidé dans mes approches méthodologiques. Je te suis très reconnaissant pour l'aide que tu m'as fournie toutes ces années, et pour ton analyse critique et érudite sur ce sujet complexe.

Je tiens également à remercier le Dr Chantal François pour sa relecture attentive du manuscrit et ses précieux enseignements.

Un grand merci également aux équipes de neurochirurgie et de neurologie qui m'ont épaulé tout le long des inclusions. En particulier le Dr Elodie Hainque pour son soutien et son aide précieuse dans le recrutement des patients.

J'adresse aussi mes remerciements et toute mon amitié au Dr Virginie Czernecki. Alors que tu m'as connu dans mes premiers jours en tant qu'interne, c'est avec grand plaisir que je continuerai de travailler à tes côtés.

J'adresse toute ma gratitude également au Pr Marie-Laure Welter, pour sa vision éclairante de la recherche clinique et pour la justesse de ses observations lors des présentations de mon travail en réunion d'équipe.

Je remercie également Dr Nathalie Georges pour ton œil juste et précis sur les résultats que j'ai pu présenter mais également pour toutes les discussions scientifiques que nous avons pu avoir.

Un immense merci à tous mes petits camarades durant ces trois années en votre compagnie et particulièrement pour ces moments de grande folie qu'étaient les journal clubs (wild), aussi bien à ceux qui se sont envolés vers de nouvelles aventures, Claire, Farah, Dijana, Marie-des-Neiges, Julia, Haissam, Matthieu, Gaétan que les désormais tauliers de l'équipe, Saoussen, Déborah, Yannick, Claire, Eliot, Anna. Sans oublier les deux meilleurs de la plateforme de marche avec qui nous construisons des labyrinthes grandioses, Edward et Clément.

J'ai une pensée toute particulière pour chacun des patients ayant aimablement accepté de participer à cette étude. Je tiens à leur exprimer mon profond respect pour leur courage et leur envie de contribuer à mieux comprendre ce dont ils souffrent.

Enfin, je tiens bien entendu à exprimer ma profonde gratitude à ma famille pour leur soutien indéfectible. Et particulièrement à Yohan, tu as su m'accompagner durant tout ce chemin de la meilleure façon comme tu sais si bien le faire.

## Résumé

La préférence pour des situations comportant du choix a été observée chez l'humain et plusieurs espèces animales. Cette préférence pour le choix présente des avantages qui semblent prévaloir sur l'obtention de récompenses plus basiques. Le choix pourrait présenter un caractère intrinsèquement motivant. Ce caractère récompensant du choix suggère l'implication du système des ganglions de la base, impliqué dans les processus de prise de décision et de pérennisation des actions, modulé fortement par le système dopaminergique, responsable de la valuation des stimuli. La dégénérescence des neurones dopaminergiques et l'altération des ganglions de la base qui en résulte, provoque les symptômes de la maladie de Parkinson. Alors que les traitements améliorent la fonction motrice, ils peuvent induire des modifications comportementales. Par exemple, certains patients parkinsoniens sous traitement dopaminergique présentent une sensibilité accrue aux récompenses extrinsèques. Certains patients traités par stimulation du noyau subthalamique (NST), noyau essentiel des ganglions de la base, peuvent souffrir d'une prise de décisions impulsives. Ce travail de recherche utilise un nouveau paradigme permettant de caractériser la préférence pour le choix comme récompense intrinsèque. Dans une première étape, le sujet a la possibilité de choisir librement ou non. Dans une seconde étape, il choisit entre deux images celle qui délivre le plus de récompenses. Pour limiter le biais d'exploration, les sujets ont réalisé un apprentissage préalable des contingences de récompenses. Des sujets sains, des patients parkinsoniens avec traitement dopaminergique et d'autres traités par stimulation cérébrale profonde du NST ont été inclus. Les patients parkinsoniens ont réalisé la tâche sans (OFF) et avec traitement (ON). Afin d'examiner si des patterns de connectivité pourraient expliquer les réponses des patients stimulés, la tractographie probabiliste des fibres corticales incluses dans le volume de tissu activé au sein du NST a été analysée. Premièrement, chez des sujets sains, nous avons établi que les humains préfèrent les possibilités de choix pour elles-mêmes, indépendamment des récompenses extrinsèques associées à ces possibilités. Chez les patients, il existe un effet aigu de la stimulation du NST, les sujets ON préférant le choix à hauteur de celle des sujets sains, indépendamment des récompenses extrinsèques qu'ils ont reçus. Chez les patients OFF, la préférence pour le choix diminue près du seuil d'indifférence (50%). Il existe une connectivité accrue avec le cortex cingulaire antérieur dorsal droit chez les patients ayant le plus d'effet. Les patients avec traitement dopaminergique présentaient une préférence pour le choix accru si leurs doses quotidiennes étaient plus élevées. Ces résultats font progresser notre compréhension des déterminants neurocognitifs de la récompense intrinsèque et de son interaction avec des processus motivationnels généraux en démasquant les mécanismes associés aux troubles comportementaux observés dans la maladie de Parkinson.

## Abstract

Humans and other animals appear to frequently base decisions on value attribution that is often not clearly linked to extrinsic outcomes. For example, they prefer opportunities to choose, even when these have no positive impact on future rewards. Thus, access to free choice may be intrinsically rewarding, a form of motivation that invigorates an agent to perform behaviors in and of themselves. This rewarding nature of choice suggests the implication of the basal ganglia system, which is strongly modulated by the dopaminergic system. Dopamine neuron degeneration is responsible for the symptoms of Parkinson's disease (PD). While treatments improve motor function, they can induce behavioral changes. For example, some PD patients treated with dopaminergic treatment show altered sensitivity to extrinsic rewards. Some patients treated with subthalamic nucleus (STN) stimulation, an essential structure in the basal ganglia system, may suffer from impulsive decisions and altered learning from extrinsic rewards. We sought to understand how extrinsic and intrinsic motivations act together during decision making. We developed a new paradigm to identify choice preference as an intrinsic reward in healthy humans as well as patients with PD. First, in healthy subjects, we established that humans prefer choice opportunities for their own sake, independent of the extrinsic rewards associated with these opportunities. We then studied PD patients ON and OFF treatment, which allowed to manipulate the dopaminergic system and basal ganglia that are involved in processing of extrinsic reward. In one patient group, treatment consisted of acute dopamine replacement therapy, and in second patient group treatment consisted of deep brain stimulation (DBS) of the subthalamic nucleus (STN). We found that PD patients ON treatment preferred opportunities to choose even though these did not afford more extrinsic rewards. However, patients lose this preference in conditions that diminish the quantity of dopamine in their brains or by reducing the activation of medial prefrontal cortical areas via deep brain stimulation. In particular, we used probabilistic tractography of cortical fibers from DBS stimulation sites to show that higher connectivity with the right dorsal anterior cingulate cortex was associated with increased choice preference induced by STN-DBS. In the group treated with dopaminergic medication, patients with higher chronic doses of dopaminergic medication showed an increased choice preference. These results advance our understanding of the neurocognitive determinants of intrinsic reward and its interplay with general motivational processes by unmasking motivational mechanisms associated with behavioral impairments observed in PD.



## Liste des abréviations

5-CSRT	Five-choice serial-reaction time task
6-OHDA	6-hydroxydopamine
AC-PC	Commissure antérieure-Commissure postérieure
BOLD	Blood-oxygen-level-dependent
CCA	Cortex cingulaire antérieur
DTI	Imagerie par Tenseurs de Diffusion
EEG	Electroencéphalogramme
FLAIR	Fluid-attenuated inversion recovery
FOD	Fiber orientation distribution
FWER	Familywise error rate
GPe	Globus Pallidus externe
GPi	Globus Pallidus interne
HCP	Human Connectome Project
IRM	Imagerie par Résonance Magnétique
LEDD	L-DOPA equivalent dose
MP2RAGE	Magnetization Prepared Two Rapid Gradient Echo
MoCA	Montreal Cognitive Assessment
MPTP	1-méthyl-4-phényl-1,2,3,6-tétrahydropyridine
NST	Noyau Subthalamique
PET	Tomographie par émission de positrons
PCL	Potentiels de champs locaux
QSM	Quantitative Susceptibility Mapping
SCP	Stimulation Cérébrale Profonde
SC	Stimulus conditionné
SI	Stimulus inconditionné
SIFT	Spherical Informed Filtering Technique
SNc	Substance Noire pars compacta
SNr	Substance Noire pars reticulata
TCI	Trouble du contrôle des impulsions

TDM	Tomodensitométrie
TFCE	Threshold-free cluster enhancement
UPDRS	Unified Parkinson Disease Rating Scale
VTA	Volume de Tissu Activé

# Table des matières

Résumé.....	5
Abstract.....	6
Liste des abréviations .....	8
Liste des illustrations.....	14
Liste des tableaux.....	17
INTRODUCTION .....	19
I. Traduire biologiquement la notion de préférence pour le choix .....	21
1. Une propriété commune à plusieurs espèces.....	22
Études chez l'animal.....	22
Etudes princeps chez l'humain : l'illusion du choix.....	27
2. Modèles psychologiques initiaux.....	29
Locus de contrôle et auto-efficacité .....	29
Théorie de l'auto-détermination .....	30
3. Apport des modèles d'apprentissage par renforcement .....	31
La théorie de l'erreur de prédiction .....	34
Le modèle Acteur-Critique.....	37
Vers de nouveaux modèles des récompenses intrinsèques .....	38
4. Conclusion intermédiaire .....	42
II. Hypothèses anatomiques de la préférence pour le choix.....	43
1. Organisation générale des ganglions de la base.....	45
Un système de voies parallèles traitant des modalités différentes.....	46
Contrôle de l'action : rôle des connexions internes des ganglions de la base.....	48

2. La maladie de Parkinson comme modèle de compréhension des ganglions de la base	51
La stimulation cérébrale profonde du NST .....	53
3. Le noyau subthalamique : une voie d'entrée majeure.....	55
Fonctions du NST : apport des études lésionnelles.....	59
Conséquences non motrices de la stimulation du NST dans la maladie de Parkinson	63
4. Les neurones dopaminergiques .....	67
Anatomie fonctionnelle .....	68
Les neurones dopaminergiques encodent une erreur de prédiction .....	70
Conséquences comportementales de la maladie de Parkinson et le traitement médicamenteux .....	73
5. Conclusion intermédiaire .....	76
HYPOTHESES PRINCIPALES.....	77
ETUDE 1 .....	79
Etude 1 : introduction .....	80
Etude 1 : méthodologie .....	81
Etude 1 : résultats .....	84
Expérience 1 : .....	84
Expérience 2 : .....	85
Expérience 3 .....	87
Modèles computationnels .....	88
Conclusion de l'étude 1 .....	89
Article de l'étude 1 .....	91
ETUDE 2.....	138
Etude 2 : introduction .....	140

Etude 2 : méthodologie .....	142
Participants .....	142
Tâche expérimentale .....	144
Déroulement de la passation.....	146
Analyse statistique.....	147
Analyse des données d'imagerie .....	148
Recalage des images .....	149
Reconstruction de la position des électrodes de stimulation.....	150
Détection de l'orientation de l'électrode .....	151
Calcul des volumes de tissu activé (VTA).....	153
Connectivité corticale .....	154
Principe général de l'imagerie de diffusion .....	154
Utilisation d'un connectome normatif.....	154
Tractographie des VTA avec le cortex .....	156
Obtention d'une carte de densité de fibres.....	157
Constitution de groupes .....	158
Analyse statistique voxel à voxel.....	158
Etude 2 : résultats .....	160
Données démographiques.....	160
Première étape : préférence pour le choix.....	160
Seconde étape : choix de l'image la plus récompensée.....	162
Temps de réactions entre les essais « choix libre » et « choix forcé » à la première étape .....	163
Temps de réactions entre les essais « choix libre » et « choix forcé » à la seconde étape .....	164

Analyse de la connectivité des VAT avec le cortex cérébral .....	165
Analyse en sous-groupes des patients DOPA .....	166
Conclusion de l'étude 2.....	168
Article de l'étude 2 .....	169
DISCUSSION .....	201
Le choix permet-il d'augmenter la perception de contrôle ?.....	203
Moyen ou fin : le choix permet-il d'être mieux récompensé ou est-il la récompense ?	204
Le choix pourrait-il être aversif ? .....	206
La préférence pour le choix reflète-t-elle des réponses impulsives chez les patients stimulés ?.....	207
Quel est le rôle du réseau fronto-subthalamique ?.....	209
Le NST est-il impliqué dans des mécanismes motivationnels ?.....	212
Les traitements dopaminergiques ont-ils une durée d'action prolongée ?.....	213
Conclusion .....	215
Limites et perspectives .....	216
Références.....	219

## Liste des illustrations

### Introduction

<b>Figure 1.</b> Labyrinthe de Voss & Homzie.....	23
<b>Figure 2.</b> Schéma du dispositif utilisé par Harlo.....	25
<b>Figure 3.</b> Reproduction de la tâche de Catania.....	26
<b>Figure 4.</b> Représentation schématique du conditionnement simple et instrumental. ....	33
<b>Figure 5.</b> Le modèle Acteur-Critique. ....	37
<b>Figure 6.</b> Nouvelle représentation de l'Acteur-Critique prenant en compte les récompenses extrinsèques et intrinsèques.....	39
<b>Figure 7.</b> Exemple d'essais d'une tâche probabiliste faisant varier le choix et le non-choix. .....	40
<b>Figure 8.</b> Déroulé de la tâche d'information de Bromberg-Martin & Hikosaka.....	44
<b>Figure 9.</b> Boucles fonctionnelles telles que décrites par Alexander .....	46
<b>Figure 10.</b> Représentation simplifiée du modèle d'Albin.....	49
<b>Figure 11.</b> Schéma des trois voies des ganglions de la base.....	51
<b>Figure 12.</b> Représentation de l'évolution des lésions de la maladie de Parkinson. ....	52
<b>Figure 13.</b> Visualisation tridimensionnelle des électrodes de stimulation au sein du NST	55
<b>Figure 14.</b> Répartition des fibres corticosubthalamiques.....	57
<b>Figure 15.</b> Représentation en 3D du NST provenant d'une analyse en cluster .....	58
<b>Figure 16.</b> Représentation de la 5-CRST.....	62
<b>Figure 17.</b> Schéma explicatif du drift diffusion model.....	65
<b>Figure 18.</b> Représentation schématique des voies dopaminergiques.....	68
<b>Figure 19.</b> Illustration des projections réciproques entre les neurones dopaminergiques et le striatum.....	69
<b>Figure 20.</b> L'erreur de prédiction. ....	71

### Etude 1

<b>Figure 1.</b> Structure de la tâche en deux étapes de préférence pour le choix.....	82
<b>Figure 2.</b> Proportion de choix libre en fonction de la récompense extrinsèque moyenne..	84
<b>Figure 3.</b> Dynamique de la proportion de choix libre essai par essai..	85
<b>Figure 4.</b> Proportion de choix libre en fonction de la récompense extrinsèque moyenne pour les choix forcés. ....	85
<b>Figure 5.</b> Dynamique de la proportion de choix libre essai par essai..	86
<b>Figure 6.</b> Résultats de la tâche altérant la contrôlabilité.....	87
<b>Figure 7.</b> Valeurs du bonus pour le choix provenant des modèles à la fois pessimistes et optimistes.....	88

## Etude 2

<b>Figure 1.</b> Exemple de reconstruction des trajectoires des électrodes au sein du NST.....	143
<b>Figure 2.</b> Représentation schématique d'une électrode directionnelle.....	144
<b>Figure 3.</b> Structure de la tâche en deux étapes.....	145
<b>Figure 4.</b> Exemple de vérification visuelle de recalage.....	149
<b>Figure 5.</b> Visualisation de chaque contact activé pour l'ensemble des patients. ....	150
<b>Figure 6.</b> Exemple d'un patient présentant une électrode directionnelle.....	152
<b>Figure 7.</b> Représentation de la connectivité d'un patient parkinsonien.....	155
<b>Figure 8.</b> Représentation schématique d'une FOD.....	156
<b>Figure 9.</b> Représentation schématique d'une technique de tractographie déterministe et probabiliste.....	157
<b>Figure 10.</b> Proportion de libre choix à la première étape. ....	162
<b>Figure 11.</b> Dynamique de sélection de l'image la plus récompensée. ....	163
<b>Figure 12.</b> Temps de réaction lors de la première étape.....	164
<b>Figure 13.</b> Analyse de connectivité entre les patients présentant une différence ON-OFF pour le choix libre.....	166
<b>Figure 14.</b> Résultats de l'analyse en sous-groupes.....	167

Discussion

**Figure 1.** Capture d'image du labyrinthe au moment de la première étape ..... 218

## Liste des tableaux

### Etude 1

**Tableau 1.** Coordonnées (mm) des contacts dans le référentiel AC-PC (CP l'origine)..... 151

**Tableau 2.** Tableau récapitulatif des principales données démographiques. .... 160



## INTRODUCTION



## **I. Traduire biologiquement la notion de préférence pour le choix**

L'idée que nous soyons maîtres de nos actions s'inscrit dans une longue tradition de courants de pensée avec lesquels nous sommes familiers et dont les bénéfices semblent incontestables. Plusieurs termes autour de cette idée s'y rattachent comme agir selon son « libre arbitre », ou bien acquérir une « autonomie ». Ces notions semblent intuitivement avantageuses pour l'individu car ils permettent d'acquérir le contrôle sur notre environnement, affirmant ainsi notre compétence dans un domaine (Bandura, 1977). Ces concepts de contrôle et de compétence sont au cœur de nombreuses études de la seconde moitié du XXème siècle. Elles ont tenté d'en analyser les différents déterminants chez l'humain, par exemple l'impact social ou l'éducation, et de proposer une théorie sur les moyens qu'ont les individus pour exercer le contrôle. D'autres équipes ont dirigé leurs recherches sur le comportement animal. Il a été ainsi observé chez plusieurs espèces que prédire ou contrôler certains événements, comme l'arrivée prochaine de nourriture, permettait de diminuer leur stress et par conséquent d'améliorer leur bien-être (Bassett & Buchanan-Smith, 2007).

Ainsi, d'une notion abstraite telle que le libre-arbitre, dont l'enjeu philosophique est au premier plan, se dégage un comportement retrouvé chez des espèces non humaines suggérant qu'avoir le contrôle sur son environnement pourrait avoir des bases biologiques. Il reste encore à définir les situations permettant d'avoir du contrôle. Parmi ces dernières, la notion de choix a fait l'objet de nombreuses études montrant que choisir librement serait un comportement non seulement désirable mais nécessaire pour la survie de nombreuses espèces.

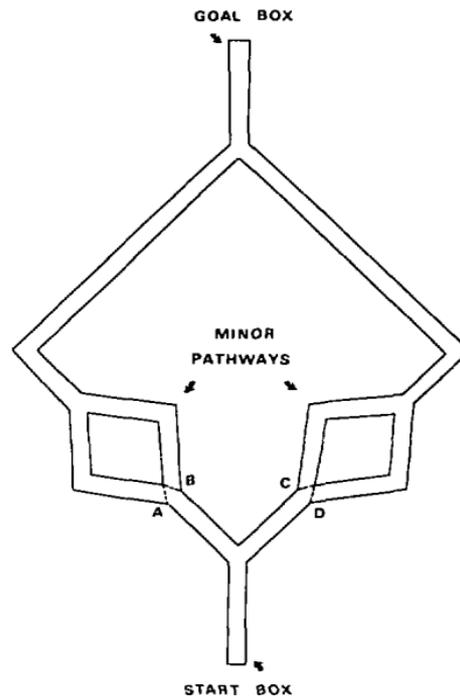
## 1. Une propriété commune à plusieurs espèces

### Études chez l'animal

Depuis les années 1970, un corpus d'études chez l'animal a confirmé une préférence naturelle des individus d'espèces distinctes vers des situations dans lesquelles il existe du choix. Le paradigme le plus utilisé dans ces études engage ces individus dans une tâche opposant des décisions avec possibilité de choisir librement entre plusieurs options et d'autres les restreignant à une seule action possible. L'élément central à toutes ces tâches repose sur une valeur strictement équivalente entre les différentes options. Autrement dit, quelle que soit l'action choisie, elle apporte un bénéfice final identique, par exemple la même récompense alimentaire. De cette manière, le seul paramètre pouvant expliquer un biais de décision est la présence ou non de choix.

Une des études pionnières reposant sur ce principe a été conduite par Voss & Homzie (1970) chez des rats se déplaçant dans un labyrinthe fermé à deux bras, conduisant tous deux au même point d'arrivée. La présence de choix était matérialisée dans un des deux bras, ces derniers se divisant une seconde fois en dessinant un cheminement strictement symétrique. Ce montage produit alors deux routes identiques amenant au même résultat (figure 1). Les rats pouvaient donc emprunter un bras "sans choix" conduisant directement à la récompense et un bras de "choix" dans lequel le rat pouvait librement emprunter l'une ou l'autre des deux directions. Les auteurs ont alors constaté qu'en moyenne, les rats empruntaient en majorité (59%) le bras dans lequel le choix était disponible alors même que cela ne leur apportait aucune récompense supplémentaire. Ils en ont conclu que

l'opportunité d'avoir le choix doit avoir une valeur propre, puisque la préférence des animaux est infléchie vers une situation économiquement équivalente à l'autre.



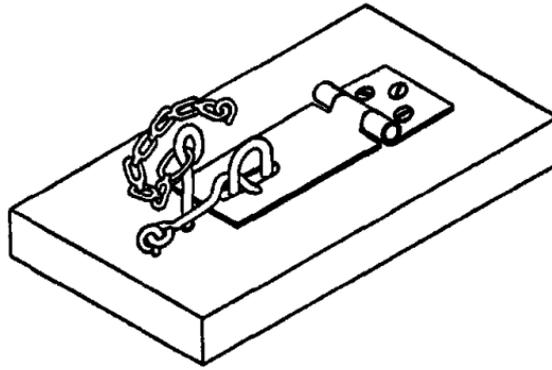
**Figure 1.** Labyrinthe de Voss & Homzie (1970). Les rats se déplaçaient librement d'une extrémité à l'autre pour obtenir une récompense alimentaire. Lors de sessions différentes, un des bras du labyrinthe A, B, C ou D était fermé créant ainsi une voie de "choix" où les deux bras pouvaient être empruntés librement, et une voie de "non choix" où une seule option était empruntable. Une récompense se trouvait à la sortie du labyrinthe (« goal box »).

Ce résultat a par la suite été reproduit chez d'autres espèces. Suivant le même principe d'une tâche en deux étapes, un biais pour les situations comportant du choix a par exemple été montré chez les primates non humains. Dans une étude réalisée par Suzuki en 1999 impliquant des primates *Macaca Fascicularis*, la présence de choix se reflétait par la possibilité de sélectionner deux images ayant des probabilités de récompense asymétriques, par exemple 80 et 50%. A l'inverse, l'option sans choix ne permettait de sélectionner qu'une seule image délivrant une récompense avec une probabilité aussi avantageuse que celle du côté choix, en l'occurrence 80%. Ainsi, les essais de choix sont désavantagés, puisque sa présence expose à l'erreur et donc au risque de sélectionner le stimulus le moins récompensant. La décision rationnelle permettant de maximiser les gains en minimisant les risques se trouve ainsi du côté du non-choix, aucune erreur ou défaut d'apprentissage des

récompenses n'étant possible. Tout comme cela avait été montré par Voss et Homzie, les primates de cette étude se sont engagés dans un comportement à l'encontre de cette stratégie visant à maximiser ses gains, et ont montré une nette préférence pour le choix.

Ce résultat a d'ailleurs été reproduit en utilisant un procédé plus complexe. Perdue et al. (2014) ont ainsi mis en évidence que le choix conférait un avantage pour s'engager dans tout type de situation motivante, n'impliquant pas forcément de récompense alimentaire, mais par exemple une activité ludique. Pour cela, les auteurs ont étudié le comportement de singes lorsqu'ils choisissaient entre plusieurs jeux pour lesquels ils avaient été entraînés au préalable. Afin de s'engager dans telle ou telle activité, les animaux avaient la possibilité de choisir par eux-mêmes, ou bien laisser l'ordinateur choisir pour eux. Là encore, les singes (rhésus et capucins) ont préféré des situations permettant de choisir librement. Ce résultat est d'autant plus marquant que le jeu sélectionné dans les essais dépourvus de choix tenait compte de la préférence de chaque individu : les singes préféraient alors choisir par eux-mêmes plutôt que de se laisser attribuer leur activité préférée, quand bien-même l'issue était identique.

L'ensemble de ces exemples suggèrent comme première explication que l'engagement volontaire dans une action afin d'obtenir une récompense serait par elle-même gratifiante, à l'inverse d'une situation octroyant passivement cette même récompense. Autrement dit, la valeur du choix pourrait trouver son origine dans ses propriétés instrumentales. Cette suggestion se rapproche en partie de ce que Harlow et al. (1950) ont qualifié d'"impulsion manipulatoire". Dans leur étude, des singes macaques rhésus manipulaient une boîte complexe ne pouvant s'ouvrir qu'en exécutant plusieurs gestes dans un ordre précis (tirer une goupille, puis un crochet et soulever une plaque mobile ; figure 2).

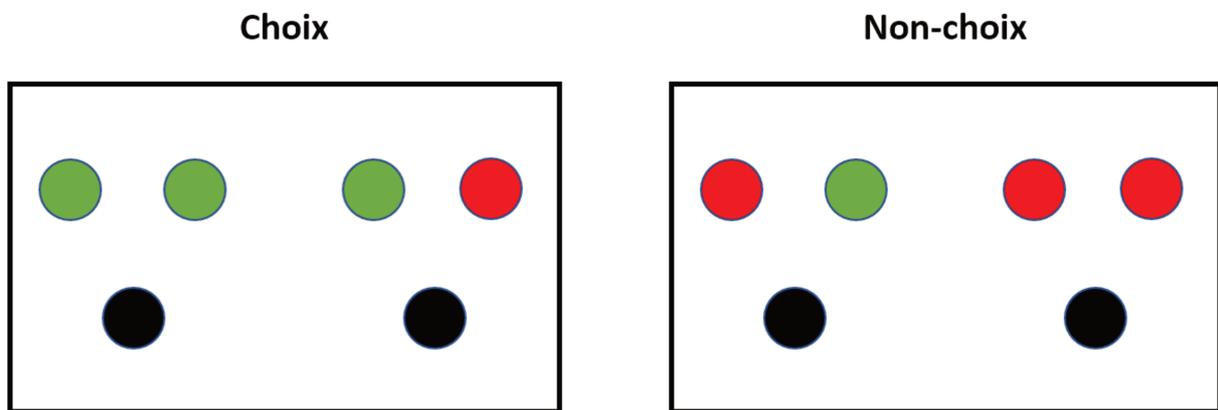


**Figure 2.** Schéma du dispositif utilisé par Harlow et al. (1950). Les singes devaient résoudre le puzzle en réalisant plusieurs actions consécutives.

Alors que les singes s'engageaient facilement dans la tâche et ouvraient avec facilité la boîte lorsqu'elle était vide, l'adjonction de nourriture venait contre toute attente diminuer les performances des animaux. « Aveuglés » par la présence de récompense alimentaire, ils essayaient alors de forcer l'ouverture du contenant par tous les moyens pour en extraire l'aliment. Si cela montre que les récompenses alimentaires peuvent être motivantes jusqu'à parasiter d'autres processus cognitifs, ce contraste comportemental suggère également que le simple fait de résoudre un problème au moyen de plusieurs manipulations était en lui-même motivant. A l'inverse, l'adjonction finale d'une récompense « extrinsèque » comme de la nourriture viendrait atténuer le caractère « intrinsèquement » gratifiant issu des manipulations de la boîte.

Enfin, cette préférence pour le choix n'a pas été mise en évidence que chez des mammifères. Une étude réalisée chez des pigeons a également affirmé une préférence constante pour les situations avec choix. Dans cette étude, des pigeons avaient la possibilité de récupérer une récompense alimentaire en touchant du bec un des quatre ronds colorés en vert ou rouge qui se trouvait devant eux (figure 3). Les ronds verts délivraient une récompense tandis que les ronds rouges n'en délivraient pas. Dans les essais choix, trois ronds verts étaient visibles tandis que dans les essais non-choix un seul l'était. Dans les deux cas, l'animal était sûr d'obtenir une récompense puisqu'elle lui était constamment délivrée lorsqu'il touchait un

rond vert. La différence entre les deux situations était alors la présence d'une alternative, c'est-à-dire de pouvoir choisir librement. Dans le cas de cette espèce, ce biais pour le choix questionne alors sur l'ancienneté phylogénétique de ce phénomène (Catania & Sagvolden, 1980). Cette étude soulève également une nouvelle caractéristique quant à la valeur du choix. Alors que l'étude de Harlow a mis en avant la dimension potentiellement gratifiante d'une tâche instrumentale, l'étude Catania semblerait avancer un attrait des individus pour la diversité. Autrement dit, les pigeons tendraient à plus s'engager dans les essais perçus comme ayant plus de variabilité, comme plusieurs stimulus identiques au lieu d'un seul.



**Figure 3.** Reproduction de la tâche de Catania et al. (1980). Les pigeons avaient la possibilité de toucher avec le bec l'un des quatre ronds colorés. Les ronds verts délivraient une récompense alors que les ronds rouges n'en accordaient pas.

En résumé, une préférence pour des situations permettant d'effectuer plusieurs choix semble être un comportement partagé par plusieurs espèces dont le système nerveux présente une large diversité évolutive. Cette universalité du choix s'observe également au travers de nombreuses situations, aussi bien pour récupérer une récompense basique comme de la nourriture ou s'engager dans des tâches plus complexes. Cela suggère alors que choisir confère une valeur pour les individus à s'attribuer par eux-mêmes leurs préférences. Plusieurs hypothèses mécanistiques émanent par ailleurs de ces études. Il est ainsi possible que les individus aient une appétence pour les situations ayant un certain degré de diversité des stimulus, quand bien même le résultat serait identique, ou alors parce qu'interagir

directement avec l'environnement, à l'inverse d'une récompense attribuée passivement, présenterait un certain degré de gratification. Ce constat a été exploré plus en détail chez les humains au travers de paradigmes plus complexes, permettant d'exposer plus précisément les composantes expliquant que le choix présente une valeur propre.

### Etudes princeps chez l'humain : l'illusion du choix

Plusieurs études chez l'humain ont ainsi montré une préférence pour les situations de choix mais ont également démontré que la conviction, même erronée, d'être en situation de choix était suffisante pour être valorisée, ce que certains auteurs appellent « l'illusion » du choix.

A titre d'exemple, les participants d'une expérience dans un contexte écologique se sont retrouvés à participer à une loterie organisée sur leur lieu de travail qui, à leur insu, était une étude expérimentale (Langer, 1975). Lorsqu'un participant venait acheter un ticket, l'expérimentateur déguisé en vendeur se comportait de deux manières différentes : soit il tendait la boîte de tickets au participant et le laissait choisir (condition de choix), soit il lui remettait lui-même le ticket (condition de non-choix). Quelques jours plus tard, l'expérimentateur demandait aux participants combien revendraient-ils leur ticket, feignant d'être en pénurie de ticket et que d'autres personnes voulaient jouer. Alors qu'ils avaient acheté leur ticket à \$1, les participants ayant préféré le choix le revendaient significativement plus cher (\$8,67) que ceux n'ayant pas eu le choix (\$1,96).

De même, Bown et al. (2003) ont rapporté un résultat similaire dans plusieurs groupes indépendants soumis à différents scénarios de vie courante (aller à la banque, choisir un bar...) en y ajoutant un "leurre" de choix. Ce leurre consistait en une proposition alternative qui n'apportait pas de bénéfice supplémentaire aux autres propositions, si ce n'est qu'elle permettait une situation de choix. Dans un des scénarios, les participants recevaient comme consigne de devoir se rendre en boîte de nuit, deux se trouvant au nord, et une au sud. Le

participant devait alors dire quelle voie il allait prendre entre les deux. Les trois options possibles étaient alors :

- A) un club bon marché mais avec de la mauvaise musique,
- B) un autre plus cher avec de la bonne musique,
- C) un troisième plus cher que les deux autres avec de la bonne musique.

Cette dernière option servait de « leurre » de choix : pour la moitié des participants il était couplé avec le club A, et pour l'autre moitié avec le club B. Alors que le choix C n'apportait aucun avantage de prix ou de qualité de musique, les participants présentaient un biais pour le choix. Ainsi, le pourcentage de choisir le club A passait de 36% à 45% lorsqu'il était proposé avec du choix alors que la sélection la plus favorable était le club B (sans choix).

Ces deux études apportent des éléments supplémentaires par rapport aux études animales. A l'inverse de l'impulsion manipulatoire de Harlow, le choix a ici un caractère perceptif et non instrumental et limite alors l'argument selon lequel le choix serait gratifiant par le simple fait qu'il engage l'individu dans une action. La perception du choix est alors suffisante pour entraîner une préférence. Elles ne permettent toutefois pas d'exclure l'hypothèse d'une préférence pour les situations choix induite par la variabilité des stimulus qu'elles sous-tendent, comme cela était le cas par exemple dans l'étude de Catania. L'interprétation couramment faite de ces résultats va cependant au-delà de cette dernière hypothèse, en mettant plus en avant la dimension de contrôle de l'individu sur son environnement, quand bien même celui-ci serait illusoire, comme dans les exemples sus-cités.

Le choix pourrait en effet être utile pour reprendre le contrôle dans des situations présentant un degré de dissonance cognitive (Festinger, 1957). Il faut comprendre par ce terme l'émergence d'une tension interne provoquée par des décisions contradictoires avec nos pensées ou nos croyances antérieures. Par le fait de choisir librement, il a été ainsi suggéré que les individus résolvent cette tension en surévaluant l'option choisie et en dénigrant

l'option rejetée. Ce phénomène, connu sous le nom du « paradigme du libre choix », suggère non pas que choisir aurait une valeur intrinsèque mais aurait une utilité pour limiter les situations dissonantes. L'existence d'une valeur propre du choix, au-delà des avantages qu'elle pourrait apporter, a toutefois fait l'objet de nombreux travaux théoriques reprenant le cadre princeps de la préférence pour le choix dans la notion de contrôle.

## 2. Modèles psychologiques initiaux

### Locus de contrôle et auto-efficacité

Plusieurs auteurs ont tenté de formuler les mécanismes inhérents à la préférence pour le choix en l'intégrant en priorité dans le concept général de perception et recherche de contrôle. Julian B. Rotter décrit par exemple l'existence d'un trait constitutif et quantifiable des individus à percevoir les événements du monde extérieur comme attribués à leurs propres actions. Cette première théorie du « locus de contrôle » a pour intérêt de se dédouaner du caractère plaisant ou non pour l'individu d'être en responsabilité de ses choix (Rotter, 1966), un étudiant pourrait par exemple recevoir une mauvaise note à un examen et estimer de façon variable que ce résultat relève de son propre fait.

Rotter insiste sur le caractère constitutif du locus de contrôle, c'est-à-dire son caractère invariable suivant les situations que l'individu traverse, dépendant plutôt de facteurs éducatifs, sociaux, ou en rapport avec des événements de vie passés. Ce trait se situe alors sur un continuum variant entre deux extrémités de locus de contrôle, du plus "interne" ("tout événement me concernant m'incombe") au plus "externe" ("ce qui doit arriver, arrivera").

Cette notion est à mettre en perspective avec celle d'auto-efficacité (Bandura, 1977), qui elle aussi explique la croyance d'un individu à être responsable des événements qu'il

traverse, mais insiste sur sa variabilité en fonction des expériences qu'il traverse. Ainsi dans une situation donnée, selon le degré d'expérience personnelle, d'apprentissage, ou des expériences d'autrui, un sujet peut développer une compétence, voire dépasser le caractère aversif d'échecs antérieurs. Il s'agit là d'une idée plus complexe incluant le caractère dynamique de l'acquisition de compétences au travers de facteurs motivationnels et que Bandura met au centre des mécanismes expliquant la perception qu'a un individu de ses propres actions sur la modification de son environnement à court terme.

Locus de contrôle et auto-efficacité semblent donc aborder des mécanismes similaires mais prenant place dans des échelles de temps et contextes différents. D'un côté se situe la première qui relève d'une influence "distale" de la perception de contrôle au travers généralement de la personnalité. Il s'agit donc d'un principe relativement stable quant aux différentes situations qu'un individu rencontre. A l'autre extrémité, Bandura souligne le caractère "proximal" de la perception de contrôle vis à vis des nouvelles situations rencontrées, c'est-à-dire pouvant évoluer en même temps que les compétences de l'individu s'enrichissent dans un domaine donné.

### Théorie de l'auto-détermination

Un modèle surplombant ces approches a été proposé en 1985 par Richard Ryan et Edward Deci dans leur théorie de l'auto-détermination. Leur théorie assimile à la fois le caractère généralisable (influence égale des traits constitutifs à l'individu sur l'ensemble des situations qu'il rencontre) et dynamique (rôle des expériences apprises ou compétences, variables au gré des situations) du besoin de contrôle chez l'humain. Les auteurs suggèrent que chez tout individu, il existe une tendance naturelle à rechercher des défis, de la nouveauté ou une opportunité d'apprendre, les poussant donc non pas à s'engager dans telle ou telle tâche pour leur finalité mais pour leur caractère intrinsèquement motivant. Ainsi, le choix devient également un facteur de motivation intrinsèque car il accroît l'autonomie du sujet (Ryan et Deci, 2000).

Selon Ryan et Deci, les situations intrinsèquement motivantes sont au même titre qu'un besoin alimentaire. Les individus retrouvent alors un équilibre homéostatique en "s'engageant dans une action pour sa satisfaction inhérente plutôt que ses éventuelles conséquences". C'est ainsi que Ryan et Deci définissent la motivation intrinsèque, par opposition à des situations apportant une satisfaction par des objets extérieurs. Ainsi, les situations de choix seraient préférables par elles-mêmes et non pour les avantages secondaires qu'elles procurent.

En considérant les mécanismes motivationnels comme centraux, la théorie de l'auto-détermination utilise un lexique s'apparentant à ceux de l'intelligence artificielle, en particulier de l'apprentissage par renforcement. Le principe fondamental de l'apprentissage par renforcement est qu'un agent - biologique ou artificiel - tend à maximiser les récompenses et minimiser les punitions (Sutton & Barto, 2018). Ces récompenses servent ensuite de signal pour mesurer la valeur subjective d'une situation ou d'une action future et aident ainsi l'individu à "y revenir pour en avoir plus" (Schultz, 2007). Cela revient à motiver ses actions en fonction de leur valeur subjective. Suivant cette définition, les situations amenant plus de contrôle pour un individu deviennent intrinsèquement motivantes. Leur exécution comporte donc une valeur appétitive au même titre que des récompenses plus classiques (dites extrinsèques) comme l'eau, la nourriture ou l'argent.

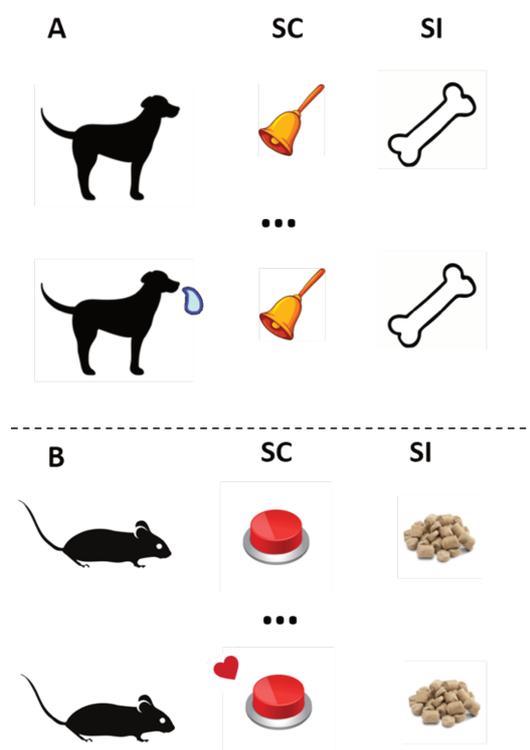
### **3. Apport des modèles d'apprentissage par renforcement**

Plusieurs auteurs ont suggéré que le choix avait sa propre valeur, indépendamment de celle obtenue par les récompenses « extrinsèques » finales. Dans une tâche probabiliste de choix d'images, Bobadilla-Suarez et al. (2017) ont montré que des sujets étaient prêts à payer pour avoir le choix. Dans leur paradigme, les sujets pouvaient soit choisir des images fournissant des récompenses par eux-mêmes, soit se faire aider par l'ordinateur moyennant un coût et

une probabilité de réussite donnée. En faisant varier les différents paramètres comme la probabilité d'être récompensé, le coût de l'aide par l'ordinateur ainsi que sa probabilité de donner la bonne réponse par ce dernier, il était possible de calculer pour chaque essai quel coût et quel bénéfice pouvait être obtenu en choisissant le choix ou en le refusant.

En moyenne, les participants payaient plus pour avoir le choix, au détriment d'une nouvelle fois d'une décision économiquement rationnelle. Il est possible d'avoir une lecture similaire des résultats comportementaux chez l'animal, notamment ceux obtenus chez les primates (Perdue et al., 2014; Suzuki, 1999). À récompense égale et quel que soit le type de récompense (activité ludique ou aliment), le fait de pouvoir choisir semble avoir plus de valeur pour les animaux.

D'autres arguments en faveur d'une valeur intrinsèque du choix ont été amenés par le champ de la psychologie computationnelle. En résumé, des données sont simulées au moyen de fonctions mathématiques dont les paramètres sont ensuite ajustés à des données expérimentales. Plusieurs de ces fonctions ont été construites pour modéliser un comportement généré par un réseau cérébral donné. Les plus connues et celles dont le niveau de preuve pour modéliser la prise de décision est le plus démonstratif concernent l'apprentissage par renforcement. Leur algorithme repose sur les principes prototypiques du conditionnement simple et instrumental (Pavlov, 1928; Skinner, 1963). Dans le premier cas (figure 4A), un stimulus neutre, dit conditionné (SC), précède suffisamment de fois une récompense (stimulus inconditionné, SI) pour que l'animal prédise la récompense dès présentation future du SC. Dans le second, le SC est une action à réaliser, l'animal apprenant alors que cette action peut être source de récompense (figure 4B).



**Figure 4.** Représentation schématique du conditionnement simple (A) et instrumental (B). Dans le premier cas (A), un stimulus neutre dit conditionné « SC » présenté suffisamment de fois avant une récompense (stimulus inconditionné « SI ») permettra à l’animal de prédire la récompense. Dans le second cas (B), une action conditionnée (SC), suivie d’une récompense (SI) suffisamment de fois permettra à l’animal de favoriser cette action dans l’attente d’une récompense.

Ces paradigmes sont au cœur de la théorie d’apprentissage par renforcement. (Sutton & Barto, 2018) et à l’origine d’une série d’algorithmes permettant de modéliser une réponse optimale en vue de maximiser ses gains à long terme. Ces fonctions ont été par la suite développées en ingénierie puis appliquées dans plusieurs domaines des neurosciences. La précision de ces modèles pour expliquer différents types de données expérimentales et leur facilité d’implémentation ont alors permis d’établir des prédictions solides et plusieurs fois reproduites sur les mécanismes de prise de décision et les systèmes cérébraux qui y sont impliqués (Niv, 2009).

## La théorie de l'erreur de prédiction

La plupart de ces modèles mettent en jeu un environnement simulé composé de différentes situations - ou états -  $s$  dans lesquels un agent peut obtenir une récompense  $r$  et ainsi prédire une valeur à long terme  $V$  au fur et à mesure des récompenses qu'il reçoit. Comme dans le conditionnement simple, il se peut que l'agent n'interagisse pas avec l'environnement et ne puisse que prédire une valeur propre à chaque état  $V(s)$ . Dans le cas d'un conditionnement instrumental, l'agent évalue alors la valeur propre  $V(s,a)$  de chaque couple état  $s$  - action  $a$  possible.

Pour cela un sujet doit :

- 1) évaluer si ses prédictions sont justes, c'est-à-dire s'il reçoit autant, plus, ou moins qu'espéré dans un état donné,
- 2) mettre à jour et modifier le modèle interne généré en fonction de cette différence - appelée erreur de prédiction -  $\delta$  lorsqu'il se trouvera à nouveau dans la même situation,
- 3) et enfin dans le cadre de tâches instrumentales, sélectionner l'action dont la valeur  $V(s,a)$  est la plus avantageuse.

L'équation ci-dessous résume les deux premiers processus, l'erreur de prédiction  $\delta$  permettant de mettre à jour la valeur  $V$  qu'un agent estime dans une situation donnée, pondérée par un paramètre  $\alpha$  (ou taux d'apprentissage) compris entre 0 et 1. Ce paramètre est dépendant des caractéristiques à la fois de l'agent (par exemple l'attention) et des stimulus (par exemple leur saillance) qu'il rencontre, et sert à moduler la proportion de l'erreur de prédiction qui sera ajoutée à la valeur mise à jour (Sutton & Barto, 2018).

Prenons l'exemple d'une personne se trouvant dans un café, à gauche se trouve la salle intérieure dans laquelle il a l'habitude d'aller tandis qu'en tournant à droite il peut accéder à la terrasse. Voyant qu'il y a du soleil, il décide de se diriger vers la droite et d'occuper une des chaises en terrasse. Alors qu'il n'avait jamais envisagé cette option auparavant, cette

personne pourra juger si cette place lui plaira à l'avenir - la valeur  $V(S_{terrasse}, a_{droite})$  sera réévaluée - en fonction de telle ou telle récompense  $r_{terrasse}$ , de ce qu'il s'imaginait initialement d'une place en terrasse -  $V(S_{terrasse}, a_{droite})$  -, et de comparer cette nouvelle estimation à ce qu'il connaît de sa place habituelle à l'intérieur  $V(S_{intérieur}, a_{gauche})$ .

$$V(S_{terrasse}, a_{droite}) \leftarrow V(S_{terrasse}, a_{droite}) + \alpha \times \delta$$

où

$$\delta = [r_{terrasse} + V(S_{terrasse}, a_{droite})] - V(S_{intérieur}, a_{gauche})$$

$$\alpha \in [0,1]$$

Ce type d'équation est dite *model free*. C'est-à-dire que l'agent n'apprend des situations associées aux actions uniquement de l'erreur de prédiction et des récompenses et non de l'environnement.

En d'autres termes, l'avantage de "s'installer en terrasse" tient uniquement compte de l'évaluation de la quantité de récompense sur le long terme que notre agent peut obtenir. Or, celle-ci peut être faible s'il n'a tenté l'expérience qu'une seule fois. En conséquence, pour que notre agent décide de s'installer dehors à l'avenir, il faudra que l'erreur de prédiction soit suffisamment ample et qu'elle soit répétée un nombre suffisant de fois pour que la valeur d'aller sur la terrasse devienne supérieure à celle de rester à l'intérieur. Il s'agit donc d'un système stable, permettant de fixer des habitudes sur le long terme, mais ayant pour corollaire son manque d'adaptabilité à des situations nouvelles.

A l'inverse, d'autres algorithmes permettent d'associer plus fortement la structure de l'environnement dans le calcul des valeurs subjectives prédites à long terme afin d'adapter rapidement la prise de décisions ("*model based*"). L'agent réalise alors une "carte" de

l'environnement et prend en compte les probabilités de se retrouver d'un état  $s$  à un autre  $s$  et d'en ajuster rapidement la valeur. Contrairement à l'algorithme "*model free*", ce modèle permet une heuristique rapide, flexible et adaptée à l'environnement. Dans le cas de notre agent, s'il se met à pleuvoir, le fait qu'il soit obligé de rentrer à l'intérieur ne dévaluera pas la place en terrasse, car il aura appris du modèle que cette situation est inhabituelle (Daw et al., 2011).

La distinction entre ces deux modèles est importante car elle implique des caractéristiques cognitives et comportementales ainsi que des réseaux neuronaux différents. Le premier fait appel à un raisonnement plus rigide, basé sur le long terme, imprimant des comportements relativement invariables pour un sujet donné, Le second conduit à plus de flexibilité et témoigne aussi de la nécessité de se représenter l'environnement avec précision pour pouvoir déduire les conséquences de ses propres actions de manière adaptée. Les deux modèles agissent très probablement de manière simultanée avec un poids différent de l'un ou l'autre en fonction des situations (Collins & Cockburn, 2020).

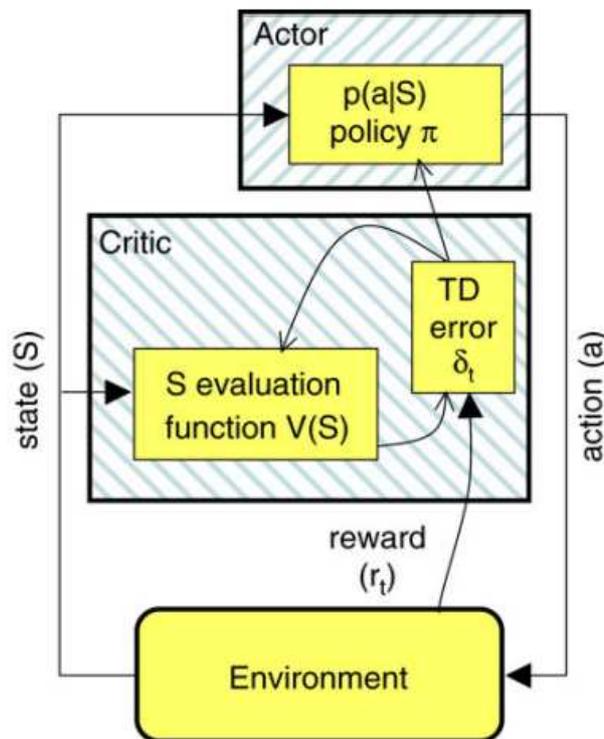
Ces deux modèles permettent d'ajuster les *valeurs* de tel ou tel évènement sur lequel l'agent appuie sa décision. La sélection des actions futures en fonction des valeurs calculées ci-dessus ("vais-je m'installer dorénavant en terrasse ?") dépend elle d'une autre famille de fonctions dont le rôle est d'optimiser les décisions sous la forme d'une probabilité conditionnelle. Étant présentement dans l'état  $s$ , quelle sera la probabilité de sélectionner  $a$  ou  $P(a/s)$ , ce qui dépendra des valeurs  $V(s,a)$  calculées en amont. Cette caractéristique est appelée dans le lexique de l'apprentissage par renforcement une police, couramment notée  $\pi$ .

Il existe donc deux fonctions distinctes, l'une permettant la prédiction des récompenses futures et des valeurs au long terme selon deux modèles, l'autre responsable de la sélection de l'action  $\pi$  que l'on peut assimiler à des systèmes neuronaux différents. L'interaction de

ces différentes actions et modes de prise de décision a fait l'objet de modélisations plus ou moins complexes.

### Le modèle Acteur-Critique

Un des modèles les plus couramment utilisés est l'Acteur-Critique. Le système Critique est responsable de l'évaluation des actions couplées aux états au moyen du calcul de l'erreur de prédiction. Le système Acteur est lui responsable de la prise de décision en appliquant une police qui prend compte des valeurs (figure 5). L'organisation de ce modèle en différents espaces a permis de les assimiler aux structures cérébrales présumées responsables de telle ou telle fonction (cf. partie II).

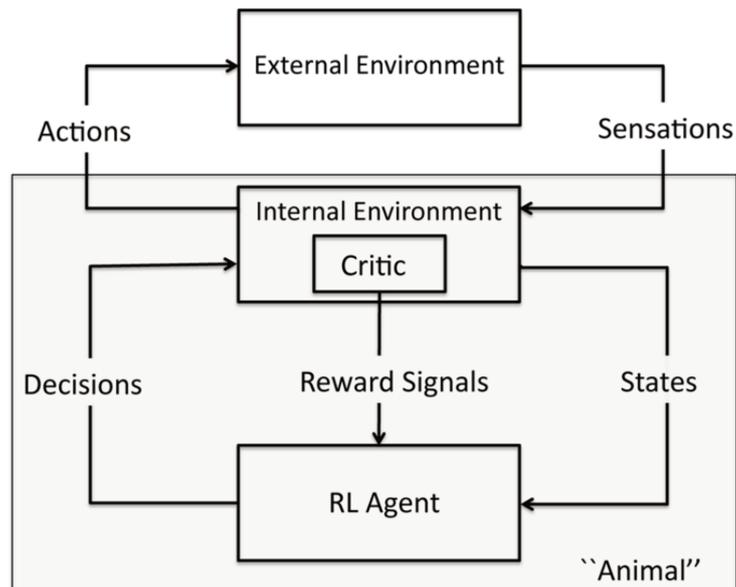


**Figure 5.** Le modèle Acteur-Critique : l'agent "critique" reçoit de l'environnement un signal de récompense qui lui sera utile pour calculer une erreur de prédiction  $\delta$  et mettre à jour la valeur  $V$  en fonction de l'état  $S$  dans lequel il se trouve. L'acteur reçoit également des informations de l'environnement sur l'état  $S$ , et les informations du Critique pour mettre à jour sa police  $\pi$  permettant la sélection de l'action  $a$ . (D'après Niv, 2009).

## Vers de nouveaux modèles des récompenses intrinsèques

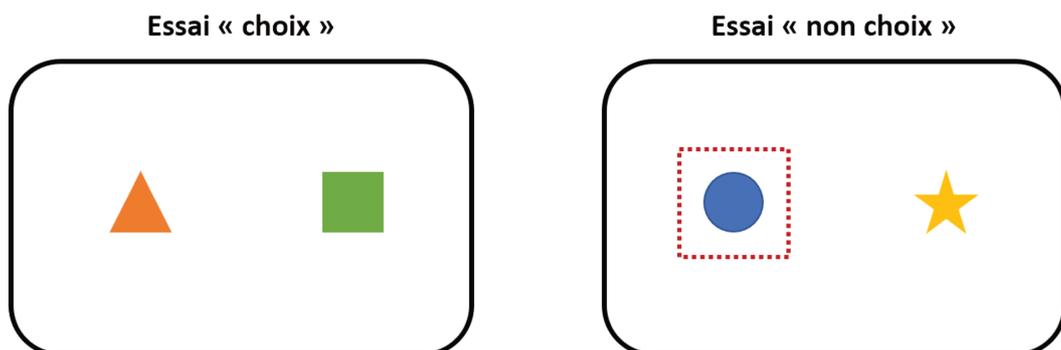
Ce modèle est l'un des plus utilisés pour expliquer des données de prise de décision mais implique communément des récompenses extrinsèques (Niv, 2009; Schultz, 2000). L'analogie avec des situations pourvoyeuses d'une récompense intrinsèque comme le choix est toutefois permise par la « théorie de la monnaie commune ». Cette théorie suggère qu'un système neuronal tendrait à renforcer la quête de tout stimulus plaisant, indépendamment de ses caractéristiques grâce à une évaluation commune des récompenses (Levy & Glimcher, 2012). L'individu peut ainsi comparer et hiérarchiser plusieurs situations ayant des enjeux différents afin de parvenir à une décision.

Certains auteurs ont alors proposé de transposer les modèles existants comme l'Acteur-Critique en tenant compte de l'existence de récompenses intrinsèques et en réajustant différents espaces du modèle (figure 6). Singh et al. (2010) proposent par exemple d'y ajouter un environnement interne dans lesquels les mécanismes en jeu dépendent uniquement de l'agent. Ils esquissent ensuite une mise à jour des équations de calcul de la valeur sur le long terme pour mieux tenir compte de cette nouvelle terminologie. Il s'agit de travaux théoriques visant à aider la réflexion sur de nouveaux modèles avec une complexité nouvelle mais n'ont pas été encore utilisés pour expliquer des données expérimentales. Ces modèles sont toutefois nécessaires car d'autres plus classiques ne suffisent pas pour expliquer de manière exhaustive les comportements intrinsèquement motivés.



**Figure 6.** Nouvelle représentation de l'Acteur-Critique prenant en compte les récompenses extrinsèques et intrinsèques. Ces dernières sont désormais traitées comme des "signaux internes" quelle que soit leur origine. L'environnement est également scindé pour tenir compte des interactions entre l'agent (décisions, état) et son environnement interne pourvoyeur de signal de récompense ainsi que de celles entre le module interne et externe de l'environnement (actions et sensations) (D'après Singh et al.,2010).

Deux études ont à ce jour proposé de modéliser le libre choix sur des données expérimentales en utilisant des algorithmes classiques. Elles proposent toutes deux une structure globale identique consistant à choisir entre deux stimulus délivrant une récompense de manière probabiliste mais proposant des essais dans lequel le sujet peut choisir entre les deux alors que dans d'autres le choix lui est imposé (Figure 7).



*Figure 7.* Exemple d'essais d'une tâche probabiliste faisant varier le choix et le non-choix entre les essais. Dans l'essai de gauche, le sujet peut sélectionner librement l'une ou l'autre image tandis que dans l'essai de droite, le design de la tâche lui impose une des deux images.

Dans l'étude de Cockburn et al. (2014), les participants apprenaient préalablement les contingences de récompense extrinsèque, certains avec du choix, d'autres sans, comme présenté dans la figure 7. Une fois que le participant avait appris les contingences, les stimuli étaient proposés par paire, mêlant parfois des stimuli choix et non choix dans une même paire. Le participant sélectionnait alors parmi les deux options celle qu'il pensait être la plus fréquemment récompensée. En utilisant cette structure, les auteurs émettent un fort a priori sur la plus-value du choix qui n'est exprimée ici qu'au moment de l'apprentissage des stimulus délivrant une récompense extrinsèque. Autrement dit, la valeur surajoutée du libre choix est "transférée" aux stimulus, et n'est pas identifiable en tant que tel, les participants ne pouvant choisir entre des essais de choix ou de non-choix. Ce modèle se rapproche alors plus des prédictions issues du paradigme du libre choix pour résoudre la dissonance cognitive (Brehm, 1956; Festinger, 1957)

L'hypothèse des auteurs était effectivement que choisir librement amplifie l'erreur de prédiction de la récompense lors de l'apprentissage, autrement dit que les récompenses finales étaient surévaluées quand elles étaient choisies par le sujet lui-même. Ils ont alors construit leur modèle en dissociant un taux d'apprentissage spécifique pour les stimulus choisis librement de celui pour les stimulus appris sans choix. Aussi bien les données comportementales non modélisées que les paramètres d'apprentissages du modèle ont mis en évidence un avantage du choix. En effet, leurs taux d'apprentissage étaient significativement plus élevés que ceux associés au "non-choix". Les sujets accordaient donc plus de valeur aux récompenses finales choisies que celles non choisies.

Prenant en compte également l'influence des contingences de récompenses entre les différentes paires de stimulus présentées, ces modèles ont également établi que le biais pour le choix apparaissait plus nettement lorsque les gains associés à un stimulus étaient plus élevés que les pertes. Autrement dit, les sujets accordaient d'autant plus de valeur aux

récompenses finales choisies que la récompense attribuée était élevée. Ce biais positif a également été mis en évidence par Chambon et al. (2019). En utilisant une tâche similaire à l'étude précédente, les taux d'apprentissage des modèles variaient entre les stimulus "choisis" comparés à ceux "non choisis". Une asymétrie des paramètres était également visible lorsqu'il était possible pour les sujets de voir la récompense du stimulus qu'ils n'avaient pas sélectionnée. Les valeurs des stimulus sélectionnés étaient alors d'autant plus élevées que les sujets savaient ce qu'ils avaient gagné mais également ce qu'ils avaient évité de perdre.

L'utilisation de la théorie de l'apprentissage par renforcement confère donc un avantage supplémentaire car il permet de mettre en évidence des comportements latents peu visibles sur les données descriptives. C'est ce qu'ont tenté de démontrer les deux études précédentes. Cette interprétation est en accord avec un avantage du choix qui, puisqu'il augmente le contrôle, valorise davantage la décision propre des sujets que celles non prises par eux-mêmes.

Bien que ces deux exemples aient permis d'apporter des éléments d'explication importants sur les avantages économiques et psychologiques du libre choix, ce dernier est évalué implicitement puisqu'il n'est visible qu'au moment de la sélection des stimulus délivrant la récompense finale. Ils renseignent donc sur les conséquences - avantageuses - de choisir librement pour obtenir plus de récompenses extrinsèques et non la valeur intrinsèque du choix à proprement parler. Cela implique alors de dissocier deux moments dans la prise de décision, à savoir 1) l'engagement du sujet dans un choix libre et 2) l'obtention d'une récompense extrinsèque. Comme vu précédemment, cette modélisation des valeurs intrinsèques est encore imparfaite.

Alors que plusieurs travaux ont tenté de modéliser sur le plan comportemental la préférence pour le choix, certains ont tenté d'explorer les structures cérébrales s'y rattachant. En se référant à l'implication de certains systèmes cérébraux dans le traitement des récompenses,

ces études se sont principalement focalisées sur le système dopaminergique et celui des ganglions de la base.

#### **4. Conclusion intermédiaire**

La préférence des situations comportant du choix est un caractère observable commun à plusieurs espèces animales et présente des avantages neuro-économiques qui semblent prévaloir sur l'obtention de récompenses extrinsèques. Pour cela, il a été suggéré d'attribuer au choix un caractère intrinsèquement récompensant. Confirmé par l'utilisation de modèles d'apprentissage par renforcement montrant que le libre choix influence les paramètres d'apprentissage de récompenses extrinsèques, la valeur propre du choix n'a pour autant pas été modélisée parfaitement par ces études antérieures. Au niveau cérébral, une vaste littérature implique le rôle de ces systèmes dans les processus d'apprentissage et dans la prise de décision. En se référant à la théorie d'une "monnaie commune" encodant tous types de récompenses, ce système serait alors impliqué dans le traitement des récompenses à la fois extrinsèques et intrinsèques.

## II. Hypothèses anatomiques de la préférence pour le choix

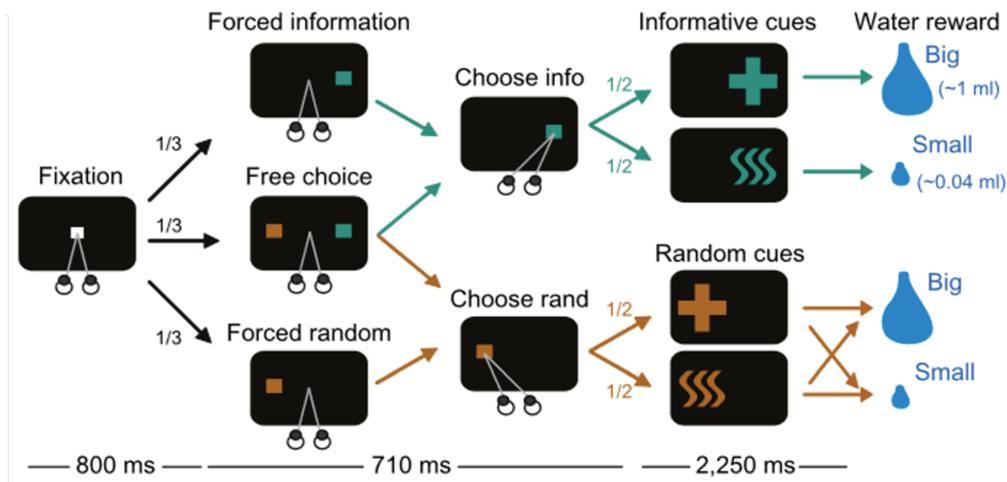
Plusieurs travaux ont permis de rattacher la préférence pour le choix à des structures cérébrales précises. Toutefois, celles-ci sont peu nombreuses et ont toutes impliqué des paradigmes permettant de quantifier indirectement la valeur du choix, au même titre que les études computationnelles précédentes.

Dans une de ces tâches, des volontaires sains devaient noter des destinations de voyage en fonction de ses préférences. Il leur était ensuite demandé de choisir entre deux destinations ayant une note proche, puis de noter à nouveau l'ensemble des destinations proposées. Premièrement, les résultats comportementaux ont montré que les participants donnaient une note supérieure aux destinations qu'ils avaient choisis par rapport à celles données à la phase initiale, suggérant une valeur ajoutée par le choix. Deuxièmement, l'analyse en IRM fonctionnelle lors de la présentation des destinations choisies montrait un signal accru dans le striatum, plus élevé d'ailleurs que le signal obtenu lorsque cette même destination était présentée avant d'avoir fait le choix (Sharot et al., 2010). L'implication du striatum a également été retrouvée dans d'autres études utilisant un paradigme similaire aux études computationnelles citées précédemment, c'est-à-dire opposant des essais comportant du choix par rapport à d'autres dans lequel le choix était forcé (Leotti & Delgado, 2014; K. S. Wang & Delgado, 2019)

Une étude électrophysiologique réalisée chez le singe a également mis en évidence l'implication des neurones dopaminergiques dans une tâche évaluant des récompenses intrinsèques. Il s'agissait d'une tâche probabiliste dans laquelle il était possible non pas de pouvoir faire un choix, mais plutôt d'avoir la possibilité d'obtenir une information supplémentaire. La tâche se déroulait en deux étapes ; dans la première, l'animal pouvait choisir s'il voulait avoir de l'information ou non, tandis que dans la seconde apparaissait aléatoirement une image qui délivrait ensuite une récompense. Dans chaque condition

(information ou sans information), deux images différentes pouvaient être affichées, différant dans leur quantité de récompense obtenue, l'une étant plus petite que l'autre.

Dans la condition information, la probabilité qu'une image entraîne une récompense était déterministe, c'est-à-dire que l'animal savait à 100% quelle récompense il allait obtenir, qu'elle soit petite ou grande. Dans la condition sans information, cette probabilité était aléatoire (50%) mais chaque image donnait également soit une grande soit une petite récompense. Dans les deux conditions, la décision d'attribuer une ou l'autre image était identique et aléatoire. Quelle que soit la condition information ou non information, la probabilité d'avoir une récompense était de 50% (Figure 8).



**Figure 8.** Déroulé de la tâche d'information. Des essais avec information, sans information, ou laissés à la décision de l'animal étaient proposés avec une fréquence identique. Dans la condition information, l'animal pouvait connaître à 100% quelle récompense il allait obtenir à l'étape suivant. Dans la condition sans information, chaque image délivrait une récompense de manière aléatoire. (D'après Bromberg-Martin & Hikosaka, 2009)

Les résultats de cette étude ont montré que premièrement, les primates préféraient la condition information de manière constante. Deuxièmement, l'enregistrement des neurones dopaminergiques lors de la présentation d'une image attribuant une récompense montrait une activité phasique accrue. Il a alors été suggéré le rôle du système

dopaminergique dans la préférence pour avoir de l'information, et comme corolaire, sur les récompenses intrinsèques.

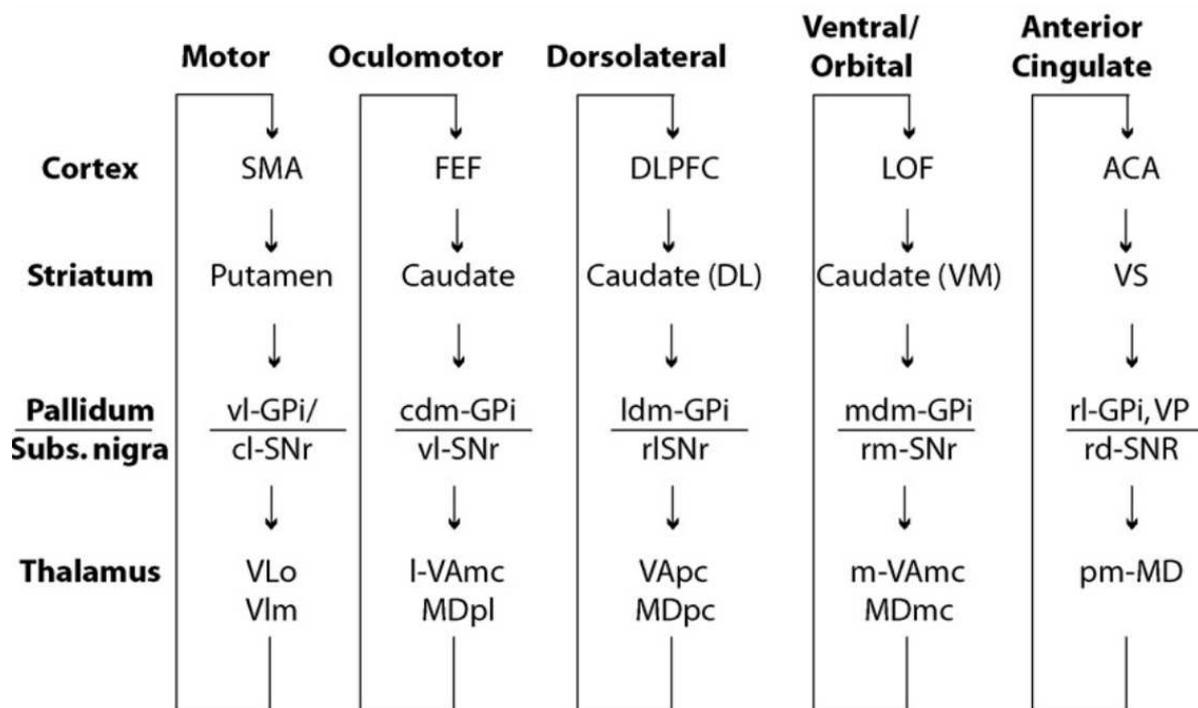
Ces résultats suggèrent donc que les structures dopaminergiques ainsi que le système des ganglions de la base auquel le striatum appartient seraient impliqués dans le traitement des récompenses intrinsèques. La suite de ce chapitre aura donc pour objet de décrire l'organisation fonctionnelle de ces structures.

## **1. Organisation générale des ganglions de la base**

Les ganglions de la base sont un ensemble de structures cérébrales profondes formant un réseau complexe d'interconnexions internes permettant de moduler les informations issus de la quasi-totalité du cortex cérébral (Selemon & Goldman-Rakic, 1985). On distingue quatre structures dans les ganglions de la base : le striatum (composé du putamen et du noyau caudé), le pallidum avec ses portions interne (GPi) et externe (GPe), le noyau subthalamique (NST), la substance noire avec sa partie réticulée (SNr) et compacte (SNc). Les informations provenant de la quasi-totalité du cortex cérébral se projettent sur deux structures d'entrée : le striatum et de façon plus restreinte sur le NST. Ces informations une fois traitées se dirigent vers deux structures de sortie qui sont le GPi et la SNr soit directement, soit indirectement, en faisant intervenir le GPe, et vont alors se projeter vers le thalamus, puis à nouveau sur le cortex préfrontal. Ces noyaux sont majoritairement composés de neurones inhibiteurs GABAergiques, à l'exception du NST, majoritairement composé de neurones glutamatergiques (son rôle est donc excitateur). La SNc contient les neurones dopaminergiques, ayant un rôle majeur modulateur du fonctionnement de l'ensemble du système des ganglions de la base. Une autre région dopaminergique, appelée aire tegmentale ventrale, disposée plus médialement par rapport à la SNc se projette préférentiellement sur le striatum ventral et le cortex préfrontal.

## Un système de voies parallèles traitant des modalités différentes

Plusieurs travaux expérimentaux chez les rongeurs et les primates non humains ont démontré l'existence de voies traitement des informations distinctes. Ces informations sont distribuées au sein de boucles grossièrement parallèles qui se maintiennent tout au long des différents noyaux des ganglions de la base. Le principe de ces travaux reposait sur le rôle présumé d'aires corticales spécifiques, et d'en disséquer les projections de relais en relais. Une des premières conclusions souligne l'importance des projections frontales comparées aux autres aires corticales. Au sein de régions distinctes du cortex frontal, Alexander et al. (1986) identifient cinq voies provenant de l'aire motrice supplémentaire, du champ oculomoteur frontal, du cortex préfrontal dorsal, du cortex orbitofrontal et du cortex cingulaire antérieur. Chacune de ces voies se projetant à des étages différents dans chaque noyau sur des territoires distincts et forment alors des boucles parallèles (figure 9).



**Figure 9.** Boucles fonctionnelles telles que décrites par Alexander et al. (1986) SMA, aire motrice supplémentaire ; vl-GPi, segment interne du globus pallidus ventrolatéral ; cl-SNr, substantia nigra pars reticulata caudolatérale ; VLo, noyau ventrolatéral du thalamus pars oralis ; VIm, noyau ventrolatéral du

thalamus pars medialis ; FEF, champs oculaires frontaux ; cdm-GPi, segment interne du globus pallidus caudodorsomedial ; vl-SNr, substantia nigra pars reticulata ventrolatérale ; l-VAmc, noyau antérieur ventral latéral du thalamus pars magnocellularis ; MDpl, sous-noyau parvocellulaire du noyau médian du thalamus ; ldm-GPi, segment interne du globus dorsomédial latéral ; rl-SNr, substantia nigra pars reticulata rostromatérale ; VApc, partie parvocellulaire du noyau thalamique antérieur ventral ; MDpc, partie parvocellulaire du noyau thalamique médiodorsal ; LOF, cortex orbitofrontal latéral ; Caudate (VM), noyau caudé ventromédial ; mdm-GPi, segment interne du globus pallidus dorsomédial médial ; rm-SNr, substantia nigra pars reticulata rostromédiale ; m-VAmc, noyau antérieur ventral médian du thalamus magnocellularis ; MDmc, sous-noyau magnocellulaire du noyau médio-dorsal du thalamus ; ACA, aire cingulaire antérieure ; VS, striatum ventral ; rl-GPi, segment interne du globus pallidus rostromatéral ; rd-SNr, substantia nigra pars reticulata rostromédiale ; pm-MD, noyau médio-dorsal postéro-médian du thalamus (D'après Grahn et al., 2008).

Cette hypothèse des cinq voies n'est cependant pas compatible avec une préservation de l'information au niveau de chaque noyau, une décroissance de volume substantielle existant entre chaque relais. A titre d'exemple, le striatum a un volume 21 fois plus élevé que le GPi et 63 fois plus élevé que le NST (Yelnik, 2002). Ce constat rend alors obligatoire une certaine convergence des informations d'un relais à l'autre. Il a donc été suggéré que les cinq voies convergeaient mais restaient relativement ségréguées au sein de trois territoires anato-mo-fonctionnels : un territoire sensori-moteur pour les boucles motrices et oculomotrices, un territoire associatif pour les boucles cognitives, et un territoire limbique pour les boucles émotionnelles et motivationnelles. Cette ségrégation en trois grands territoires anato-mo-fonctionnels au sein du complexe striato-pallidal a aussi été montrée chez l'homme à l'aide d'immuno-marquage et utilisée en clinique notamment en neurochirurgie fonctionnelle (Yelnik et al., 2007).

La ségrégation, même relative, a été très débattue ces dernières années. Au sein de ces trois grandes voies anato-mo-fonctionnelles, la résolution de ces réseaux parallèles ne s'effectue pas au niveau des grandes régions corticales, mais bien au sein de multiples sous-régions organisées elles-aussi en boucles. Au moyen de traçage rétrograde utilisant un vecteur viral trans-synaptique, il a été mis en évidence de multiples régions distinctes au sein du GPi et de la SNr (Middleton & Strick, 2002). Ces multiples sous-régions conservent toutefois une

certaine répartition topographique. Ainsi, deux régions corticales contiguës se projettent dans le GPi vers des territoires voisins mais leurs frontières sont floues. Par exemple, certains neurones du striatum provenant des régions dorsolatérales (« motrices ») se projettent vers des parties associatives de la SNr (François, Yelnik, Percheron, & Fénelon, 1994). Il a été également mis en évidence des collatérales axonales étendues, notamment au sein de neurones striato-pallidaux, suggérant que des neurones striataux situés dans un territoire peuvent réaliser des connexions de manière non ségréguée et dans plusieurs régions fonctionnellement distinctes du GPi (Parent et al., 2000). Plutôt qu'une ségrégation en trois territoires anatomo-fonctionnels, il existe un gradient de connectivité, comme retrouvé par un marqueur immunohistochimique (la calbindine), intensément exprimé dans les territoires limbiques. Dans les autres territoires, ce marqueur s'atténue dans les territoires associatifs pour devenir absent du territoire sensorimoteur du complexe striato-pallidal du singe et de l'homme (François, Yelnik, Percheron, & Tandé', 1994; Karachi et al., 2002).

Le traitement de l'information par les ganglions de la base est donc multimodal au sein de voies relativement ségréguées mais dont les frontières entre les trois grands territoires anatomo-fonctionnels sont floues au sein de chaque noyau permettant un traitement complexe et intégré de l'information.

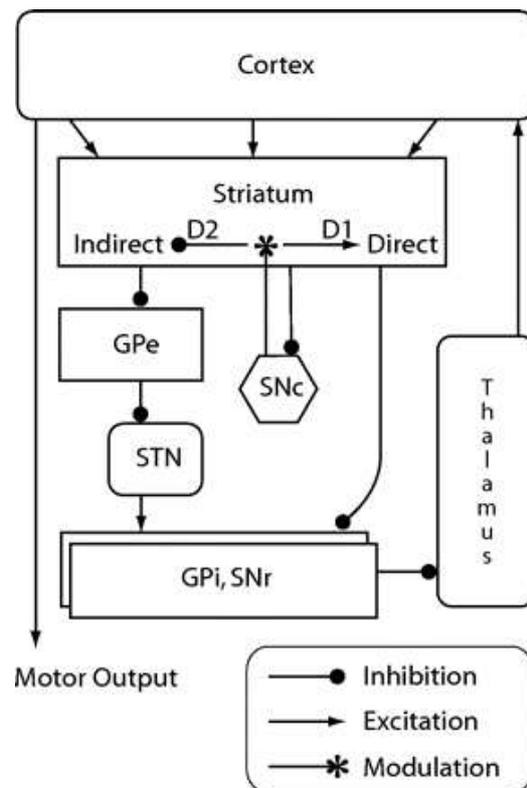
### Contrôle de l'action : rôle des connexions internes des ganglions de la base

Différents travaux de la fin des années 1980 ont permis d'établir un modèle explicatif du fonctionnement des ganglions de la base. Ce modèle initial décrit par Albin décrit deux voies de polarité opposées :

- La voie directe est activatrice. Les neurones du striatum se projettent sur le GPi et la SNr.

- La voie indirecte est inhibitrice. Les neurones du striatum se projettent sur le GPe, qui à leur tour se projettent sur le NST, puis sur le GPi et la SNr.

Ces deux voies sont modulées par les neurones dopaminergiques de la SNc se projetant sur le striatum, via des récepteurs D1 (activateurs) pour la voie directe, et D2 (inhibiteur) pour la voie indirecte (figure 10.).



**Figure 10.** Représentation simplifiée du modèle d'Albin. (D'après Moyer & Danish, 2007)

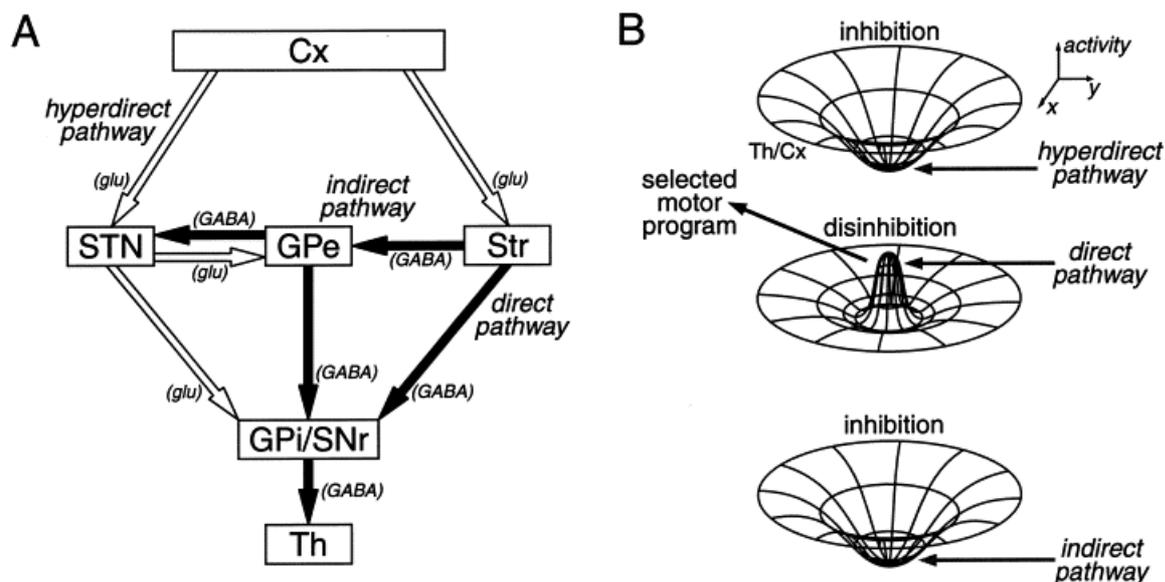
Se basant sur plusieurs travaux électrophysiologiques d'enregistrement unitaire de neurones des ganglions de la base (DeLong, 1972; Georgopoulos et al., 1983), mais également sur les observations de patients ayant des lésions des ganglions de la base, le rôle de ce système a été considéré historiquement comme impliqué dans le contrôle du mouvement.

Des études électrophysiologiques chez l'animal ciblant ces différentes structures ont alors tenté d'affiner le rôle de chacune de ces voies dans la préparation du mouvement. Il a ainsi

été constaté que l'activité du GPe et du GPi intervenait dans des paramètres cinétiques et d'amplitude du mouvement (Georgopoulos et al., 1983), suggérant alors un rôle de la voie indirecte de « pilotage » du geste tout le long de celui-ci.

Une hypothèse alternative s'est basée sur l'activité tonique au repos du GPi. Il est alors proposé que l'ensemble de la voie indirecte est dans un état basal inhibiteur, qui ne se lèverait que pour un comportement désiré. Les ganglions de la base permettraient alors d'amplifier l'action voulue par la voie directe et simultanément inhiber des actions parasites par la voie indirecte (Mink, 1996). Cette dernière notion est toutefois problématique compte tenu de l'asymétrie de réponse attendue entre les voies directes et indirectes. En effet, la différence en nombre de relais synaptiques entre ces deux voies induit une action différée de la voie indirecte, ce qui ne permet pas un contrôle suffisant rapide pour sélectionner de façon optimale des actions voulues et inhiber des actions parasites.

A la fin des années 1990, Nambu propose un nouveau modèle permettant d'intégrer la voie cortico-subthalamique qui avait été décrite récemment par des travaux d'anatomie. Nambu constate en enregistrant les neurones du NST et du GPi chez le primate, qu'une stimulation corticale induisait une réponse brève et précoce du NST ainsi que du GPi de manière différée. Il décrit une voie corticale « hyperdirecte », ne se projetant pas sur le striatum. Il propose alors une nouvelle version du fonctionnement de ces trois voies dans le contrôle de l'action : l'activation de la voie hyperdirecte permettrait l'inhibition rapide de schémas moteurs interférant, permettant de faciliter l'émergence de l'action voulue de manière précoce et flexible par la voie directe. La voie indirecte interviendrait de manière différée en permettant de mettre fin à l'action (Nambu et al., 2000).



**Figure 11. A.** Schéma des trois voies d'après le modèle de Nambu. Les flèches blanches et noires représentent les projections glutamatergiques (glu) excitatrices et GABAergiques (GABA) inhibitrices, respectivement. Cx, cortex cérébral ; GPe, segment externe du globus pallidus ; GPi, segment interne du globus pallidus ; SNr, substantia nigra pars reticulata ; STN, noyau subthalamique ; Str, striatum ; Th, thalamus. **B.** Diagramme schématique expliquant le changement d'activité dans le thalamus et/ou le cortex (Th/Cx) suite aux entrées séquentielles par les voies cortico-subthalamo-pallidales (milieu) et cortico-striato-pallidales directes (bas). (D'après Nambu et al., 2002).

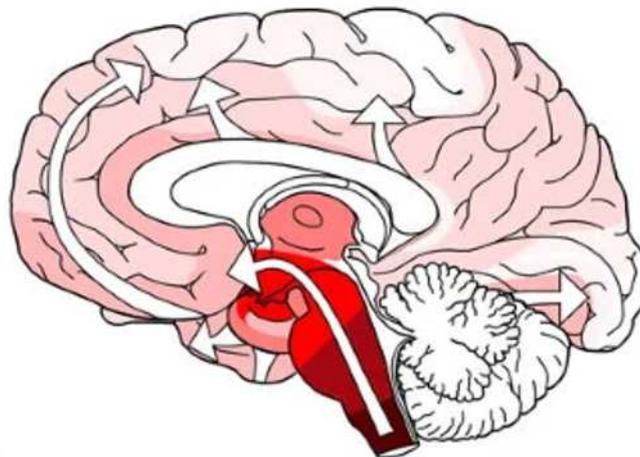
La dopamine étant un neuromodulateur, le système dopaminergique influence l'ensemble du fonctionnement des ganglions de la base. En son absence, les perturbations des différentes voies conduisent à une hyperactivité du NST, de la SNr et du GPi, et donc une inhibition majeure des informations traitées. Compte-tenu de la déplétion massive de dopamine dans la maladie de Parkinson, cette maladie représente un modèle d'étude prototypique des ganglions de la base.

## 2. La maladie de Parkinson comme modèle de compréhension des ganglions de la base

La maladie de Parkinson est la seconde maladie neurodégénérative la plus fréquente dans les pays à haut revenu. En 2017, la prévalence était estimée à 14 pour 100 000 personnes

dans la population totale, et 160 pour 100 000 chez les sujets âgés de plus de 65 ans (Kaplin & Montel Williams, 2019). En France, le nombre de patients traités pour une maladie de Parkinson était estimé à environ 160 000 personnes en 2015 (Santé-Publique France). Il s'agirait de la deuxième cause de handicap neurologique après les accidents vasculaires cérébraux.

Les mécanismes lésionnels provoquant la mort des neurones dopaminergiques sont complexes. La principale hypothèse met en cause le rôle de l'alpha-synucléine, une protéine neuronale produite physiologiquement. Son accumulation dans différentes structures du cerveau est considérée comme un des facteurs lésionnels de la maladie de Parkinson (Spillantini et al., 1997). Le pattern de progression des lésions est relativement stéréotypé, débutant probablement dans le tronc cérébral et s'étendant avec les années vers le cortex (Braak et al., 2003).



**Figure 12.** Représentation de l'évolution des lésions de la maladie de Parkinson, débutant dans le tronc cérébral dans les stades les plus précoces (rouge foncé), pour diffuser à l'ensemble des structures cérébrales au fil des années. Dans les stades les plus avancés, les lésions sont également présentes dans le cortex. (D'après Braak et al., 2003).

Les signes moteur (akinésie, tremblement, rigidité) sont les plus connus et apparaissent lorsque les lésions s'étendent vers le mésencéphale et atteignent les neurones dopaminergiques de la substance noire compacte (SNc ; Damier et al., 1999) et de manière plus hétérogène ceux de l'aire tegmentale ventrale (Alberico et al., 2015). La phase d'état

est la période caractéristique pendant laquelle les principaux signes de la pathologie sont visibles. On distingue principalement un retard à l'initiation et une diminution de l'amplitude du mouvement ce qui occasionne une lenteur invalidante. Les autres signes les plus caractéristiques sont le tremblement de repos et la rigidité.

Apparaissent également tôt dans la maladie des symptômes non moteurs. Ainsi, les patients peuvent se plaindre dès les premières années de la maladie de difficultés mnésiques, ou de trouble de l'humeur (Pont-Sunyer et al., 2015).

Ces symptômes peuvent être atténués en substituant la perte dopaminergique par un traitement médicamenteux (Fox et al., 2018). Ces traitements peuvent être regroupés en deux grandes classes que sont les agonistes de la dopamine et la L-DOPA. Le premier a une affinité plus importante pour les récepteurs D2, le second est un précurseur de la dopamine, métabolisé en dopamine dans le système nerveux central. Leur effet sur la motricité est démontré depuis des décennies, mais peut être pourvoyeur d'effets comportementaux, comme une sensibilité aux récompenses accrue (Voon et al., 2009). L'apparition différée de fluctuations des symptômes au cours de la journée nécessite la prise régulière du traitement médicamenteux. Dans des situations où ce contrôle n'est pas obtenu de manière satisfaisante par les traitements oraux, il est possible d'avoir recours à des thérapies délivrées de manière continue, parmi lesquelles la stimulation cérébrale profonde du NST.

### La stimulation cérébrale profonde du NST

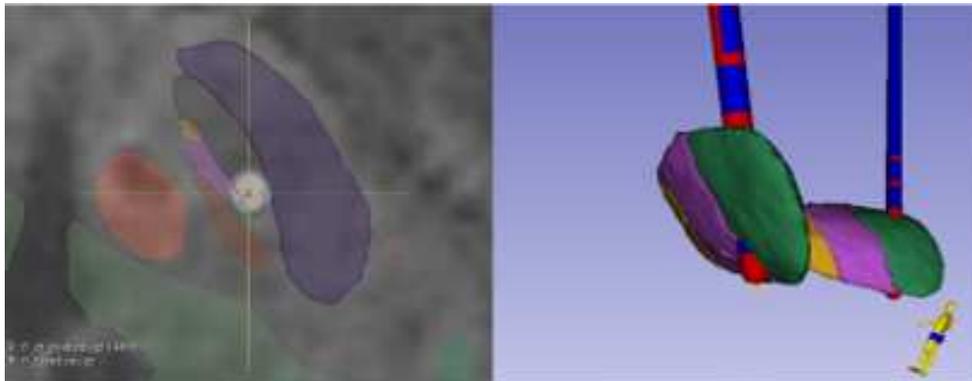
Des études pionnières chez le primate ont montré que les lésions, puis la stimulation à haute fréquence du NST chez le singe MPTP, provoque une diminution nette des symptômes moteurs (Benazzouz et al., 1993; Bergman et al., 1990). Le NST a alors été considéré comme une cible de choix pour la neurostimulation chez les patients parkinsoniens. En modulant l'activité du NST par un courant électrique à haute fréquence, la neurostimulation corrigerait l'hyperactivité du NST et améliorerait la balance excitation / inhibition des

structures de sortie des ganglions de la base. Sans que l'on ne connaisse exactement les mécanismes d'action, la stimulation cérébrale profonde (SCP) du NST améliore de façon majeure les symptômes dopa-sensibles des patients atteints de maladie de Parkinson (Limousin et al., 1998).

Les résultats similaires entre les études lésionnelles chez l'animal et les effets de la stimulation à haute fréquence ont conduit à penser que la stimulation du NST était principalement inhibitrice. Des travaux en électrophysiologie ayant enregistré les structures en aval du NST pendant sa stimulation ont d'ailleurs corroboré cette hypothèse, puisqu'une diminution simultanée de l'activité GPi était observée (Boraud et al., 1996). D'autres en revanche ont observé des patterns d'activation du GPi lors de la stimulation (Hashimoto et al., 2003). Une des hypothèses avancées reposerait sur un effet inhibiteur du courant délivré se propageant sur les fibres de projections environnantes (provenant du GPe et du striatum, inhibiteurs). Une autre explication a été avancée par McIntyre et al., (2004), suggérant une action inhibitrice sur les corps cellulaires mais une action activatrice sur les axones. Cette hypothèse a été explorée par un modèle mathématique simulant de manière exhaustive des neurones et le champ électrique émis par une électrode de stimulation. La stimulation électrique agirait alors plutôt comme modulateur de l'activité du NST et, outre sa position au sein du NST, la variabilité des effets produits par la stimulation dépendrait donc également du groupe axonal activé localement et de ses projections.

En pratique clinique, cette technique s'est développée dans de nombreuses équipes et est actuellement un traitement de routine ayant concerné plusieurs centaines de patients dans le monde (Deuschl et al., 2006; Welter et al., 2014). La tolérance cognitive de ce traitement a été sujet de plusieurs études, concluant globalement à une relative préservation des grandes fonctions cognitives, sans handicap majeur surajouté par la chirurgie (Ardouin et al., 1999; Jahanshahi et al., 2000). Toutefois, de nombreux tableaux cognitifs et neuropsychiatriques retrouvés chez une proportion de patients ont conduit à questionner l'impact de la stimulation du NST. En effet, comme les autres structures des ganglions de la

base, le NST présente une structure interne composée de territoires fonctionnels, différentes si on considère les afférences pallidales externes ou bien les afférences corticales. Cette double organisation en sous-territoires selon le type d'afférence vers le NST est d'ailleurs observable de manière distincte dans les différents travaux anatomiques réalisés sur ce sujet. Alors que les fibres limbiques, motrices et associatives provenant du GPe se répartissent sur des territoires du NST facilement délimités, celles provenant du cortex cérébral se projettent plutôt selon un gradient sur l'axe postéro-antérieur, sans qu'une délimitation aussi précise puisse être réalisée. Ces observations quant à la structure anatomique du NST questionnent dans tous les cas son implication dans des fonctions non motrices. Le prochain chapitre aura pour objet de définir plus précisément le NST sur le plan anatomique et fonctionnel, avec une emphase sur la voie cortico-subthalamique, les particularités de sa délimitation en sous-territoires ainsi que les conséquences majeures de sa modulation par stimulation chirurgicale sur le comportement humain.



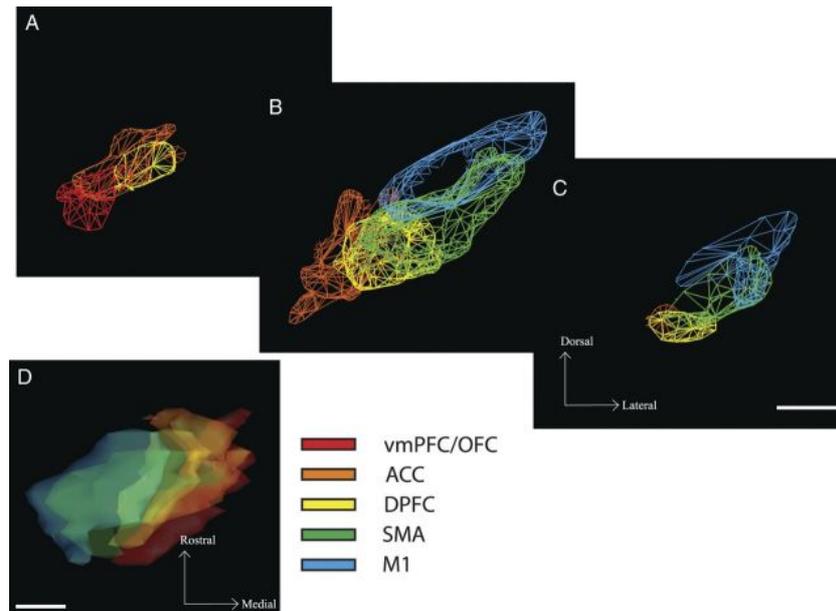
**Figure 13.** Visualisation tridimensionnelle des électrodes de stimulation au sein du NST, délimité selon le sous-territoires définis par ses afférences pallidales externes. Plan axial (gauche). Vue globale (droite). Vert : territoire sensorimoteur. Violet : territoire associatif. Jaune : territoire limbique

### 3. Le noyau subthalamique : une voie d'entrée majeure

Le NST est une petite structure (3 x 5 x 13 mm, 200 mm<sup>3</sup> chez l'humain) ayant une place stratégique au sein des ganglions de la base. Compte tenu des projections corticales directes

monosynaptiques qu'il reçoit, le NST module rapidement le fonctionnement des sorties des ganglions de la base.

Plusieurs études anatomiques chez le primate non humain ont permis de mettre en évidence son organisation en trois grands territoires anatomo-fonctionnels (sensorimoteurs, limbiques et associatifs) si l'on considère les projections pallidales externes (Karachi et al., 2005). En ce qui concerne les projections corticales, le NST reçoit sur toute son étendue des afférences motrices et pré-motrices (Nambu et al., 2000). Des projections du cortex cingulaire antérieur vers le pôle antérieur et médian ont aussi été mises en évidence tandis qu'une seule étude a montré une projection en provenance du cortex préfrontal dorsolatéral mais sans possibilité de la quantifier. Cette même étude a montré que les projections limbiques occupaient certes une petite place au pôle antérieur du NST mais surtout qu'elle se dirigeait médialement, dépassant largement les limites du noyau lui-même (Haynes & Haber, 2013; Karachi et al., 2005). Cette observation a ainsi permis de démontrer que le NST n'est pas un noyau dit « fermé » mais qu'il présente au contraire en son bord antérieur et médian une continuité avec les structures adjacentes. Ces structures médiales sont communément regroupées sous le terme de région subthalamique médiale et triangle de Sano, appelée anciennement hypothalamus postérieur (Sano, 1962). Elles reçoivent un contingent important de fibres corticales provenant de régions limbiques, comme cela a été récemment démontré chez l'humain *in vivo* lors d'une étude de tractographie de la voie cortico-subthalamique.



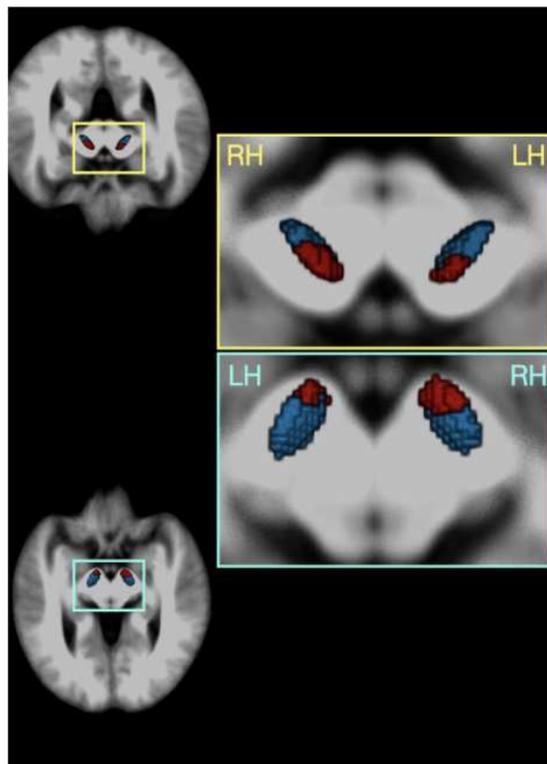
**Figure 14.** Répartition des fibres corticosubthalamiques. A-C : vue coronale, D : vue supérieure. vmPFC : ventromedian prefrontal cortex (limbique), OFC : orbitofrontal cortex (limbique), ACC : anterior cingulate cortex (limbique), DPFC : dorsal prefrontal cortex (associatif), SMA : supplementary motor areas (moteur), M1 : primary motor cortex (moteur). (D'après Haynes & Haber, 2013)

En utilisant des séquences d'IRM de diffusion à haut champ et avec une bonne résolution, Temiz et al. (2019) ont ainsi mis en évidence que les fibres provenant des cortex sensori-moteurs se projettent sur la région dorsolatérale du NST et que les fibres limbiques se projettent sur l'extrémité antérieure et ventromédiale du noyau. Comme observé chez le primate, les frontières entre ces régions sont mal définies, les projections limbiques se distribuant le long d'un gradient décroissant de l'avant vers l'arrière et inversement pour les projections sensorimotrices qui projettent sur l'ensemble du noyau. Les auteurs ont par ailleurs confirmé l'existence d'une connectivité en provenance du cortex préfrontal dorsolatéral comme cela avait été décrit chez le primate (Haynes & Haber, 2013) bien que ces fibres « associatives » ne soient représentées que par un contingent minime de fibres au sein du NST.

Au vu de ces résultats, il semblerait que la voie hyperdirecte se projette sur le NST sur deux régions principales, l'une antérieure majoritairement connectée à des régions préfrontales et limbiques, et une large région sensorimotrice postérieure. Cette hypothèse a par ailleurs

été étudiée récemment en IRM de diffusion, en utilisant une analyse en cluster de connectivité (Temiz et al., 2022).

L'ensemble de ces constatations apportent une connaissance supplémentaire sur l'organisation de la voie hyperdirecte chez l'humain. Comme décrit précédemment, il a été mis en évidence une connectivité importante des régions limbiques corticales sur la petite région médiane du NST et surtout vers la région subthalamique médiane adjacente s'étendant au-delà du NST jusqu'au troisième ventricule. Le pôle antérieur et médian du NST apparaît donc comme une entité relativement à part au sein du noyau, recevant des afférences limbiques corticales et du pallidum externe, mais aussi spécifiquement des afférences dopaminergiques et sérotoninergiques (Lavoie & Parent, 1990). Cela renforce alors l'hypothèse affirmant que le NST présente un contingent "ouvert" à son bord médial en continuité avec la région subthalamique médiane, hautement impliquée dans le traitement de fonctions émotionnelles.



**Figure 15.** Représentation en 3D du NST provenant d'une analyse en cluster ayant identifié une sous-région antérieure connectée au cortex préfrontal (rouge) et une sous-région postérieure connectée majoritairement au cortex sensorimoteur (bleue ; d'après Temiz et al., 2022)

Au final, l'absence de limite nette entre les différents territoires anatomo-fonctionnels et le haut degré de ségrégation volumique des fibres corticales sur le NST suggèrent une très grande convergence des informations cognitives, émotionnelles et motrices, et fait donc du NST un noyau fortement intégratif.

### Fonctions du NST : apport des études lésionnelles

Compte-tenu de la thématique comportementale mise en avant dans cette thèse, la description fonctionnelle du NST présentera brièvement son rôle dans la motricité, mais sera axée en priorité sur des aspects non moteurs.

#### Implication dans le contrôle du mouvement

Historiquement, les premières descriptions entre la fin du XIX<sup>ème</sup> et le début du XX<sup>ème</sup> siècle de ce qui a été nommé *le Corpus Luysii* suggéraient une implication dans le contrôle moteur proximal, une lésion de ce dernier entraînant chez les patients une hémichorée controlatérale violente. Le terme d'hémiballisme sera ensuite utilisé pour désigner spécifiquement ces mouvements stéréotypés proximaux et amples (Académie nationale de médecine, 2003). Plusieurs observations cliniques provenant majoritairement de patients présentant un infarctus circonscrit, et expérimentales après lésions chez l'animal ont également confirmé cette observation (Postuma & Lang, 2003). Ainsi l'injection de bicuculline (un antagoniste GABA) dans le NST dorsolatéral (territoire sensorimoteur) provoque un hémiballisme réversible chez le singe (Crossman et al., 1984), confirmant le rôle du NST dans le contrôle proximal de la motricité. Les effets de l'injection d'un antagoniste GABA semblent paradoxaux du fait de son action potentiellement activatrice du NST, il a été plutôt suggéré que l'importance des doses injectées aurait un effet de dépolarisation de tous les neurones du NST.

## Implication dans le contrôle de l'action et la sélection des comportements

Les données de la littérature ont montré que des lésions expérimentales du NST n'induisent pas seulement des altérations du mouvement lui-même mais plutôt sa préparation. Ainsi chez le rat, des lésions bilatérales (Baunez & Robbins, 1997) ainsi que la stimulation à haute fréquence du NST (Baunez et al., 2007; Desbonnet et al., 2004) altèrent certains processus attentionnels et d'inhibition. Les micro-injections de bicuculline ou de muscimol (un agoniste GABA) au sein du NST antéro-médial entraînent des troubles de l'attention et une hyperactivité chez le singe, sans altération du mouvement en tant que tel. Encore plus antérieur dans le NST, il est aussi possible de provoquer des stéréotypies de toilettage, témoignant du contrôle complexe qu'exerce le NST dans la sélection de comportements automatisés au sens large (Karachi et al., 2009). Ces travaux suggèrent que le NST est impliqué dans la programmation de l'action volontaire.

Les descriptions historiques cliniques de lésions de la région du NST rapportent en effet qu'en plus d'un hémiballisme, les patients peuvent présenter un comportement désinhibé ou impulsif, traduisant une altération du contrôle de l'action. L'impulsivité est un phénomène complexe dont les manifestations sont multiples. Un comportement peut être défini comme impulsif si par exemple l'individu ne parvient pas à inhiber une action déjà initiée, ou alors à différer une action immédiate, ou encore de sélectionner convenablement entre deux situations conflictuelles laquelle est la plus optimale. Différents paradigmes expérimentaux permettent de mettre en avant un type donné d'impulsivité avec différents degrés de complexité cognitive. Deux types de déficits sont principalement observés : il peut s'agir de réponses prématurées, traduisant l'incapacité de l'individu à refreiner une décision prématurée, ou alors de persévérations, pouvant être définies par l'incapacité de l'individu de se désengager d'une action et ainsi d'y mettre un terme lorsque cela est approprié. Des lésions réversibles (chimiques ou électriques), ou des lésions simultanées de plusieurs systèmes (comme le système dopaminergique et celui des ganglions de la base) ont été

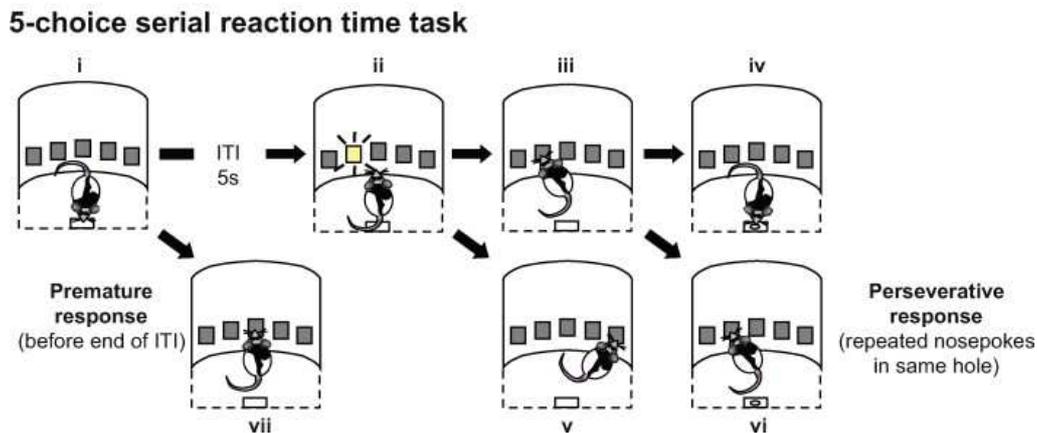
réalisées chez le rongeur utilisé comme modèle pour ce type de paradigme, permettant de proposer des hypothèses mécanistiques.

Ainsi, dans une tâche de temps de réaction dans laquelle des rats devaient maintenir un appui sur un levier pour être récompensés, des lésions bilatérales du NST induisent une augmentation nette des réponses prématurées (les rats se désengageaient précocement de leur action). Dans cette même tâche, des rats traités par 6-OHDA entraînant une destruction des neurones dopaminergiques présentaient eux, comme attendu, un relâchement tardif plus important. Il est intéressant de noter que ces mêmes rats après lésion du NST présentaient également un comportement plus impulsif par rapport à des rats non lésés, bien que le ralentissement moteur induit par les lésions dopaminergiques ait été rétabli (Baunez et al., 1995).

Dans une tâche nécessitant plus de ressources attentionnelles comme la 5-CSRT, les effets d'une lésion du NST sont plus complexes. En résumé, après avoir démarré un essai en appuyant sur un bouton, les rats doivent ensuite attendre qu'une lumière brève apparaisse sur un des cinq stimuli disposés devant eux pour pouvoir s'y engager. S'ils ont réussi l'essai, la récompense se situera au niveau du bouton initial (figure 13). Ce type de paradigme impose aux rats de maintenir une attention divisée (avoir une vue globale sur l'ensemble des stimuli) puis d'avoir une attention focale (identifier rapidement le stimulus allumé).

Lors d'une lésion définitive du NST, des rats avec et sans lésion des neurones dopaminergiques présentaient une altération diffuse des performances à la tâche traduisant un déficit attentionnel : ils réalisaient ainsi plus d'erreurs pour identifier le bon stimulus et leurs réponses étaient prématurées (Baunez & Robbins, 1997). En utilisant une action réversible sur le NST par stimulation à haute fréquence, il a été observé le même type de résultat (Baunez et al., 2007). Un résultat plus particulier est l'apparition d'un comportement de persévération sur l'un des stimulus ou bien sur la boîte renfermant la récompense alimentaire aussi bien chez des rats dont la lésion est définitive, que ceux ayant

une stimulation à haute fréquence du NST (Baunez et al., 2007; Baunez & Robbins, 1997). Là encore, ces comportements de persévération persistaient avec la présence combinée d'une inactivation du système dopaminergique et du NST.



**Figure 16.** Représentation de la 5-CRST. L'engagement dans un essai débute par un appui sur un levier situé derrière les stimulus (i) ; après un délai programmé dans la tâche, un des 5 stimulus s'illumine (ii) et l'animal doit alors s'y diriger (iii). Si l'essai est réussi, il obtient une récompense au même endroit où se situait le levier d'activation de l'essai (iv). Plusieurs types d'erreurs peuvent être observés : les animaux peuvent soit réaliser des réponses prématurées (vii), soit avoir un comportement de persévération (v et vi ; d'après Eagle & Baunez, 2010).

L'interprétation d'un comportement de persévération, surtout lorsqu'il est médié vers un stimulus récompensant, présente quelques différences avec un comportement impulsif de types réponses prématurées. En effet, une dimension motivationnelle se surajoute à la composante d'impulsivité du fait de ce comportement répétitif, une analogie étant possible avec les comportements compulsifs chez l'humain.

#### Des arguments pour un rôle motivationnel du NST

Le rôle du NST dans des processus motivationnels, notamment médiés par une récompense, a émergé récemment. Premièrement, il semblerait pouvoir traiter différentes sortes de récompenses comme suggéré par l'effet des lésions du NST chez le rat. Dans une tâche

renforçant l'appui sur levier soit par une récompense alimentaire, soit par cocaïne, des lésions du NST avaient un effet opposé. L'effet motivationnel était alors mesuré par l'augmentation progressive du nombre d'appuis à effectuer pour obtenir une récompense ; ainsi, plus la récompense accumulée est élevée à la fin de l'épreuve, plus l'animal aura été motivé pour la récupérer. Alors que la préférence pour une récompense alimentaire était accrue par les lésions du NST, celles pour la cocaïne, diminuait après une lésion du NST (Baunez et al., 2005).

L'activité neuronale d'un tel effet motivationnel et discriminant des récompenses a d'ailleurs été caractérisée par des enregistrements montrant des décharges neuronales distinctes suivant le type de récompense, ou distinguant un stimulus récompensant d'un autre aversif (Breyse et al., 2015; Lardeux et al., 2013).

Au final, l'activité du NST faciliterait l'émergence d'actions optimisées au détriment d'actions impulsives et potentiellement délétères. L'effet lésionnel chez des animaux avec ou sans déplétion dopaminergique semblent donner des résultats similaires. D'autres modèles expérimentaux ont par ailleurs suggéré que le NST soit également impliqué dans la valuation des stimulus récompensant. A l'inverse de l'effet inhibiteur principal qui caractérise sa fonction, il semblerait donc exister un système promouvant l'action au sein du NST et ceci médié par des récompenses. Des comportements similaires ont également été constatés chez des patients stimulés dans le cadre de la maladie de Parkinson.

## Conséquences non motrices de la stimulation du NST dans la maladie de Parkinson

Depuis l'émergence de la stimulation du NST, plusieurs observations cliniques ont fait état de modifications cognitives ou neuropsychiatriques comme un défaut d'inhibition (Jahanshahi et al., 2000; Witt et al., 2004) ou des manifestations d'hypomanie post-

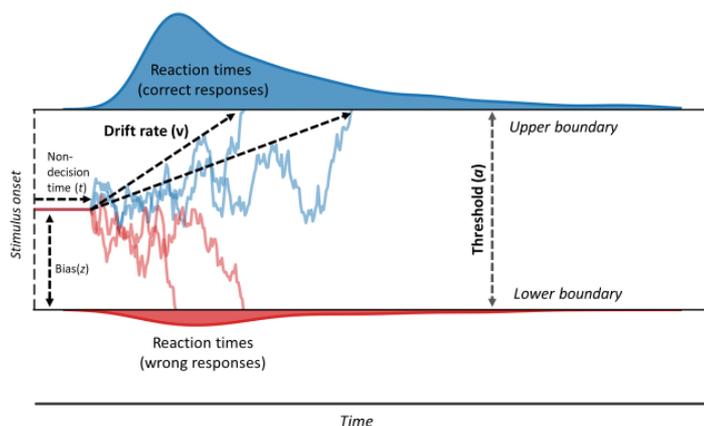
opératoire (Prange et al., 2022; Welter et al., 2014). Il a même été constaté une augmentation du risque suicidaire même si ces causes sont probablement multifactorielles (Voon et al., 2008).

#### Un effet spécifique dans des tâches de conflit

Ces observations mettent donc au premier plan, comme chez l'animal, une action de contrôle du NST, dont l'inactivation serait pourvoyeuse d'impulsivité. Frank et al. (2007) ont proposé un modèle explicatif du rôle du NST dans la prise de décision impulsive, notamment en présence de conflit. Lors d'une tâche probabiliste proposée à des patients stimulés pour une maladie de Parkinson, les patients ON (stimulation active) confrontés à un essai conflictuel (par exemple deux paires d'images dont la récompense est proche), avaient même tendance à accélérer leur décision comparés à des sujets sains du même âge (Frank et al., 2007). Dans son modèle, Frank et al. suggèrent que le NST est impliqué dans la résolution de conflit en « stoppant » tout signal dans laquelle la meilleure décision n'est pas évidente, et nécessite plus de temps de traitement de l'information. En accord avec les données issues de l'expérimentation animale, les constatations de Frank et al. sont concordantes avec l'induction de comportements impulsifs par la modulation du NST, limitant la capacité des patients à différer une action prématurée et ainsi d'en accomplir une autre de manière différée et plus adaptée.

L'hypothèse selon laquelle le NST serait impliqué dans l'optimisation de l'action en inhibant des comportements immédiats en vue d'une réponse optimisée a été enrichie par l'utilisation de modèles computationnels d'accumulation de l'information comme le *drift diffusion*. Dans ce modèle, plusieurs paramètres interviennent pour décider si une option a plus d'intérêt d'être choisie qu'une autre (figure 14). Ainsi, un système inhibiteur limite la prise de décision au passage d'un « seuil » qui doit être franchi pour que celle-ci soit prise. Lors de cette inhibition, un système de traitement de l'information évalue distinctement les différentes options possibles pour faire « dériver » la quantité d'évidence accumulée vers

une décision finale ; ce deuxième paramètre s'appelle le « taux de dérive » (Cavanagh et al., 2011).



**Figure 17.** Schéma explicatif du *drift diffusion model* sur lequel est modélisé le « seuil » (*threshold*)  $a$ , « taux de dérive » (*drift rate*)  $v$ , ainsi que les valeurs du modèle de prise d'information amenant soit à une bonne décision en bleu, ou une mauvaise en rouge. Sont représentés également dans ce modèle un éventuel biais  $z$  correspondant à l'a priori du sujet vers l'une ou l'autre des décisions. (D'après Vinding et al., 2018).

En utilisant le même type de tâche avec conflit que celle de Frank et al. en 2007, Cavanagh et al (2011) ont observé que le paramètre de « seuil décisionnel » était conjointement corrélé à l'augmentation de l'activité EEG thêta dans le cortex préfrontal médian ainsi que celle du NST en enregistrement des potentiels de champs locaux. Il existerait donc un réseau fronto-subthalamique, responsable de la détection et le traitement de signaux de conflit. Lorsqu'un choix comporte plusieurs options peu discriminables, l'activité de ce réseau fronto-subthalamique a alors un rôle inhibiteur, limitant l'émergence d'une réponse prépotente ou possiblement erronée afin d'accumuler le maximum d'évidence pour optimiser la décision. Dans la maladie de Parkinson, ce seuil serait plus conservateur, limitant de manière excessive un plus grand nombre de comportements. A l'inverse, chez les patients stimulés, ce seuil est abaissé de telle façon qu'il provoque l'émergence de réponses prématurées à risque d'être suboptimales, c'est-à-dire impulsive (Frank et al., 2007).

Vers un système encodant la valeur des stimulus ?

L'observation d'un effet pro-motivationnel de la stimulation du NST chez certains patients vient paradoxalement remettre en perspective son effet exclusivement pro-impulsif. De nombreuses observations dans les années 2000 ont rapporté l'émergence d'une apathie post-opératoire, semblant présenter une cinétique croissante sur les mois suivants et rendant ainsi moins probable un mécanisme aigu lié par exemple au geste chirurgical (Funkiewicz, 2002, Abbes 2018, Drapier et al., 2006). L'hypothèse d'une influence directe de la stimulation des réseaux limbiques du NST sur les réseaux frontaux a été mise en avant, ce qui a d'ailleurs été objectivé par une étude en PET-TDM avec une corrélation significative entre le degré de sévérité de l'apathie et le métabolisme régional des régions frontales (Le Jeune et al., 2009).

L'explication la plus fréquemment évoquée ne viendrait pas d'un effet direct de la stimulation mais plutôt des diminutions parfois majeures et rapides du traitement dopaminergique dans les semaines suivant la chirurgie, ce qui peut être pourvoyeur de syndrome de sevrage (Lhommée et al., 2018; Thobois et al., 2010).

Une autre hypothèse, semblable à celles retrouvées chez les animaux, serait un effet propre du NST qui serait capable de traiter les récompenses. Dans une tâche faisant varier la magnitude des récompenses (5, 20 centimes ou 1 euro) et l'effort à effectuer pour l'obtenir au moyen d'un dynamomètre, il a été montré un effet propre du NST dans l'évaluation coût-bénéfice. Cette tâche impliquait des patients opérés chez qui un enregistrement des potentiels de champs locaux (PCL) était réalisé. Il a été ainsi montré une activité dans une bande entre 1 et 10 Hz au sein du NST corrélant aussi bien à l'effort réalisé, le taux d'acceptation des sujets à un essai (c'est-à-dire en quelles proportions étaient-ils motivés pour réaliser un essai, même difficile) ou la récompense reçue. Il s'agirait donc d'un

argument électrophysiologique suggérant que le NST serait activé par la valeur des stimulus (Zénon et al., 2016).

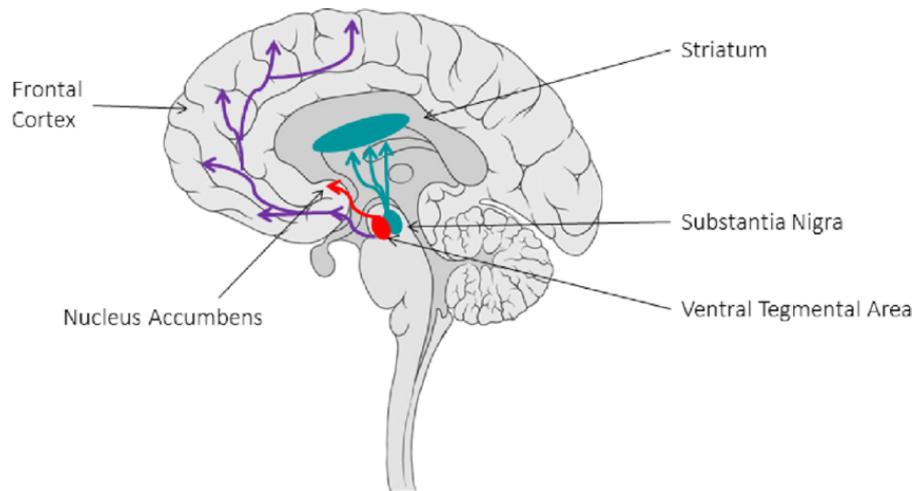
Une étude similaire utilisant le même paradigme a cette fois-ci analysé le comportement de patients ON et OFF (stimulation éteinte), et a montré un effet similaire sur le taux d'acceptation des essais, de la récompense, et du degré d'effort lorsque les patients étaient ON. Cet effet peut paraître paradoxal pour démontrer un effet pro-motivationnel mais peut être expliqué par l'hypothèse que la neurostimulation n'aurait pas un simple effet inhibiteur sur le NST (Atkinson-Clement et al., 2019).

En résumé, le NST est une structure d'importance majeure dans la prise de décision, notamment par sa connectivité directe avec certains éléments clés du cortex préfrontal. Il permettrait l'application d'un seuil pour discriminer deux options de valeur proche afin d'optimiser l'action sélectionnée et finalement exécutée. De plus, certains arguments suggèrent l'implication du NST dans des processus de valuation des stimulus récompensant. Cette hypothèse innovante est à mettre en perspective avec le lien entre le NST et les systèmes évaluant la récompense. En effet, ces systèmes bien connus depuis plusieurs décennies mettent plutôt en avant le système dopaminergique et ses projections avec les ganglions de la base et le cortex préfrontal.

#### **4. Les neurones dopaminergiques**

Le système dopaminergique mésencéphalique est composé de la SNc et d'une région adjacente médiale, appelée aire tegmentale ventrale. Ce système envoie des projections particulièrement diffuses vers l'ensemble du cerveau, en particulier vers le cortex préfrontal ainsi que l'ensemble des ganglions de la base (Björklund & Dunnett, 2007). On distingue deux voies venant moduler les ganglions de la base. La voie nigrostriée allant de la SNc au striatum, et la voie méso-cortico-limbique issue de l'aire tegmentale ventrale (figure 18), se

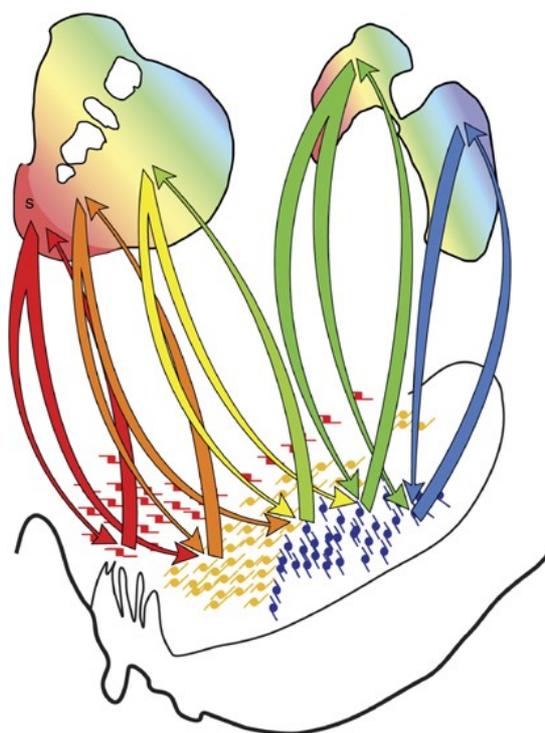
projetant sur le cortex préfrontal et cingulaire antérieur, ainsi que sur le noyau accumbens, partie la plus médiane et ventrale du striatum.



**Figure 18.** Représentation schématique des voies dopaminergiques avec la voie nigrostriée en vert, la voie mesocorticale en violet, et la voie mésolimbique se terminant sur le noyau accumbens en rouge (D'après Reneman et al., 2021).

## Anatomie fonctionnelle

Les neurones de la SNc et de l'aire tegmentale ventrale sont organisés en plusieurs sous-territoires anatomo-fonctionnels en fonction de leur projection sur le striatum. Globalement, l'aire tegmentale ventrale se projette sur la partie la plus médiane et ventrale du striatum (émotionnelle et motivationnelle) et la SNc se projette sur sa partie dorsolatérale (sensorimotrice et associative ; Haber, 2014, figure 19). Les travaux réalisés par S.N. Haber sur des cerveaux de primates non humains ont également mis en évidence une boucle de rétrocontrôle, la voie striatale se projetant sur les neurones dopaminergiques. Le rôle fonctionnel d'une telle voie est encore débattu.



**Figure 19.** Illustration des projections réciproques entre les neurones dopaminergiques et le striatum. Rouge, projections entre l'aire tegmentale ventrale et le striatum ventromédial/noyau accumbens. Orange, projections entre la SNc et les parties centrales du striatum. Bleu, projections entre la SNc et le striatum dorsolatéral. (D'après Haber, 2014).

Ainsi, la partie la plus ventrale du striatum recevant des neurones de l'aire tegmentale ventrale est impliquée dans des fonctions limbiques, tandis que la SNc se projetant dans des territoires centraux et dorsolatéraux est impliquée dans des processus associatifs et sensorimoteurs. Bien que cette notion soit à nuancer par l'existence d'un gradient entre aire tegmentale ventrale et SNc et non de territoires ségrégués de manière stricte, ils permettent de mettre en évidence deux grandes fonctions distinctes au sein d'une topographie différente. D'une part, la SNc occuperait des fonctions de sélection de l'action, tandis que les neurones de l'aire tegmentale ventrale s'occuperaient du calcul de la valeur des stimulus.

Dans les paragraphes suivants, et pour des raisons similaires à celles exposées dans le chapitre sur le NST, il sera décrit plus en détail les fonctions non motrices du système dopaminergique et leur altération dans la maladie de Parkinson.

## Les neurones dopaminergiques encodent une erreur de prédiction

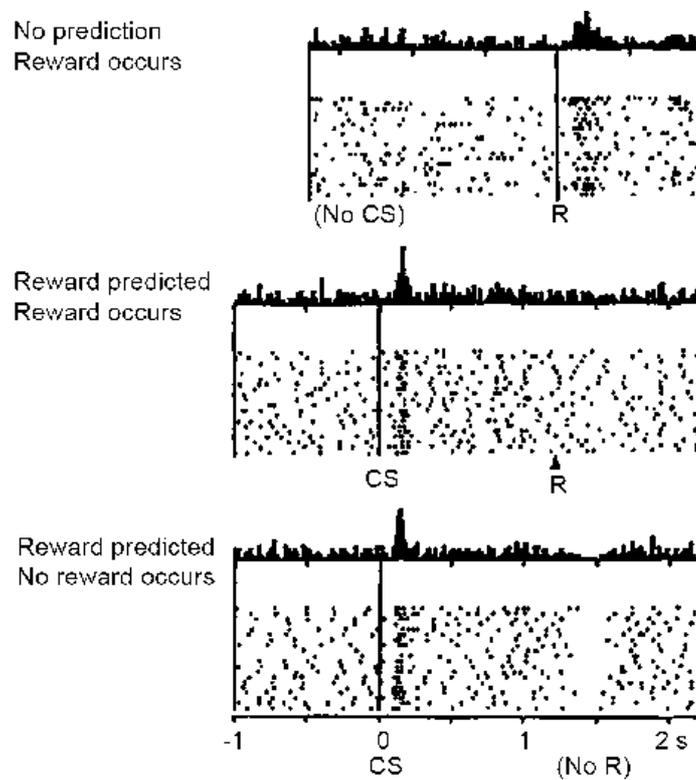
De nombreux travaux pionniers dans l'enregistrement des neurones dopaminergiques chez le primate ont permis de décrire avec précision leurs différentes caractéristiques, en fonction de la présentation ou non d'un stimulus externe, de la réalisation d'un mouvement ou que celui-ci soit auto-généré ou induit par un stimulus. Il a été observé une activité lente et modérée des neurones précédant l'initiation d'un geste et se poursuivant dans une moindre mesure lors de sa réalisation suggérant leur implication dans la génération interne du mouvement.

Un tout autre pattern bien plus fréquent correspond à une activité brève, phasique dépendant spécifiquement de la présentation de récompenses (Romo & Schultz, 1990; Schultz, 1986). Une certaine perplexité transparait des auteurs dans ces premiers travaux car cette activité phasique a aussi été observée non seulement lorsque la récompense était présentée mais également en présence d'un stimulus neutre, comme lors de l'ouverture de la boîte contenant la récompense. Autrement dit, l'activité dopaminergique renseignait aussi bien sur la présence d'une récompense que sur des stimulus qui lui étaient associés. Les auteurs ont alors suggéré un "transfert" de la réponse neuronale depuis la prise de la nourriture à la survenue d'un événement sans valeur propre mais associé dans le temps avec le stimulus récompensant.

Ce signal correspondrait donc à un signal d'apprentissage ; en effet dans des tâches de conditionnement (à la fois simple et instrumental) avec enregistrements neuronaux chez le singe, l'apparition d'une récompense était suivie du signal phasique, la récompense n'étant pas prédite. Lorsque la récompense est prédite par un stimulus conditionné, le signal neuronal s'est propagé au moment de l'apparition du stimulus conditionné et non plus de la

récompense. De plus, lorsque le stimulus conditionné n'est pas suivi d'une récompense, il est observé une extinction des neurones (figure 20).

En résumé, les neurones dopaminergiques sont activés lorsqu'une situation est « meilleure que prédite » (récompense non prédite), et sont inhibés lorsque la situation est « moins bonne que prédite » (récompense prédite mais non présente). Ce signal est alors un substrat neuronal de l'erreur de prédiction de la récompense des modèles d'apprentissage par renforcement (Montague et al., 1996; Schultz et al., 1997).



**Figure 20.** Activité des neurones dopaminergiques en fonction de la présence ou non d'évènements appétitifs. Modification de l'activité de neurones dopaminergiques lors de la présentation d'une récompense non prédite (activation phasique suivant la récompense, figure du haut), prédite et suivie d'une récompense (activation phasique suivant le stimulus conditionné, figure du milieu) et prédite mais suivie de l'absence de récompense (inhibition phasique suivant le stimulus conditionné non suivi d'une récompense, figure du bas). CS : stimulus conditionné (*conditioned stimulus*) ; R : récompense (*reward*). (D'après Schultz et al., 1997)

De plus, ce signal encode également une information. Dans le même type de paradigme, la probabilité ou la magnitude de récompense était proportionnelle à l'amplitude de réponse des neurones (Fiorillo et al., 2003; Tobler et al., 2005). Il s'agit donc d'un signal précis, encodant des informations quantitatives sur la différence entre ce qui est attendu et ce qui est obtenu.

La découverte que les neurones dopaminergiques ne signalent pas la présence de récompenses mais une erreur de prédiction a permis de formaliser avec précision l'activité de ces neurones au moyen de différents algorithmes computationnels. Ce lien robuste entre un système neuronal et sa modélisation a été probablement un des plus étudiés et reproductibles en neurosciences, tant sur des données comportementales, de neuroimagerie ou d'électrophysiologie (Niv, 2009; Schultz, 2000).

L'utilisation de ces algorithmes fait également concorder les composants de modèles computationnels avec l'anatomie fonctionnelle du striatum. Ainsi, l'activité en IRM fonctionnelle du striatum ventral et dorsal de sujets sains réalisant une tâche d'apprentissage sont corrélés à des paradigmes différents. L'activité du striatum ventral est ainsi rattachée à des tâches de conditionnement simple (perceptif), tandis que celle du striatum dorsal au conditionnement instrumental (O'Doherty et al., 2003, 2004). Une comparaison avec le modèle Acteur-Critique peut également être réalisée : le striatum ventral, composante « Critique », encoderait la valeur des stimulus, tandis que le striatum dorsal, composante « Acteur », participerait plus à la sélection de l'action.

Il en est de même pour la valence de l'erreur de prédiction dont l'activité striatale est similaire à celle des neurones dopaminergiques enregistrés, correspondant à une augmentation d'activité lorsque l'erreur est positive, et une diminution d'activité lorsqu'elle est négative (la récompense obtenue est moins importante que prédite ; Delgado et al., 2000; Pessiglione et al., 2006; Seymour et al., 2007). Cet effet a également été pharmacologiquement prouvé par l'administration de dopamine par voie exogène (L-

DOPA) à des sujets sains, améliorant leurs capacités d'apprentissage, tandis que l'administration d'un antagoniste de la dopamine l'altérait (Pessiglione et al., 2006).

Ainsi, puisque la dopamine participe à l'apprentissage de la valeur de stimulus pourvoyeur d'une récompense, la dénervation dopaminergique observée dans la maladie de Parkinson serait également responsable d'une altération dans des processus d'apprentissage.

### Conséquences comportementales de la maladie de Parkinson et le traitement médicamenteux

Alors que la lésion-clé de la maladie de Parkinson est une perte des neurones de la SNc, il s'avère qu'une majorité de patients présentent également une dégénérescence concernant aussi les neurones de l'aire tegmentale ventrale. Reposant sur des études histologiques de cerveau de patients atteints de maladie de Parkinson ont mis en évidence une perte de 50% des neurones de l'aire tegmentale ventrale, et également jusqu'à 65% de perte de dénervation du striatum ventral, principale projection de ces neurones (Alberico et al., 2015; Chinaglia et al., 1992). Même si cette perte n'est pas aussi importante que celle observée dans la SNc (aux alentours de 80%), cela suggère qu'il est possible d'observer chez ces patients une altération comportementale spécifique.

Ainsi, en utilisant un paradigme d'apprentissage probabiliste chez les sujets sains et chez des patients parkinsoniens, les capacités d'apprentissage médiés par des récompenses sont significativement altérées chez les patients non traités. En revanche, l'administration de L-DOPA permet d'en améliorer les performances (Palminteri et al., 2009). Toutefois, cet effet de la prise de dopamine exogène chez les patients parkinsoniens peut paradoxalement altérer les processus optimaux de renforcement. Il a ainsi été montré que la prise de L-DOPA chez les patients améliorerait effectivement leurs capacités d'apprentissage par des

récompenses, mais altérerait leur capacité à se désengager de stimulus dévalués (Cools et al., 2001).

Ce paradoxe a été formalisé au moyen d'un modèle plus complet du fonctionnement des ganglions de la base dans lequel chaque structure et chaque voie est mathématiquement représentée (un voie "Go" -directe- et une voie 'No Go' -indirecte). Ce modèle permet ainsi de simuler la disparition de l'activité des neurones dopaminergiques, sa supplémentation par la prise de médicaments dopaminergiques ainsi que toutes ses conséquences sur l'ensemble des structures du réseau, et donc du comportement final (Frank, 2005).

Frank et al. (2004) ont ainsi observé que des patients réalisant une tâche probabiliste avec (ON) et sans traitement (OFF) arrivaient mieux à apprendre à éviter des pertes en OFF. A l'inverse, les patients ON arrivaient mieux à apprendre à accumuler des gains. L'un des résultats majeurs est que la réciproque est vraie, les patients ON ont des difficultés à éviter les pertes et les patients OFF à accumuler des gains. En utilisant son modèle des ganglions de la base, Frank et al. expliquent ce résultat par la variation de l'activité tonique des neurones dopaminergiques. Ainsi, en ON, le traitement dopaminergique atténuerait l'inhibition phasique observée lorsque l'erreur de prédiction est négative. Les stimulus aversifs ou dévalués dans le temps sont donc moins bien appris. A l'inverse, en OFF, l'erreur de prédiction positive est atténuée mais l'apprentissage des « punitions » serait accentué (Frank et al., 2004).

Ce modèle explique les difficultés des patients traités à se détacher d'un comportement déjà appris mais apporte également un mécanisme du trouble du contrôle des impulsions (TCI) fréquemment observé chez une partie des patients parkinsoniens et probablement sous-diagnostiqué. Il s'agit de la répétition d'un comportement appétitif, émergeant souvent des préférences du patient lui-même (achats, jeux d'argent, sexe, nourriture...) mais prenant une dimension incontrôlable et dont les conséquences psycho-sociales de cette perte de

contrôle, parfois dramatique, semble ne pas être considérée ou très atténuée par le patient (Evans et al., 2009; Voon et al., 2017).

Plusieurs facteurs prédisposants ont été mis en évidence l'âge, le sexe et surtout la durée d'exposition aux traitements dopaminergiques (Corvol et al., 2018; Weintraub et al., 2010a). Le rôle particulier des agonistes dopaminergiques semblerait être prépondérant comparé à la L-DOPA. Les raisons en sont complexes, probablement multiples.

D'un côté les récepteurs dopaminergiques-cibles de ces traitements sont différents en structure et en localisation. Ainsi les agonistes dopaminergiques ont une affinité pour des sous-classes de récepteurs présents aussi bien sur les neurones du striatum que sur ceux du cortex préfrontal médial (impliqué dans le contrôle inhibiteur, de type D2). La L-DOPA en revanche aurait une affinité supérieure pour les sous-classes de récepteurs présents uniquement sur les neurones du striatum (type D1). D'autre part, leurs propriétés pharmacocinétiques et pharmacodynamiques diffèrent : ainsi la L-DOPA aurait probablement une action sur l'amplitude de la réponse neuronale, s'apparentant à la réponse phasique, alors que les agonistes induiraient une activité continue sur les récepteurs, majorant ainsi la composante tonique (Frank et al., 2004; Voon et al., 2017).

Un autre argument expliquant l'émergence hétérogène de tels symptômes chez des parkinsoniens viendrait également des lésions différentes de l'aire tegmentale ventrale chez ces patients. Il a été ainsi supposé que les patients présentant moins de lésions de l'aire tegmentale ventrale seraient à risque de présenter des TCI, l'absorption de dopamine exogène venant suractiver ce réseau tout en corrigeant la perte de dopamine sur les parties motrices du striatum (Weintraub & Mamikonyan, 2019).

L'argument pharmacologique d'une majoration de l'activité tonique des neurones est concordant avec le modèle de Frank et al. (2004). Dans sa tâche, les sujets devaient non seulement sélectionner les images les plus récompensées (« Go ») mais éviter celles le moins

récompensées (« No Go »). En utilisant un paradigme identique, il a été ainsi montré que les patients parkinsoniens présentant des TCI comparés à des patients n'en ayant pas apprenaient mieux dans un contexte de gain tandis que leurs performances étaient moins bonnes pour éviter les pertes (Voon et al., 2010). De même, suivant un autre paradigme impliquant la prise de risque, les patients ayant un TCI expriment un biais vers des comportements plus risqués, suggérant qu'ils sous-évaluaient le risque, considéré comme un stimulus aversif (Voon et al., 2011).

Au final, la vulnérabilité des structures sous-tendant l'apprentissage par des récompenses dans la maladie de Parkinson explique les modifications comportementales observées chez certains patients, notamment ceux souffrant d'un TCI.

## **5. Conclusion intermédiaire**

Le système dopaminergique intègre les informations collectées par les individus lorsqu'ils se trouvent dans une situation où ils réalisent une action et la comparent à leur prédiction sur le résultat final. Ce signal est alors transmis aux ganglions de la base pour modéliser une valeur subjective en fonction du caractère récompensant (résultat meilleur que prévu) ou aversif (résultat moins bon que prévu) et en déduire un comportement d'approche ou de retrait. Au sein des ganglions de la base, le noyau subthalamique occupe une place particulière permettant de sélectionner rapidement une action adaptée lorsque deux situations présentent des valeurs subjectives plus ou moins conflictuelles. La destruction des neurones dopaminergiques dans la maladie de Parkinson entraîne un déséquilibre aussi bien de valuation que de sélection de l'action, du fait notamment de l'hyperactivité du NST qui en résulte.

## HYPOTHESES PRINCIPALES

Les hypothèses de ce travail peuvent être résumé en deux parties. Premièrement, nous émettons l'hypothèse que la préférence pour le choix est une récompense intrinsèque. Deuxièmement, nous pensons qu'elle est traitée comme telle par le système des ganglions de la base et le système dopaminergique.

Ces deux aspects sont successivement traités dans les deux études suivantes. Dans la première, des sujets sains ont réalisé une tâche permettant d'évaluer de manière indépendante leur préférence pour le choix. Dans la seconde, des patients atteints d'une maladie de Parkinson ont réalisé la même tâche que celle utilisée dans l'étude 1. Les patients étaient divisés en deux groupes. L'un recevant uniquement un traitement dopaminergique oral, l'autre bénéficiant d'une stimulation du NST.

Les différentes hypothèses sont reprises plus en détail en début de chaque étude.



## ETUDE 1

**Arguments comportementaux pour une valeur intrinsèque  
du choix chez des sujets sains**

## Etude 1 : introduction

Comme nous l'avons vu, il existe une préférence naturelle pour le choix libre dans plusieurs espèces, même lorsque les situations de choix forcé amènent à un résultat identique. Le choix libre serait donc perçu comme une récompense au même titre que des récompenses plus basiques comme la nourriture par exemple (Catania & Sagvolden, 1980; Suzuki, 1999; Voss & Homzie, 1970). Il en va de même chez l'humain où prime cette notion du choix libre, matérialisée récemment par des analyses neuro-computationnelles (Chambon et al., 2020; Cockburn et al., 2014).

Plusieurs hypothèses psychologiques ont été proposées, dont celle qu'avoir le choix amène plus de contrôle et que ce contrôle sur l'environnement et les événements était intrinsèquement motivant (Ryan & Deci, 2000). Or, les paradigmes expérimentaux précédents n'ont pas permis de confirmer cette hypothèse. En effet, il s'agissait le plus souvent de tâches probabilistes entremêlant des essais dans lequel le sujet pouvait choisir librement à ceux où un choix leur était imposé. Cette configuration ne permet pas de capturer le moment où le patient exprime lui-même ses préférences. Une autre configuration est celle du labyrinthe de Voss & Homzie (1970) dans lequel il existe deux étapes. Dans la première, les animaux choisissent s'ils voulaient ou non s'engager dans le bras du labyrinthe avec choix, dans la seconde il se dirigeait directement vers une récompense extrinsèque. Là encore, ce type de configuration est soumis au biais d'exploration de l'individu. En effet, étant naïf de l'environnement, les bras avec choix seront d'autant plus préférés que l'animal voudra apprendre des différentes options. Choisir apparaît alors non pas comme motivant en tant que tel mais comme un moyen de diminuer l'incertitude.

Pour contrôler ces deux biais, nous avons réalisé une tâche en deux étapes sur le modèle du labyrinthe de Voss & Homzie mais comportant une première étape de sélection d'un choix « libre » ou d'un choix « forcé », et une deuxième étape de récupération de récompenses

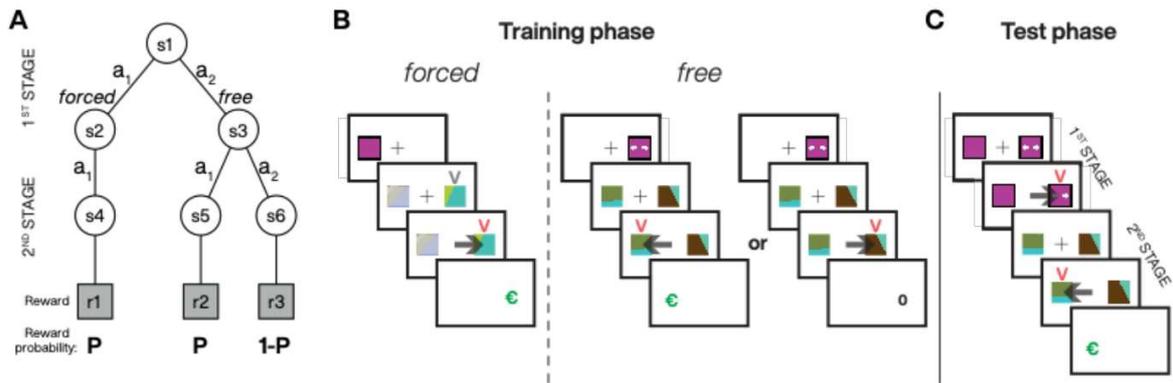
extrinsèques distribuées de manière probabiliste. Avant cela, les sujets ont été entraînés à apprendre les différentes contingences de récompenses extrinsèques afin de contrôler une préférence pour le choix libre qui serait biaisée par un comportement d'exploration.

## **Etude 1 : méthodologie**

Les sujets qui ont participé à cette étude sont au nombre de 94 adultes entre 18 et 45 ans. Les sujets n'avaient pas de pathologie neurologique et psychiatrique et parlaient français couramment. La passation de la tâche se réalisait sur ordinateur dans une pièce calme.

La tâche a été construite comme une structure en deux étapes. Lors de la première étape, deux images étaient affichées à l'écran, représentant des options "libres" ou "forcées", et les sujets devaient faire leur choix. Lors de la deuxième étape, deux images différentes étaient affichées. Dans l'option "libre", les sujets pouvaient choisir n'importe laquelle des deux images, tandis que dans l'option "forcée", ils devaient choisir celle qui avait été présélectionnée par l'ordinateur. Notons que l'ordinateur sélectionnait toujours la même image et que la sélection de l'image opposée était désactivée pendant la condition "forcée".

Après avoir sélectionné l'image de la deuxième étape, le résultat de la récompense extrinsèque était affiché avec une probabilité  $P$  pour une des images et  $1-P$  pour l'autre côté choix libre, et toujours  $P$  côté choix forcé (Figure 1).



**Figure 21.** Structure de la tâche en deux étapes comportant les 6 états (s) et les 6 actions (a) ainsi que leurs probabilités de récompense extrinsèque. A : diagramme illustrant les différentes voies que les sujets empruntent au cours de la tâche. B : illustration schématique d'un essai pendant l'entraînement. Pour chaque bloc de probabilité, les sujets s'entraînaient sur deux blocs séparés : un avec choix libre, un avec choix forcé. C : illustration schématique d'un essai pendant le test.

Au début de chaque bloc, c'est-à-dire avant la phase de test, les sujets ont effectué une phase d'entraînement afin d'apprendre les contingences entre les différentes images et leurs probabilités de récompense associées. Au cours de cette phase, il n'est pas proposé aux sujets de choisir entre des options libres ou forcées et ils ont donc effectué successivement des essais libres et forcés obligatoires

Trois expériences sont réalisées :

Dans la première, les contingences de récompense extrinsèques varient pour adresser la dimension exploratoire du choix : il existait ainsi trois blocs de probabilité différente,  $P=0.5/P=0.5$  (stochastiques),  $P=0.75/P=0.25$  (probabiliste),  $P=1/1-P=0$  (déterministe). Dans l'option « choix forcé », la récompense est toujours délivrée avec une probabilité  $P$ . 58 sujets ont participé à cette tâche.

Dans la seconde expérience, la probabilité de récompense extrinsèque à la suite du choix forcé augmente progressivement, permettant ainsi de déterminer un point d'inflexion dans

les préférences des sujets entre choix libre et choix forcé et ainsi donner une « valeur relative » du choix libre par rapport au choix forcé. 36 sujets ont participé à cette tâche.

Dans la troisième, la contrôlabilité de la seconde étape est altérée. Ainsi lorsque le sujet sélectionne une des images, l'ordinateur sélectionne l'image opposée dans une proportion donnée d'essais. 36 sujets ont participé à cette tâche.

Les données des expériences 1 et 2 ont servi à construire un modèle d'apprentissage par renforcement. Dans ces modèles, la valeur intrinsèque du choix était paramétrée par un « bonus », c'est-à-dire une constante qui se surajoutait à la récompense obtenue, uniquement pour les essais « choix libre ». Si l'on reprend l'équation présentée en introduction définissant l'erreur de prédiction, le bonus correspond à un nombre scalaire ajouté à la récompense, nommée arbitrairement  $\rho$ .

$$V(s,a) \leftarrow V(s,a) + \alpha \times \delta$$

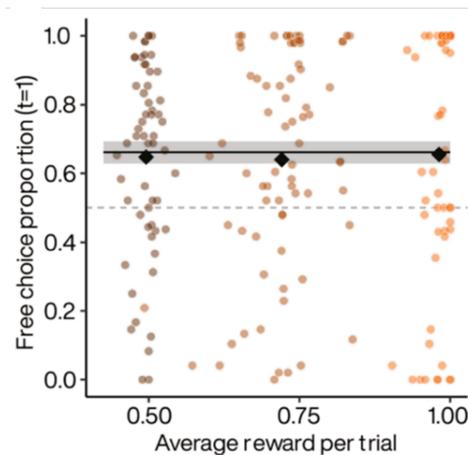
$$\delta = [r + \rho + V(s,a)] - V(s,a)$$

Deux types de raisonnement des sujets ont été explorés par deux modèles différents. L'un, dit « optimiste », se traduisait par une modélisation de l'erreur de prédiction en n'envisageant que la « meilleure des options ». Par exemple, au cours de la tâche, le sujet « optimiste » évaluera la valeur du choix libre sur l'image la plus récompensée, et il fera de même pour le choix forcé. Un deuxième modèle dit « pessimiste » a également été réalisé. Dans ce genre modèle, le sujet prend ses décisions à la fois sur les meilleures et les pires options. Par exemple, au cours de la tâche, le sujet « pessimiste » évaluera la valeur du choix libre sur l'image la plus récompensée ainsi que la moins bonne. Pour les essais « choix forcé », il ne pourra évaluer leur valeur que sur une image. En d'autres termes, ce modèle a tendance à diminuer la valeur du choix.

## Etude 1 : résultats

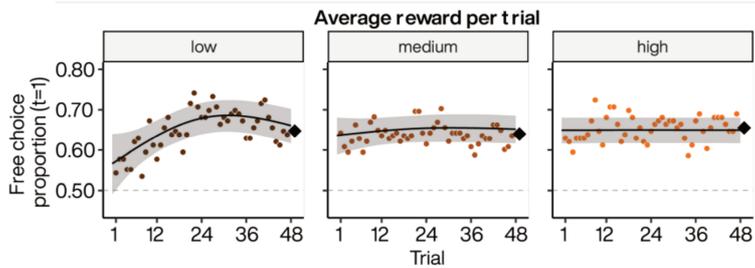
### Expérience 1 :

Les réponses des sujets ont été analysées au moyen d'un modèle mixte généralisé. Dans ce modèle était inclus non pas la probabilité de récompense extrinsèque théorique, mais la moyenne des récompenses accumulées par le sujet pour chaque bloc. Les sujets exprimaient une préférence pour le choix avec 64% en moyenne avec très peu de variabilité entre les probabilités de récompense extrinsèque (entre 64 et 66% ; figure 2).



**Figure 22.** Proportion de choix libre en fonction de la récompense extrinsèque moyenne. Les sujets sont représentés par les points colorés. Les losanges noirs correspondent à la moyenne sur l'ensemble des sujets par bloc. La ligne noire correspond à l'estimation du modèle GLMM avec intervalle de confiance à 95% (bandeau gris).

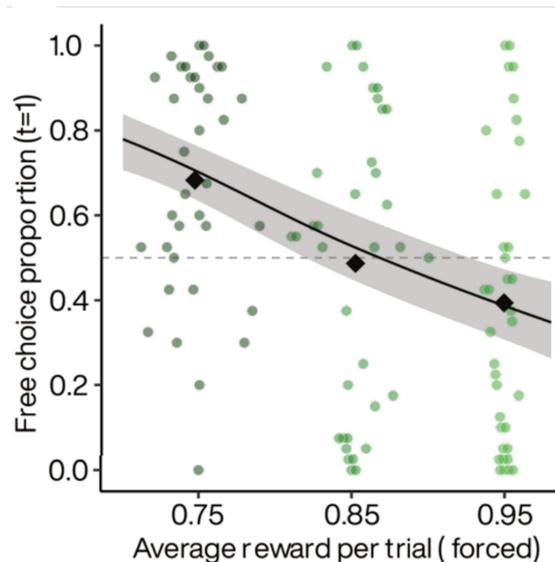
Il existe une absence de variabilité au cours de la tâche de la préférence pour le choix libre, les sujets ayant tendance à exprimer leur préférence dès le début de la tâche et de s'y tenir, excepté pour la plus basse récompense où l'expression du choix libre augmente progressivement le long des essais.



**Figure 23.** Dynamique de la proportion de choix libre essai par essai. Low/medium/high : récompense accumulée basse, moyenne et grande respectivement. Chaque point coloré correspond à un sujet, les losanges symbolisent la moyenne des sujets entre tous les essais. La ligne noire correspond à l'estimation du modèle (GAMM) accompagné de son intervalle de confiance.

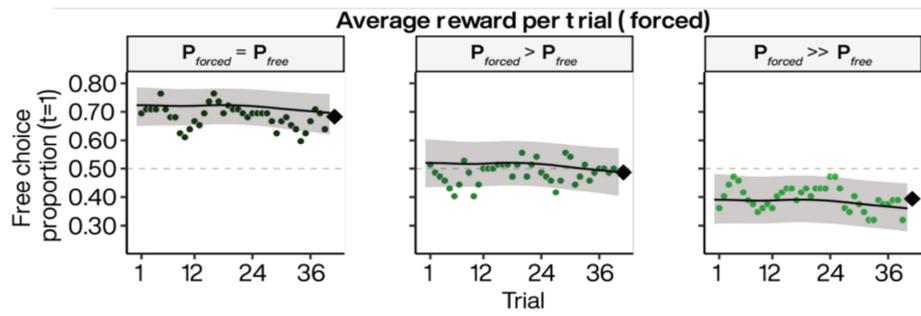
### Expérience 2 :

Pour les essais choix libre, la probabilité de récompense extrinsèque était de  $P = 0.75 / P-1 = 0.25$ . Pour les essais choix forcé, la probabilité de récompense extrinsèque variait sur trois blocs à 0.75, 0.85 et 0.95. On observe une diminution croissante de la proportion de choix libre : 68, 49 et 39% pour chaque bloc. Le point d'indifférence entre choix libre et choix forcé était à  $P = 0.88$  de récompense extrinsèque liées au choix forcé.



**Figure 24.** Proportion de choix libre en fonction de la récompense extrinsèque moyenne pour les choix forcés. Les sujets sont représentés par les points colorés. Les losanges noirs correspondent à la moyenne sur l'ensemble des sujets par bloc. La ligne noire correspond à l'estimation du modèle GLMM avec intervalle de confiance à 95% (bandeau gris).

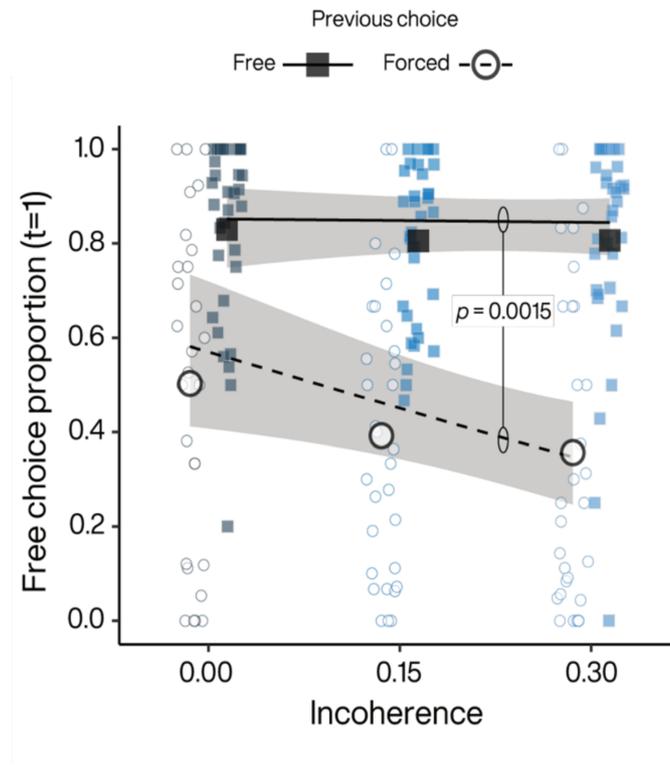
La dynamique de la proportion du choix essai par essai montre une décision constante au cours du temps, bien qu'il existe quelques fluctuations



**Figure 25.** Dynamique de la proportion de choix libre essai par essai. Chaque graphique dépend de l'asymétrie des récompenses finales entre les choix « libres » et « forcés ». A gauche, les récompenses sont identiques entre les deux bras, au milieu le choix libre est dévalué moyennement, à droite le choix libre est très dévalué. Chaque point coloré correspond à un sujet, les losanges symbolisent la moyenne des sujets entre tous les essais. La ligne noire correspond à l'estimation du modèle (GAMM) accompagné de son intervalle de confiance.

## Expérience 3

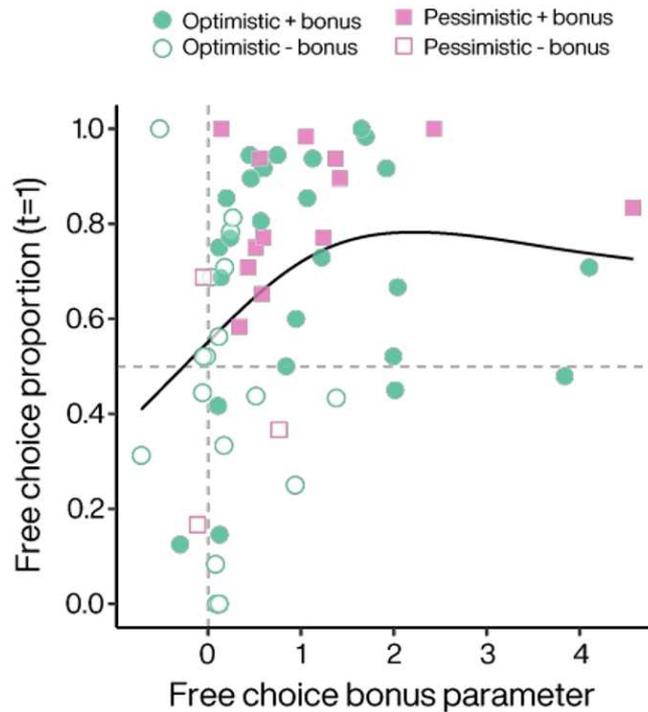
Dans cette expérience, l'ordinateur choisissait la réponse opposée au sujet dans une proportion des essais à la deuxième étape, aussi bien pour les essais de choix forcé que les essais de choix libre, ce qui est nommé ici incohérence. Cette perte de contrôlabilité permettait d'interroger si le sentiment de contrôle motivait les sujets à choisir le choix libre. Lorsque l'incohérence augmente, les sujets ont tendance à permuter d'un choix forcé à un choix libre, ce qui traduit un besoin de regagner le contrôle.



**Figure 26.** Résultats de la tâche altérant la contrôlabilité. Probabilité pour chaque sujet de permuter ou d'un essai choix forcé à un essai choix libre à l'essai suivant (cercle vide), ou bien de rester sur un choix libre à l'essais suivant (carré plein). Les points et carrés colorés représentent les sujets. En abscisse est représenté la proportion d'incohérence par bloc (15% d'essais incohérents ou 30%,0% étant un contrôle semblable à l'expérience 1).

## Modèles computationnels

La modélisation a montré que l'ajout d'un bonus permet de mieux expliquer les données et ce de manière quantitative : plus un sujet a une proportion pour le choix libre élevée, plus la valeur du bonus le sera également. De plus, chaque sujet s'ajustait plus à un modèle qu'à l'autre, certains étant plus pessimistes, d'autres plus optimistes



**Figure 27.** Valeurs du bonus pour le choix provenant des modèles à la fois pessimistes (rose) et optimistes (vert). Les figures pleines indiquent lorsque le meilleur modèle pour le sujet incluait un bonus.

## Conclusion de l'étude 1

Dans cette étude, il a été observé que les sujets préféraient majoritairement le choix libre, indépendamment du contexte. L'expérience 1 le démontre puisque ce choix libre était indépendamment de la probabilité de recevoir une récompense extrinsèque. Quand l'incertitude est absente (comme dans le bloc  $P=1$ ), cette préférence pour le choix est maintenue, allant à l'encontre des hypothèses du choix libre comme moyen de limiter la dissonance cognitive (inexistante dans ce bloc). De plus, les sujets ayant été entraînés avant d'effectuer la tâche, cette préférence pour le choix libre se maintient même hors d'un contexte exploratoire.

La valeur du choix libre vient des résultats de l'expérience 2 et de la modélisation. La valeur élevée du point d'indifférence auquel le choix libre est équivalent au choix forcé est un argument pour une valeur quantitative du choix libre. Les modèles viennent ajouter un argument supplémentaire puisque les données sont plus ajustées lorsqu'un bonus est ajouté au modèle. Le bonus serait donc la valeur quantitative du choix libre.

Enfin, dans l'expérience 3, l'ordinateur introduisait une perte de contrôlabilité du sujet. Il a été montré que plus la tâche était incontrôlable (en augmentant le niveau d'incohérence des réponses), plus le sujet avait tendance à préférer le choix. Ainsi, choisir librement permet au sujet de regagner du contrôle dans des situations où celui-ci venait à manquer.

Cette étude démontre et quantifie la valeur intrinsèque du choix chez le sujet sain et la rattache au besoin de contrôle.



## Article de l'étude 1

### Choice seeking is motivated by the intrinsic need for personal control

Available at: <https://doi.org/10.1101/2022.09.20.508669>

#### *Authors*

Jérôme Munuera<sup>1,2,\*</sup>, Marta Ribes Agost<sup>2</sup>, David Bendetowicz<sup>1</sup>, Adrien Kerebel<sup>‡</sup>, Valérian Chambon<sup>2,‡,\*</sup>, Brian Lau<sup>1,‡,\*</sup>

#### *Affiliations*

1. Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Inserm, CNRS, APHP, Paris, France

2. Institut Jean Nicod, Département d'études cognitives, ENS, EHESS, CNRS, PSL University, 75005 Paris, France

‡. Co-last authors

\*. Corresponding authors: [jerome.munuera@icm-institute.org](mailto:jerome.munuera@icm-institute.org); [valerian.chambon@ens.fr](mailto:valerian.chambon@ens.fr); [brian.lau@upmc.fr](mailto:brian.lau@upmc.fr)

Keywords: decision-making; reward; reinforcement learning; human; agency

#### Abstract:

When deciding between options that do or do not lead to future choices, humans often choose to choose. We studied choice seeking by asking subjects to decide between a choice opportunity or performing a computer-selected action. Subjects preferred choice when these options were equally rewarded, even deterministically, and were willing to trade

extrinsic rewards for the opportunity to choose. We explained individual variability in choice seeking using reinforcement learning models incorporating risk sensitivity and overvaluation of rewards obtained through choice. Degrading perceived controllability diminished choice preference, although willingness to repeat selection of choice opportunities remained unchanged. In choices following these repeats, subjects were sensitive to rewards following freely chosen actions, but ignored environmental information in a manner consistent with a desire to maintain personal control. Choice seeking appears to reflect the intrinsic need for personal control, which competes with extrinsic reward properties and external information to motivate behavior.

Author summary:

Human decisions can often be explained by the balancing of potential rewards and punishments. However, some research suggests that humans also prefer opportunities to choose, even when these have no impact on future rewards or punishments. Thus, opportunities to choose may be intrinsically motivating, although this has never been experimentally tested against alternative explanations such as cognitive dissonance or exploration. We conducted behavioral experiments and used computational modelling to provide compelling evidence that choice opportunities are indeed intrinsically rewarding. Moreover, we found that human choice preference varied according to individual risk attitudes, and expressed a need for personal control that competes with maximizing reward intake.

Preference for choice has been observed in humans(1–6) as well as other animals including rats(7), pigeons(8) and monkeys(9,10). This free-choice premium can be behaviorally measured by having subjects perform trials in two stages: a decision is first made between the opportunity to choose from two terminal actions (*free*) or to perform a mandatory terminal action (*forced*) in the second stage(7). Although food or fluid rewards follow terminal actions in non-human studies, choice preference in humans can be elicited using hypothetical outcomes that are never obtained(3,11). Thus, choice opportunities appear to possess or acquire value in and of themselves. It may be that choice has value because it represents an opportunity to exercise control, which is itself intrinsically rewarding(1,4,12). Personal control is central in numerous psychological theories, where constructs such as autonomy(13,14), controllability(15,16), personal causation(17), effectance(18), perceived behavioral control(19) or self-efficacy(15) are key for motivating behaviors that are not economically rational or easily explained as satisfying basic drives such as hunger, thirst, sex, or pain avoidance(20).

There are alternative explanations for choice seeking. For example, subjects may prefer choice because they are curious and seek information(21,22), or they wish to explore potential outcomes to eventually exploit their options(23), or because they seek variety to perhaps reduce boredom(24) or keep their options open(3). By these accounts, however, the expression of personal control is not itself the ends, but rather a means for achieving an objective that once satisfied reduces choice preference. For example, choice preference should decline when there is no further information to discover in the environment, or after uncertainty about reward contingencies have been satisfactorily resolved.

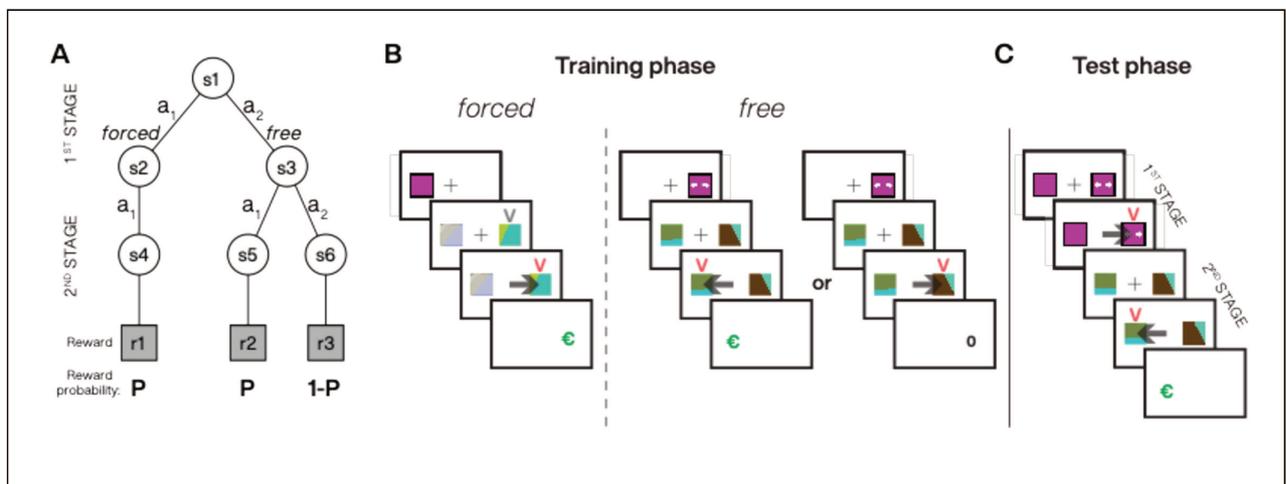
Choice seeking may also arise due to selection itself altering outcome representations. Contexts signaling choice opportunities may acquire distorted value through choice-induced preference change(25). By this account, deciding between equally valued terminal actions generates cognitive dissonance that is resolved by post-choice revaluation favoring the chosen action(25,26). This renders the free option more valuable than the forced option since revaluation only occurs for self-determined actions(27,28). Alternatively, subjects may develop distorted outcome representations through a process related to the winner's or optimizer's curse(29), whereby optimization-based selection upwardly biases value estimates for the chosen action. One algorithm subject to this bias is Q-learning(30), where action values are updated using the maximum value to approximate the maximum expected value. In a two-stage task, the free action value is biased upwards due to considering only the best of two possible future actions, while the forced action value remains unbiased since there is only one possible outcome(31). Again, the expression of personal control is not itself the ends for these selection-based accounts, and both predict that choice preference should be reduced when terminal rewards associated with the free option are clearly different.

Data from prior studies does not arbitrate between competing explanations for choice-seeking. Here, we used behavioral manipulations and computational modelling to explore the factors governing human preference for choice. In the first experiment, we altered the reward contingencies associated with terminal actions in order to rule out curiosity, exploration, variety-seeking, and selection-based explanations for choice seeking. In the second experiment, we used a titration procedure to measure the value of choice

relative to an extrinsic reward available in the environment (i.e., money). We then used reinforcement learning models to show that optimistic learning (considering the best possible future outcome) was insufficient to explain individual variability in choice seeking. Rather, subjects adopted different *decision attitudes*, the desire to make or avoid decisions independent of the outcomes(11), which were balanced against differing levels of risk sensitivity. Finally, in the third experiment, we sought to test whether choice preference was motivated by personal control beliefs. We manipulated the perceived controllability of the task and found that subjects' willingness to repeat a free choice was not affected by the lack of objective controllability over reward outcome. Importantly, subjects were sensitive to past rewards only in trials where state outcomes could be attributed to self-determined choice, and ignored rewards on trials where there was an apparent loss of control. Together, our results support the hypothesis that human preference for choice opportunities derives from the intrinsic motivation for personal control.

## Results:

Subjects performed repeated trials with a two-stage structure (Fig. 1). In each trial, subjects made a 1<sup>st</sup>-stage choice between two options defining the 2<sup>nd</sup>-stage: the opportunity to choose between two fractal targets (*free*) or performing an obligatory selection of another fractal target (*forced*). Extrinsic rewards (€) were delivered only for terminal (i.e., 2<sup>nd</sup>-stage) actions. If subjects chose the *forced* option, the computer always selected the same fractal target for the subjects. If subjects chose the *free* option, they had to choose between two fractal targets associated with two different terminal states. We fixed reward contingencies in blocks of trials, and used unique fractal targets for each block. We divided each block into an initial training phase (Fig. 1B) followed by a test phase (Fig. 1C) to ensure that the subjects learned the associations between the different fractal targets and extrinsic reward probabilities.



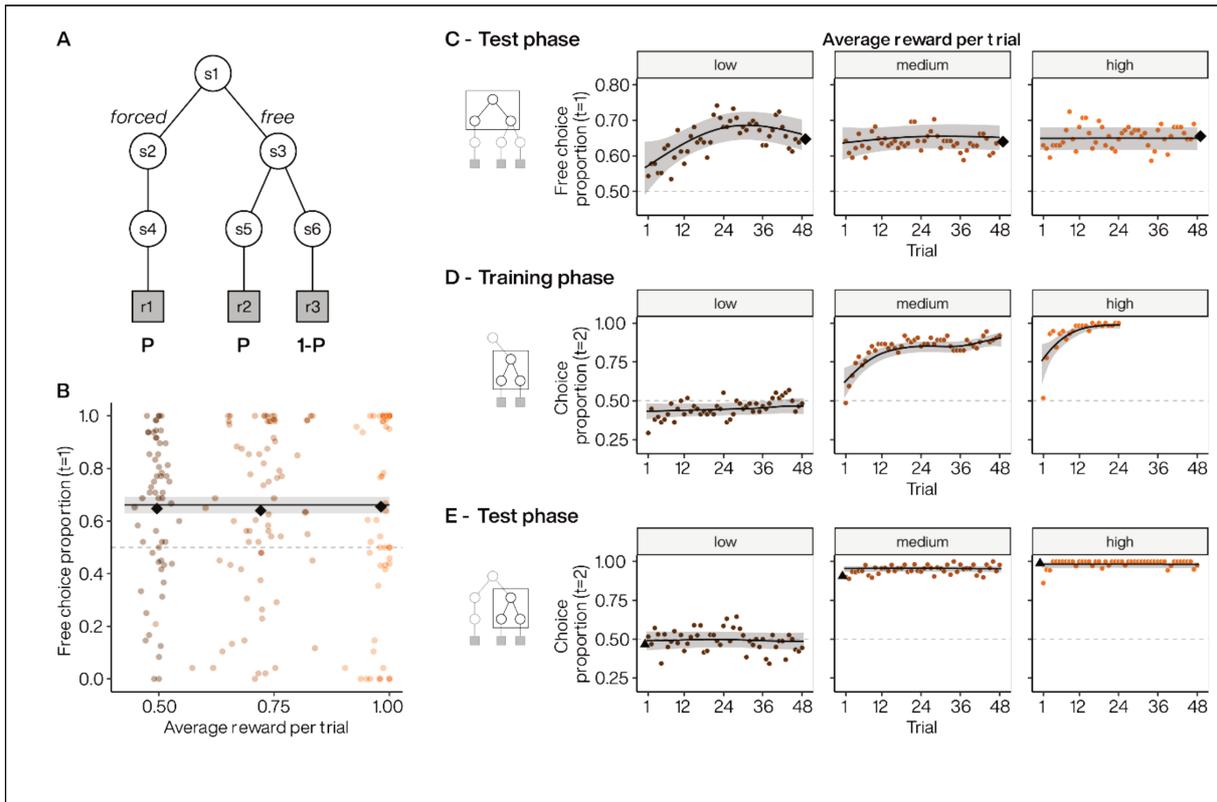
**Figure 1.** Two-stage task structure. **A.** State diagram illustrating the 6 possible states (s), actions (a) and associated extrinsic reward probabilities (e.g.,  $P = 0.5, 0.75$  or  $1$  for blocks 1 to 3, respectively); s2 and s3 were represented by two different 1<sup>st</sup>-stage targets (e.g., colored squares with or without arrows for *free* and *forced* trials, respectively) and s4 to s6 were associated to three different 2<sup>nd</sup>-stage targets (fractals). **B.** Sequence of events during the training phase where the subjects learned the contingencies between the fractal targets and their reward probabilities ( $P$ ) associated with the *forced* (no choice) and *free* (choice available) options. When training the reward contingencies associated with the *forced* option, subjects' actions in the 2<sup>nd</sup>-stage had to match the target indicated by a grey V-shape and was always the same (s4). When training the reward contingencies associated with the *free* option, no mandatory target is present at the 2<sup>nd</sup>-stage (s5 or s6 can be chosen) but one of the targets is more rewarded when  $P > 0.5$ . Black arrows represent the selection of the target by the subject. **C.** Sequence of events during the test phase: subjects first decided between the *free* or *forced* option and then experienced the associated 2<sup>nd</sup>-stage. Rewards, when delivered, were represented by a large green euro symbol (€).

### Free choice preference across different extrinsic reward probabilities

In experiment 1, we varied the overall expected value by varying the probability of extrinsic reward delivery ( $P$ ) across different blocks of trials. These probabilities ranged from 0.5 to 1 across the blocks (i.e., low to high), and the programmed probabilities in *free* and *forced* 2<sup>nd</sup>-stage rewards were equal (Fig. 2A). For example, in high probability blocks, we set the probabilities of the *forced* terminal action and of one of the *free* terminal actions (a1) to 1, and set the probability of the second *free* terminal action (a2) to 0. Therefore, the maximum expected value was equal for the *free* and *forced* options.

Subjects chose to choose more frequently, selecting the *free* option in 64% ( $n=58$ ) of test trials on average (Fig. 2B). The level of preference did not differ significantly across blocks ( $p = 0.857$ , low = 65%, medium = 64%, high = 66%). We found that subjects immediately expressed above chance preference for the *free* option (Fig. 2C) despite never having actualized 1<sup>st</sup>-stage choices during training. Looking within a block, we found that subjects' preference remained constant across trials in medium and high reward probability

blocks ( $p = 0.22$  and  $0.6823$  for nonlinear smooth by trial deviating from a flat line, respectively; Fig. 2C, middle and right panels). In low probability blocks, subjects started with a lower choice preference that gradually increased to match that observed in the medium and high probability blocks ( $p = 0.0014$  for nonlinear smooth by trial; Fig. 2C left panel). The lower reward probability may have prevented subjects from developing accurate reward representations by the end of the training phase, which may have led to additional sampling of the three 2<sup>nd</sup>-stage targets (two in *free* and one in *forced*) in the beginning of the test phase.



**Figure 2.** Choice preference across different absolute extrinsic reward probabilities. **A.** Experiment 1 task design where maximal extrinsic reward probabilities increased equally across *free* and *forced* options. **B.** Subject preference for *free* option during 1<sup>st</sup>-stage. Colored points indicate individual subject mean choice preference per block, plotted against the obtained average reward. Black diamonds indicate the average of subject means per block. Line indicates the estimated choice preference from a GLMM, with 95% CI. **C.** Dynamics of *free* option preference across test phase blocks for low (left), medium (middle) and high (right) absolute extrinsic reward probabilities. Each point represents the average *free* option preference as a function of trial within a block. Diamonds: as in B. Lines indicate the estimated choice preference from a GAMM, with 95% CI. **D to E.** Dynamics of the selection of the most rewarded 2<sup>nd</sup>-stage targets in *free* option for low (left),

medium (middle) and high (right) during the training (D) and test (E) phases. Note that in left panels, the probability of extrinsic rewards is equal for two 2<sup>nd</sup>-stage targets ( $P=0.5$ ). We labelled the best choice as 1 when  $P > 0.5$ . Triangles represents the final average selection at the end of the training phases. Lines: as in C.

### Second-stage performance following *free* selection

We investigated participants' 2<sup>nd</sup>-stage choices following *free* selection to exclude the possibility that choice preference arose because reward contingencies had not been learned. During the training phase, when  $P > 0.5$ , participants quickly learned to choose the most rewarded fractal targets (at  $P=0.5$ , all fractal targets were equally rewarded) (Fig. 2D). During the test phase, participants continued to select the same targets (Fig. 2E), confirming stable application of learned contingencies ( $p > 0.1$  for nonlinear smooth by trial deviating from a flat line for all blocks).

Choice preference was not explained by subjects obtaining more extrinsic rewards following selection of *free* compared to *forced* options. Obtained reward proportions were not significantly different in the low (following selection of *free* vs. *forced*, 0.516 vs. 0.536,  $p = 0.276$ ) or medium (0.746 vs. 0.762,  $p = 0.322$ ) probability blocks. In contrast, in high probability blocks, subjects received significantly fewer rewards on average after *free* selection than after *forced* selection (0.989 vs. 1,  $p = 0.0016$ ). In this block, reward was fully deterministic, and *forced* selection always led to a reward, whereas *free* selections could lead to missed rewards if subjects chose the incorrect target.

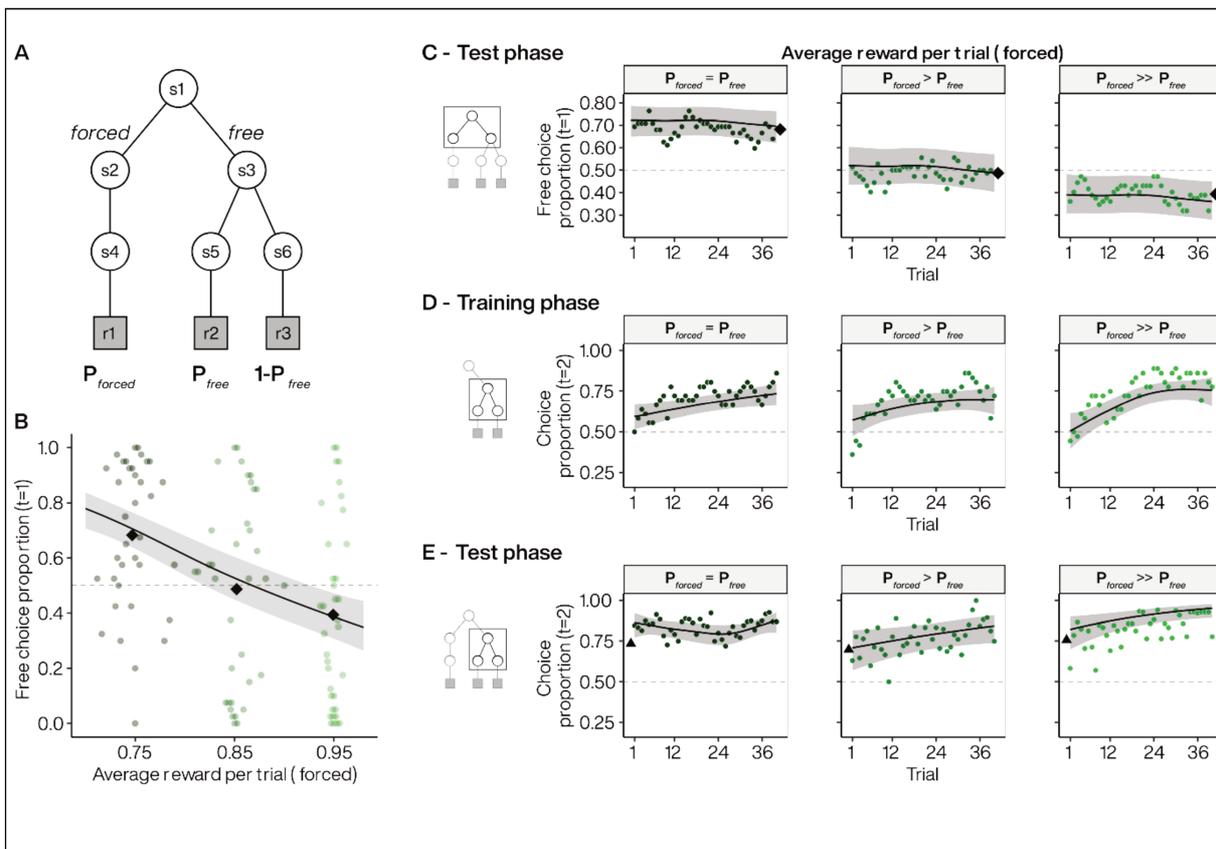
### Trading extrinsic rewards for choice opportunities

Since manipulating the overall expected reward did not alter choice seeking behavior at the group-level, we investigated the effect of changing the relative expected reward between 1<sup>st</sup>-stage options. In experiment 2, we tested a new group of 36 subjects for whom we decreased the objective value of the *free* versus *forced* options. This allowed us to assess the point at which these options were equally valued and potentially reversed to favor the initially non-preferred (*forced*) option (Fig. 3A). Thus, we titrated the value of choice opportunity by increasing the reward probabilities following *forced* selection (block 1:  $P_{forced} = 0.75$ ; block 2:  $P_{forced} = 0.85$ ; block 3:  $P_{forced} = 0.95$ ), while keeping the reward probabilities following *free* selection fixed ( $P_{free|a1} = 0.75$ ,  $P_{free|a2} = 0.25$  for all blocks).

As in experiment 1, we found that subjects preferred choice when the extrinsic reward probabilities of the *free* and *forced* options were equal (block 1: 68% 1<sup>st</sup>-stage choice in favor of *free*; Fig. 3B, dark green). Increasing the reward probability associated with the *forced* option significantly reduced choice preference ( $p = 0.00344$ , Fig. 3B) to 49% (block 2) and 39% (block 3). We estimated the population preference reversal point at  $P_{forced} = 0.88$ , indicating that indifference was obtained on average when the value of the *forced* option was 17% greater than that of the *free*. We found that subjects' preference remained constant across trials when reward probabilities were equal ( $p = 0.875$  for nonlinear smooth by trial; Fig. 3C, left panel). Although reduced overall, the selection of the *free* option also did not vary across trials in blocks 2 and 3 ( $p = 0.737$  and  $0.078$  for nonlinear smooth by trial, respectively). Furthermore, as in experiment 1, subjects acquired preference for the most rewarded 2<sup>nd</sup>-stage targets during the learning phase (Fig. 3D) and continued to express this preference during the test phase in all three blocks (Fig. 3E). Thus, the decrease in choice

preference was not related to failure to learn the reward contingencies during the training phase.

Although decreasing the relative value of the *free* option reduced choice preference, most subjects did not switch exclusively to the *forced* option. Even in block 3, where the *forced* option was set to be rewarded most frequently ( $P_{forced} = 0.95$  versus  $P_{free} = 0.75$ ), 32/36 subjects selected the *free* option in a non-zero proportion of trials. Since exclusive selection of the *forced* option would maximize extrinsic reward intake, continued *free* selection indicates a persistent appetency for choice opportunities despite their diminished relative extrinsic value.



**Figure 3.** Choice preference across different relative extrinsic reward probabilities. **A.** Experiment 2 task design where extrinsic reward probability is always at  $P = 0.75$  for the highly rewarded target in *free* options but vary from 0.75 to 0.95 across 3 blocks for *forced* options. **B.** Subject preference for *free* option during 1<sup>st</sup>-stage. Colored points indicate individual subject mean choice preference per block, plotted against the average reward in *forced* option. Black diamonds indicate the average of subject means per block. Line indicates the

estimated choice preference from a GLMM, with 95% CI. **C.** Dynamics of *free* option preferences across test phase blocks when extrinsic reward probabilities of *forced* options were set at 0.75 (left), 0.85 (middle) and 0.95 (right). Each point represents the average *free* option preference as a function of trial within a block. Diamonds: as in B. Lines indicate the estimated choice preference from a GAMM, with 95% CI. **D to E.** Dynamics of the selection of the most rewarded 2<sup>nd</sup>-stage targets in *free* option when extrinsic reward probabilities of *forced* options are set at 0.75 (left), 0.85 (middle) and 0.95 (right) during the training (D) and test (E) phases. Triangles represents the final average selection at the end of the training phases. Lines: as in C.

### Reinforcement-learning model of choice seeking

We next sought to explain individual variability in choice behavior using a value-based decision-making framework. We first used mixed logistic regression to examine whether rewards obtained from 2<sup>nd</sup>-stage actions influenced 1<sup>st</sup>-stage choices. We found that obtaining a reward on the previous trial significantly increased the odds that subjects repeated the 1<sup>st</sup>-stage selection that ultimately led to that reward ( $p < 0.0001$ , odds ratio rewarded/unrewarded on previous trial:  $1.92 \pm 95\% \text{ CI } [1.40, 2.60]$ ). This suggests that subjects continued to update their extrinsic reward expectation based on experience during the test phase. We therefore leveraged the framework of temporal-difference reinforcement learning (TDRL) to provide a model-based characterization of the emergence of choice preference.

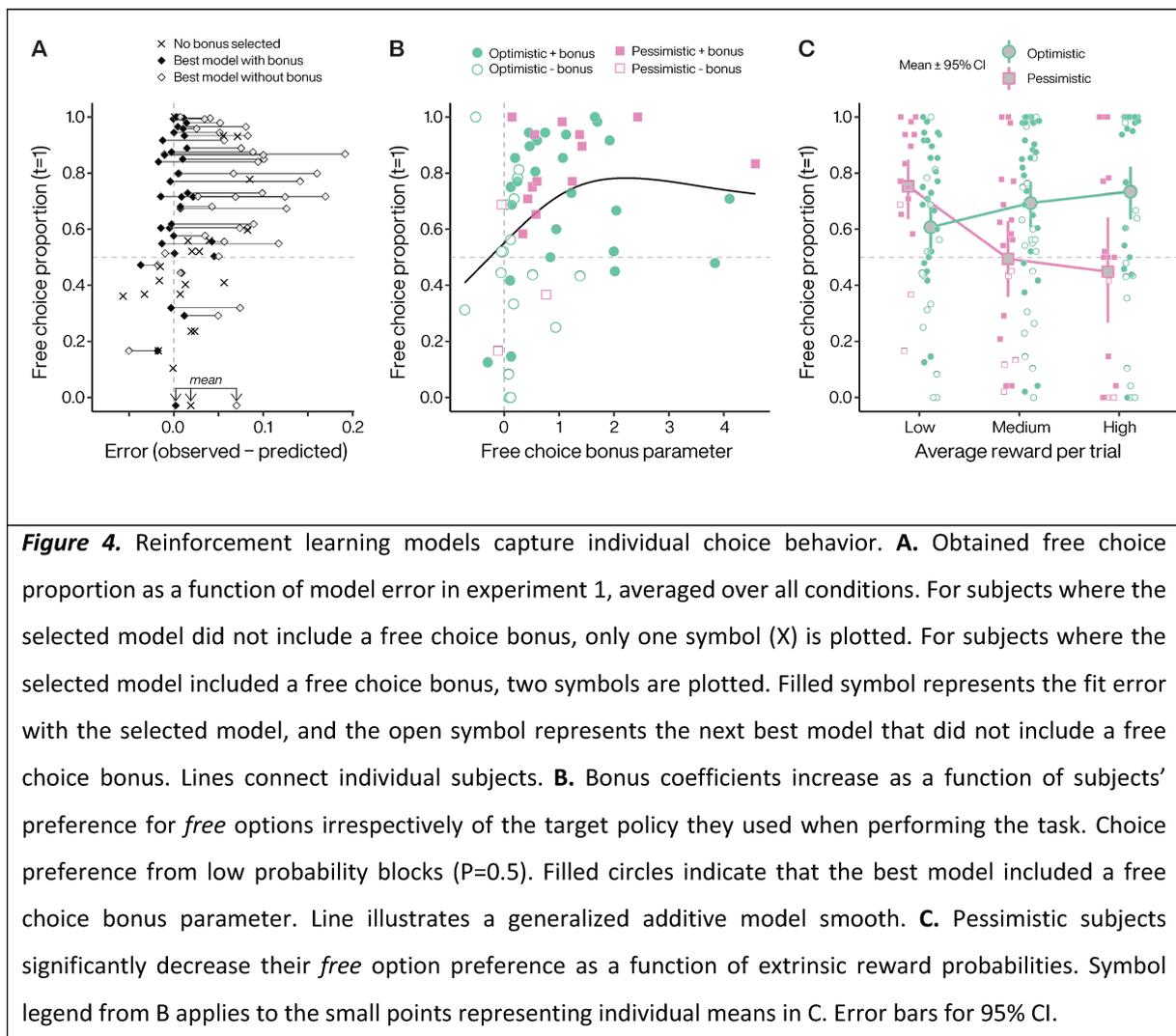
We fitted TDRL models to individual data using two distinct features to capture individual variability across different extrinsic reward contingencies. The first feature was a free choice bonus added to self-determined actions as an intrinsic reward. This can lead to overvaluation of the *free* option via standard TD learning. The second feature modifies the form of the future value estimate used in the TD value iteration, which in common TDRL variants is, or approximates, the best future action value (Q-learning or SARSA with

softmax behavioral policy, respectively). We treated both Q-learning and SARSA together as optimistic algorithms since they are not highly discriminable with our data (Supplementary Fig. 1). We compared this optimism with another TDRL variant that explicitly weights the best and worst future action values (Gaskett's  $\beta$ -pessimistic model(32)), which could capture avoidance of choice opportunities through increased weighting of the worst possible future outcome (pessimistic risk attitude). For example, risk is maximal in the high reward probability block in experiment 1 since selection of one 2<sup>nd</sup>-stage target led to a guaranteed reward (best possible outcome) whereas selection of the other target led to guaranteed non-reward (worst possible outcome).

We found that it was necessary to incorporate the overvaluation of rewards obtained from *free* actions to predict choice preference in experiment 1 (Fig. 4A). Moreover, the magnitude of the bonus was significantly associated with increasing choice preference during the 1<sup>st</sup>-stage of the trials ( $p = 0.0005$  for nonlinear smooth, Fig. 4B). Therefore, optimistic or pessimistic targets alone were insufficient to explain individual choice preference across different extrinsic reward contingencies. We found that a pessimistic target best fitted about 28% (16 of 58) of the subjects in experiment 1. Moreover, most pessimistic subjects (13 of 16) were best fitted with a model including a free choice bonus to balance risk and decision attitudes across reward contingencies. In experiment 1, we introduced risk by varying the difference in extrinsic reward probability for the best and worst outcome following *free* selection. The majority of so-called 'pessimistic subjects' preferred choice when extrinsic reward probabilities were low, but their weighting of the worst possible outcome decreased this preference as risk increased (Fig. 4C, pink). Thus,

pessimistic subjects avoided the *free* option despite rarely or never selecting the more poorly rewarded 2<sup>nd</sup>-stage target during the test phase.

We also fitted the TDRL variants to individual data from experiment 2, and found that a free choice bonus was also necessary to explain choice preference across extrinsic reward contingencies in that experiment. Four subjects (of 36) were best fitted using the  $\beta$ -pessimistic target (see Supplementary Fig. 2) although this may be a conservative estimate since we did not vary risk in experiment 2.



### Influence of action-outcome coherence on choice seeking behavior

We next asked whether choice preference was related to personal control beliefs. To do so, we manipulated the coherence between an action and its consequence over the environment. In experiment 3, we tested the relationship between preference for choice opportunity and the physical coherence of the terminal action by directly manipulating the perceived controllability of 2<sup>nd</sup>-stage actions. We modified the two-stage task to introduce a mismatch between the subject's selection of the 2<sup>nd</sup>-stage target and the target ultimately displayed on the screen by the computer (Fig. 5A). We did this by manipulating the probability that a 2<sup>nd</sup>-stage target selected by a subject would be swapped for the 2<sup>nd</sup>-stage target that had not been selected. That is, on coherent trials, a subject selecting the fractal on the right side of the screen would receive visual feedback indicating that the right target had been selected. On incoherent trials, a subject selecting the fractal on the right side would receive feedback that the opposite fractal target had been selected (i.e., the left target).

To ensure that all other factors were equalized between the two 1<sup>st</sup>-stage choices, we implemented target swaps following both *free* and *forced* selections by adding an additional state to our task (Fig. 5A). In one block of trials, the incoherence was set to 0 and every subject action in the 2<sup>nd</sup>-stage led to a coherent selection of the second target. In the other blocks, we set incoherence to 0.15 or 0.3, resulting in lower perceived controllability between target choice and target selection (e.g., 85% of the time, pressing the left key will select the left target, and in 15% the right target). We set all of the extrinsic reward probabilities associated with the different fractal targets to  $P = 0.75$ . Since all 2<sup>nd</sup>-stage actions had the same expected value, the experiment was objectively uncontrollable

because the probability of reward was independent of all actions(16). Moreover, equal reward probabilities ensured that outcome diversity(33,34), outcome entropy(35), and instrumental divergence(36) did not contribute to choice preference since these were all equal between the *forced* and *free* options.

The same group of participants who performed experiment 2 also performed experiment 3 (n=36). Choice preference was high (70%) in block 1 when coherence was not altered, similar to block 1 from experiment 2 where extrinsic reward was equal between *free* and *forced* options. The only difference between these two blocks was that choosing the *forced* option resulted in the obligatory selection of the same fractal (experiment 2) or one of two fractals randomly selected by the computer (experiment 3), which indicates that subjects' choice preference was not related to action variability per se following *forced* selection. Moreover, we found that choice preference was significantly correlated ( $r = 0.358$ ,  $p = 0.03175$ ) between block 1 of experiments 2 and 3, highlighting a within-subject consistency in choice preference.

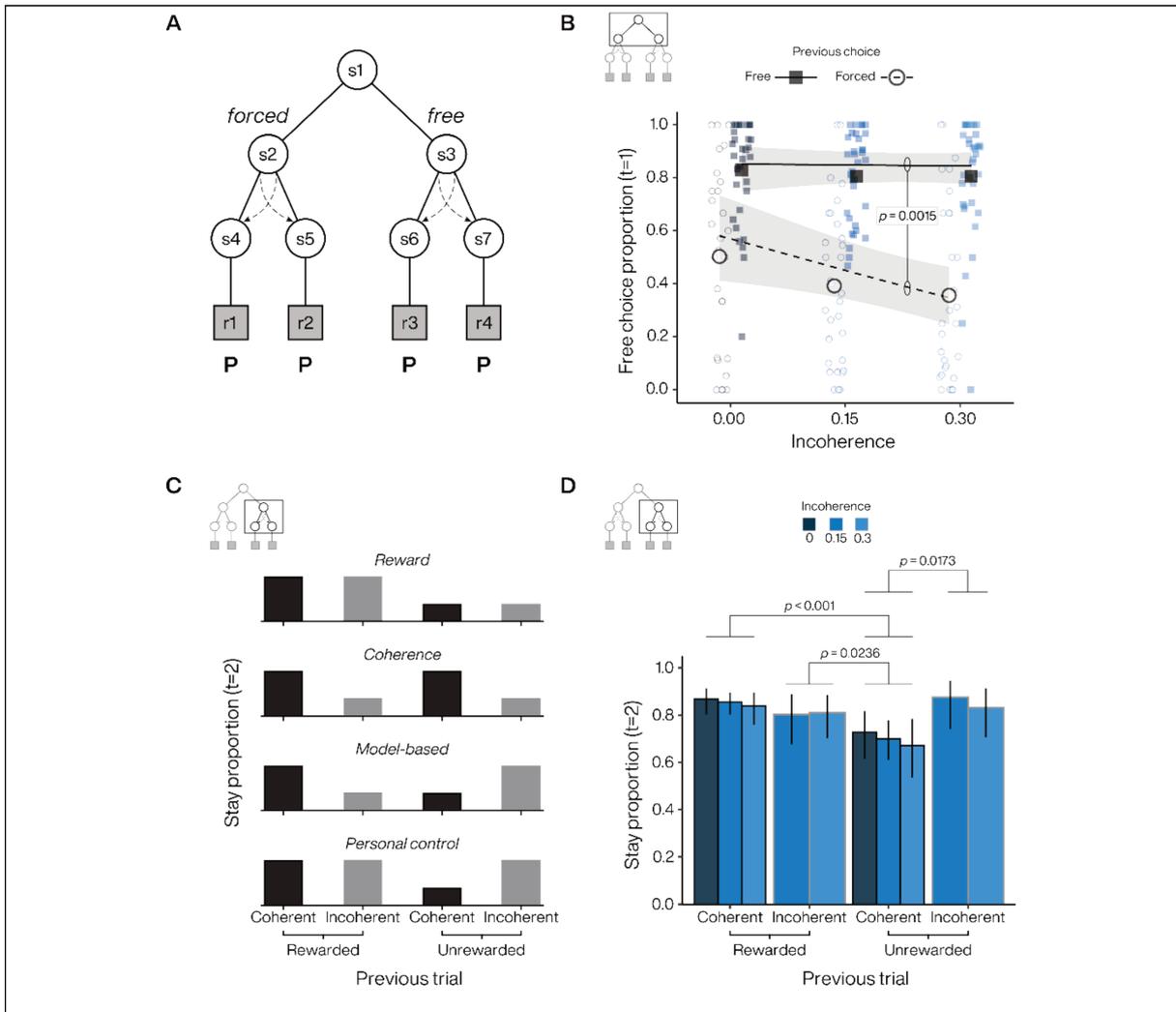
Increasing the incoherence of the 2<sup>nd</sup>-stage actions progressively reduced choice preference (block 2 and 3: 67% and 64% in favor of *free* respectively). As in experiments 1 and 2, choice preference was expressed immediately after the training phase and remained constant throughout the different blocks (Supplementary Fig. 3). We found that the decline in choice preference depended on the 1<sup>st</sup>-stage choice on the previous trial. Specifically, following coherent trials, we found that there was a significant interaction between the previous 1<sup>st</sup>-stage choice (*free* or *forced*) and the degree of incoherence ( $p = 0.0015$ , Fig. 5B). The difference in slopes was due to decreasing propensity to choose the *free* option

following *forced* selection on the previous trial ( $p = 0.0111$ ), with no change in the propensity to choose the *free* option following *free* selection on the previous trial ( $p = 0.8706$ ). Thus, as incoherence increased, subjects tended to stay more with the *forced* option, while maintaining a preference to repeat *free* selections.

The sustained repetition of *free* selections across the different levels of incoherence suggests that subjects may have been seeking to regain control of the environment through self-determined 2<sup>nd</sup>-stage choices. Although the task was objectively uncontrollable since all terminal action-target sequences were associated with the same reward probability, subjects may have developed structure beliefs based on local reward history and target swaps, which could be reflected in 2<sup>nd</sup>-stage patterns of choice. Thus, subjects may have followed a strategy based on reward feedback by repeating only actions associated with a previous reward (illusory maximization of reward intake; Fig.5C, first panel). Alternatively, they could have followed a strategy based on action-outcome incoherence feedback and thus avoided trials associated with a previous target swap (illusory minimization of incoherent states; Fig. 5C, second panel). However, subjects may have also employed another classic strategy known as “model-based” where agents use their (here illusory) understanding of the task structure built from all the information provided by the environment (Fig.5C, third panel)(37). Under this strategy, subjects try to integrate both the reward and target-swap feedback to select the next target in order to maximize reward. For example, an incoherent but rewarded trial would lead to a behavioral switch because the subject has integrated the information provided by the environment (i.e., the target swap induced by the computer), signaling that the other target is actually rewarded (see

second bar on third panel of Fig. 5C). Finally, an alternative strategy could rely on maximizing personal (i.e., internal) control, where the subject is the (illusory) agent of the entire sequence of events (i.e., action-state-reward) and would therefore ignore reward outcomes when they are not associated with the selected action-state (Fig.5C, fourth panel).

Results of the stay behavior during 2<sup>nd</sup>-stage choice following *free* selection suggests that subjects seek personal control when choosing between the different fractal targets (Fig.5D). Indeed, when their action was consistent with the state they were choosing (i.e., the coherent fractal target feedback), they took the reward outcome into account to adjust their behavior on the next trial, either by staying on the same target when the trial was rewarded or by switching to the other one when no reward was delivered. However, subjects were insensitive to the reward outcome during incoherent trials as they maintained the same strategy (staying) during subsequent trials, regardless of whether they were previously rewarded or not. This strategy reflects an attempt to regain personal control over the environment at the expense of the task goal of maximizing reward intake.



**Figure 5.** Perceived controllability alters choice preference. **A.** Task design where a 7<sup>th</sup> state, associated to the *forced* options, has been added to manipulate the incoherence in both *free* and *forced* options. At incoherence = 0, the visual feedback presented to the subject matches their selected target. Extrinsic reward probabilities set at  $P=0.75$  for all the 2<sup>nd</sup>-stage targets. **B.** First-stage probabilities to stay or switch in free options after a *free* and *forced* trial respectively, as a function of the different incoherence blocks. **C.** Second-stage stay probabilities for the different action-state-reward trial type. Each sub-panels represent a putative strategy followed by the subject. **D.** Estimated 2<sup>nd</sup>-stage stay probabilities. Error bars for 95% CI. P-values are displayed for significant pairwise comparisons and adjusted for multiple comparisons.

## Discussion

Animals prefer situations that offer more choice to those that offer less. Although this behavior can be reliably measured using the two-stage task design popularized by Voss and Homzie(7), their conclusion that choice has intrinsic value is open to debate. To rule out alternative explanations for choice-seeking, we performed three experiments in which we clearly separated learning of reward contingencies from testing of choice preference. Our experiments point to a sustained preference for choice opportunities that express an intrinsic need for personal control. Moreover, this need may compete with potentially valuable information for maximizing outcomes or even extrinsic rewards per se.

In the first and second experiments, we varied the reward probabilities associated with terminal actions following *free* and *forced* selection. Consistent with previous studies, subjects preferred the opportunity to make a choice when expected rewards were equal between terminal actions ( $P = 0.5$ ). Surprisingly, subjects also preferred choice when we increased the value difference between terminal actions in the *free* option, while keeping the *maximum* expected reward equal in the free and forced options ( $P > 0.5$ ). This sustained preference for choice is potentially economically suboptimal since making a free choice carries the risk of making an error leading to lowered reward intake. The persistence of this preference for free choice even when reward delivery was deterministic ( $P = 1$ ), makes it unlikely that this preference was due to an underestimation of forced trials due to poor learning of reward contingencies.

Subjects appeared to have understood the reward contingencies as evidenced by

their consistent preference for the highest-rewarded 2<sup>nd</sup>-stage fractal, which was acquired during the training phase and expressed during the test phase. This stable 2<sup>nd</sup>-stage fractal selection, together with the immediate expression and maintenance of 1<sup>st</sup>-stage choice preference, renders unlikely accounts based on curiosity, exploration or variety seeking since varying the probability of rewards did not modulate choice preference about two third of the subjects (i.e., optimistic subjects).

Selection-based accounts also have trouble explaining the pattern of results we observed. The idea that post-choice revaluation specifically inflates expected outcomes after choosing the free option can explain choice-seeking when all terminal reward probabilities are equal. However, post-choice revaluation cannot explain choice preference when the terminal reward probabilities in the *free* option clearly differ from one another, since revaluation appears to occur only after choosing between closely valued options(28,38). That is, there is no cognitive dissonance to resolve when reward contingencies are easy to discriminate, and no preference for choice should be observed when the choice is between a surely (i.e., deterministically) rewarded action and a never rewarded action. The existence of choice preference in the deterministic condition ( $P = 1$ ) also cannot be explained by an optimistic algorithm such as Q-learning, since the maximum action value is equal to the maximum expected value, and the value of the free option is not biased upwards under repeated sampling(31).

Although standard Q-learning could not capture variability across different terminal reward probabilities, we found that combining two novel modifications to TDRL models was able to do so. The first feature was a free choice bonus—a fixed value added to all

extrinsic rewards obtained through free actions—that can lead to overvaluation of the free option via standard TD learning. This bonus implements Beattie and colleagues' concept of *decision attitude*, the desire to make or avoid decisions independent of the outcomes(11). The second feature modifies the form of the future value estimate in the TD value iteration. Zorowitz and colleagues(31) showed that replacing the future value estimate in Q-learning with a weighted mixture of the best and worst future action values(32) can generate behavior ranging from aversion to preference for choice. The mixing coefficient determines how optimism (maximum of future action values, total risk indifference) is tempered by pessimism (minimum of future action values, total risk aversion). In experiment 1, we found that 28% of subjects were best fitted with a model incorporating pessimism, which captured a downturn in choice preference with increasing relative value difference between the terminal actions in the *free* option. Importantly however, individual variability in the TD future value estimates alone did not explain the pattern of choice preference across target reward probabilities, and a free choice bonus was still necessary for most subjects. Thus, the combination of both a free choice bonus (decision attitude) and pessimism (risk attitude) was key for explaining why some individuals shift from seeking to avoiding choice. This was unexpected because the average choice preference in experiment 1 was not significantly different across reward manipulations, highlighting the importance of examining behavior at the individual level. Here, we examined risk using the difference between the best and worst outcomes as well as relative value using probability (see(39)). It may be the case that variability is also observed in how individuals balance the intrinsic rewards with other extrinsic reward properties that can influence choice preference, such

as reward magnitude(39).

Why are choice opportunities highly valued? It may be that choice opportunities have acquired intrinsic value because they are particularly advantageous in the context of the natural environment in which the learning system has evolved. Thus, choice opportunities might be intrinsically rewarding because they promote the search for states that minimize uncertainty and variability, which could be used by an agent to improve their control over the environment and increase extrinsic reward intake in the long run(40,41). Developments in reinforcement learning and robotics support the idea that both extrinsic and intrinsic rewards are important for maximizing an agent's survival(42–44). Building intrinsic motivation into RL agents can promote the search for uncertain states and facilitate the acquisition of skills that generalize better across different environments, an essential feature for maximizing an agent's ability to survive and reproduce over its lifetime, i.e. its evolutionary fitness(42).

The intrinsic reward of choice may be a specific instance of more general motivational constructs such as autonomy(13,14), personal causation(17), effectance(18), learned helplessness(45), perceived behavioral control(19) or self-efficacy(15), which are key for motivating behaviors that are not easily explained as satisfying basic drives such as hunger, thirst, sex, or pain avoidance(20). Common across these theoretical constructs is that control is intrinsically motivating only when the potential exists for agents to determine their own behavior, which when realized can give rise to a sense of agency and, in turn, strengthens the belief in the ability to exercise control over one's life(46). Thus, individuals with an *internal* locus of control tend to believe that they, as opposed to external

factors such as chance or other agents, control the events that affect their lives. Crucially, the notion of locus of control makes specific predictions about the relationship between preference for choice—choice being an opportunity to exercise control—and the environment: individuals should express a weaker preference for choice when the environment is adverse, stressful or unpredictable(47). This prediction is consistent with what is known about the influence of environmental adversity on control externalization: individuals exposed to greater environmental instabilities tend to believe that external and uncontrollable forces are the primary causes of events that affect their lives, as opposed to themselves(48). In other words, one would expect belief in one's ability to control events, and thus preference for choice, to decline as the environment is perceived as increasingly unpredictable.

In our third experiment, we sought to test whether it was specifically a belief in personal control that motivated subjects, by altering the perceived controllability of the task environment. To do so, we introduced a novel change to the two-stage task where in a fraction of trials subjects experienced random swapping of the terminal states (fractals). Thus, subjects were subjected to trials where the terminal state was incoherent with their choice, and thus experienced alterations in their ability to predict the state of the environment following their action. Incoherence occurred with equal probability following free and forced actions in order to equate for any value associated with swapping itself. We found a significant reduction in the propensity to switch from forced to free choice following action-target incoherence, suggesting that altering the perceived controllability of the task causes choice to lose its attractiveness. This reduction in choice preference

following incoherent trials is reminiscent of a form of locus externalization, and is consistent with the notion that choice preference is driven by a belief in one's personal control. In this experiment, we focused on the value of personal control, and therefore held other decision variables such as outcome diversity(33,34), outcome entropy(35), and instrumental divergence (36,49). Further experiments are needed to understand how these variables interact with personal control in the acquisition of potential control over the environment.

Interestingly, when subjects selected the *free* option, the subsequent choice was sensitive to the past reward when the terminal state (the selected target) was coherent and the reward could therefore be attributed to the subject's action. In contrast, subjects' choices were insensitive to past reward when the terminal state was incoherent. Furthermore, the probability of sticking with the previous 2<sup>nd</sup>-stage choice following incoherent trials, whether rewarded or not, was not different from the probability of sticking with the previously *rewarded* 2<sup>nd</sup>-stage choice following coherent trials. Thus, subjects appeared to ignore information about action-state-reward contingencies that was externally derived, and instead appeared to double down by repeating their past choice as if they sought to maintain or regain personal control. This behavior is consistent with many observations suggesting that when individuals experience situations that threaten or reduce their personal control, they implement compensatory strategies to restore their perceived control to its baseline level(50,51).

Computationally, however, this compensatory strategy is at odds with a pure model-based strategy(37), where an agent could exploit information about action-state-reward

contingencies whether it derived from their own choices (internal control) or from the environment (external control). Rather, it is consistent with work showing that choice-seeking could emerge when self-determined actions amplify subsequent positive reward prediction errors(5,52), and more generally with the notion that events are processed differently depending on individuals' beliefs about their own control abilities. Thus, positive events are amplified only when they are believed to be within one's personal control, whereas they are treated impartially when they are not(52), or when they come from an uncontrollable environment(53).

Together, our results suggest that choice seeking may represent one critical facet of intrinsic motivation and is associated with the desire of personal control. They also suggest that the need for personal control can compete with maximization of extrinsic reward provided by externally driven actions. Indeed, subjects favor positive outcomes associated to internally driven action even if reward rate is lower than for action performed under the instruction of an external agent. In general, the perception of being in personal control could then account for several aspects of our daily life such as enjoyment during game(54) or motivation to perform demanding task(55). Since our results shown inter-individual difference, it would be nonetheless important in the future to phenotype subject behaviors during choice-making to investigate how these individual traits can explain attitude difference when facing decision and their consequences, as exemplified by the variety of attribution and explanation styles of individuals in the general population(56,57).

**Materials and Methods:**

**Participants.** Ninety-four healthy individuals (mean age = 30  $\pm$ SD 7.32 years, 64 females) responded to posted advertisements and were recruited to participate in this study. Relevant inclusion criteria for all participants were being fluent in French, not treated for neuropsychiatric disorders, having no color vision deficiency and being aged between 18 and 45 years old. Out of these 94 subjects, 58 participated to experiment 1 and 36 to experiments 2-3. We gave subjects 40 euros for participating. The sample size was chosen based on previous studies that used similar two-alternative decision making tasks(52,58,59).

**Ethics statement.** The local ethics committee (Comité d'Évaluation Éthique de l'Inserm) approved the study (2019-CER2-MR-004). Participants gave written informed consent during inclusion in the study, which was carried out in accordance with the declaration of Helsinki (1964; revised 2013).

**General procedure.** The paradigm was written in Matlab, using the Psychophysics Toolbox extensions(60,61). It was presented on a 24 inches screen (1920 x 1080 pixels, aspect ratio 16:9). Subjects seat ~57 cm from the center of the monitor. Our behavioral task design was designed as a value-based decision paradigm. All participants received written and oral instructions. They were told that the goal of the task was to gain the maximum number of rewards (a large green euro). They were informed about the differences between the different trial types and that the extrinsic reward contingencies experienced during the training phases remained identical during the test phases. After instructions, participants

received a pre-training session of a dozen trials (pre-train and pre-test phases) in order to familiarize them with the task design and the keys they would have to press.

In our experiments, subjects performed repeated trials with a two-stage structure. In the 1<sup>st</sup>-stage they made an initial decision about what could occur in the 2<sup>nd</sup>-stage. Selecting the *free* option led to a subsequent opportunity to choose and selecting the *forced* option led to an obligatory computer-selected action. In the 2<sup>nd</sup>-stage, we presented subjects with two fractal images, one of them being more rewarded following *free* selection in experiment 1 (except for P=0.5) and experiment 2. In experiments 1 and 2, the computer always selected the same fractal target following *forced* selection. Experiment 3 all fractal targets were equally rewarded and the computer randomly selected one of the two fractal targets following *forced* selection (50%). Following *forced* selection, the target to select with a key press was indicated by a grey V-shape above the target. Pressing the other key on this trial type did nothing and the computer waited for the correct key press to proceed further in the trial sequence. Either at the 1<sup>st</sup>- or 2<sup>nd</sup>-stage, after the subject's selection of the target, a red V-shape appears immediately after above the target to indicate the one they had selected (in experiment 3 blocks this red V-shape remains 250ms on the screen and eventually jumped with the target, see below).

**Experimental conditions.** In experiment 1, fifty-eight subjects performed trials where the maximal reward probabilities were matched following *free* and *forced* selection. We varied the overall expected value across different blocks of trials, each of them being associated to

different programmed extrinsic reward probabilities ( $P$ ). Forty-eight subjects performed a version with 3 blocks (experiment 1a) with different extrinsic reward probabilities ranging from 0.5 to 1 (block 1:  $P_{forced} = P_{free} = 0.5$ ; block 2:  $P_{forced} = 0.75$ ,  $P_{free|a1} = 0.75$ ,  $P_{free|a2} = 0.25$ ; block 3:  $P_{forced} = 1$ ,  $P_{free|a1} = 1$ ,  $P_{free|a2} = 0$ ; where  $a1$  and  $a2$  represent the two possible key presses associated with the fractal targets). Ten additional subjects performed the same task with 4 different blocks (experiment 1b) associated to extrinsic reward probabilities also ranging from 0.5 to 1 ( $P = 0.5$  or  $0.67$  or  $0.83$  or  $1$  from block 1 to 4 respectively.) We did not observe any substantial difference between these two subject groups, and pooled them for analyses.

Experiment 2 was similar to experiment 1 (six states) except programmed extrinsic reward associated with the *forced* option were higher than than the *free* option in two out of three blocks ( $P_{forced} = 0.75, 0.85$  or  $0.95$ ). Reward probabilities following *free* selection did not change across the three blocks ( $P_{free|a1} = 0.75$ ,  $P_{free|a2} = 0.25$ )

Experiment 3 consisted of a 7-state version of the two-stage task. Here, we manipulated the coherence between the subject selection of a 2<sup>nd</sup>-stage (fractal) target and the target ultimately displayed on the screen by the computer. Irrespectively of the target finally selected by the computer or the subjects, the extrinsic reward probability associated to all the 2<sup>nd</sup>-stage targets in *free* and *forced* trials was set at  $P=0.75$ . Importantly, adding the 7<sup>th</sup> state in this last task version allowed the computer to swap the fractal 2<sup>nd</sup>-stage targets following both *free* and *forced* selection. Thus, subjects did not perceive the weak coherence as a feature specific to the *free* condition.

We associated unique fractal targets with each action in the 2<sup>nd</sup>-stage, and a new set

was used for each block in all experiments. Colors of the 1<sup>st</sup>-stage targets were different between experiments. Positive or negative reward feedback, as well as the side of the 1<sup>st</sup>-stage and 2<sup>nd</sup>-stage target positions, were pseudo-randomly interleaved on the right or left of screen center. Feedback was represented by the presentation (reward) or not (non-reward) of a large green euro image.

In experiment 1, when  $P < 1$ , participants performed a minimum of 48 trials per block in the training phases (*forced* and *free*) and the test phases. For  $P = 1$ , participants performed a minimum 24 trials for training phases (*forced* and *free*) and 48 trials for test phase. The order of the blocks were randomly interleaved. In experiments 2 and 3 they performed a minimum of 40 trials for each block. Here, subjects started by performing experiment 3 followed by experiment 2. This was to ensure that the value of *free* trials was not devalued by experiment 2 (titration) when performing experiment 3. In experiment 3, subjects always started by the block with no target swaps (incoherence = 0), and in experiment 2 by the block with equal extrinsic reward probability (equivalent to the block  $P = 0.75$  of experiment 1). All the other blocks were randomly interleaved.

**Trial structure.** During the training phase, for each trial, a first fixation point appeared in the center of the screen for 500ms, followed by the one of the first two targets of the different trial types for an additional 750ms, either (*forced* or *free*) to the left or right of the fixation point ( $\sim 11^\circ$  from the center of the screen on the horizontal axis,  $3^\circ$  wide). Immediately after, the first target was turned off and two fractal targets appeared at the same eccentricity than the first target to the left and right of the fixation point. The subjects

could then choose by themselves or had to match the target (depending on the trial type) using a key press (left or right arrow keys for left and right targets, respectively). After their selection, a red V-shape appeared for about 1000ms above the selected target (trace epoch). Note that in experiment 3, the V-shape was initially light red and turned on for 250ms above the actual fractal target selected by the subjects. It was then turned in dark red for 750ms. If the trial was incoherent, after 250ms, the red V-shape jumped and thus reappeared simultaneously with the other target on the other side of the screen also for 750ms. Finally, the fixation point was turned-off and the outcome was displayed during 750ms before the next trial. For the test phase, the timing was equivalent except for the decision epoch related to the first stage where participants could choose their favorite trial type (*free* and *forced* targets positioned randomly, left or right) after 500ms of fixation point presentation. When a selection was made, the first target remained for 500ms, associated to a red V-shape over the selected 1<sup>st</sup>-stage target – indicating their choice. The second stage started with a 500ms epoch where only the fixation point was presented on the screen, followed by the fractal target presentation. During the first and second action epochs, no time pressure was imposed on subjects to make their choice, but if they pressed one of the keys during the first 100ms after target presentation ('early press'), a large red cross was displayed in the center of the screen for 500ms and the trial was repeated.

**Computational modelling.** We fitted individual subject data with variants of temporal-difference reinforcement learning (TDRL) models. All models maintained a look-up table of state-action value estimates ( $Q(s, a)$ ) for each unique target and each action across all

conditions within a particular experiment. State-action values were updated at each stage ( $t \in \{1,2\}$ ) within a trial according to the prediction error measuring the discrepancy between obtained and expected outcomes:

$$\delta_t = r_{t+1} + Z(s_{t+1}, a_{t+1}) - Q(s_t, a_t)$$

where  $r_{t+1} \in \{0,1\}$  indicates whether the subject received an extrinsic reward, and  $Z(s_{t+1}, a_{t+1})$  represents the current estimate of the state-action value. The latter could take three possible forms:

$$Z(s_{t+1}, a_{t+1}) = \begin{cases} Q(s_{t+1}, a_{t+1}) & \text{SARSA} \\ \max_{a'} Q(s_{t+1}, a') & \text{Q-learning} \\ \beta \cdot \max_{a'} Q(s_{t+1}, a') + (1 - \beta) \cdot \min_{a'} Q(s_{t+1}, a') & \beta\text{-pessimistic} \end{cases}$$

Although Q-learning and SARSA variants differ in whether they learn off- or on-policy, respectively, we treated both of these algorithms as optimistic. Q-learning is strictly optimistic by considering only the best future state-action value, whereas SARSA can be more or less optimistic depending on the sensitivity of the mapping from state-action value differences to behavioral policy. We compared Q-learning and SARSA variants with a third state-action value estimator that incorporates risk attitude through a weighted mixture of the best and worst future action values (Gaskett's  $\beta$ -pessimistic model(32)). As  $\beta \rightarrow 1$  the pessimistic estimate of the current state-action value converges to Q-learning.

The prediction error was then used to update all state-action values according to:

$$Q(s_{t+1}, a_{t+1}) \leftarrow Q(s_{t+1}, a_{t+1}) + \alpha \cdot \delta_t$$

where  $\alpha \in [0,1]$  represents the learning rate.

We tested whether a free choice bonus could explain choice preference by modifying the obtained reward as follows:

$$r_{t+1} = r_{t+1}^{\text{extrinsic}} + \rho$$

where  $\rho \in (-\text{inf}, +\text{inf})$  is a scalar parameter added to any extrinsic reward following any action taken following selection of the free option.

Free actions at each stage were generated using a softmax policy as follows:

$$\pi(s, a^1) = \frac{\exp(Q(s, a^1)/\tau)}{\exp(Q(s, a^1)/\tau) + \exp(Q(s, a^2)/\tau)}$$

where increasing the temperature,  $\tau \in [0, +\text{inf})$ , produces a softer probability distribution over actions. The forced option, on the other hand, always led to the same fixed action. We used a softmax behavioral policy for all TDRL variants, and in the context of our task, the Q-learning and SARSA algorithms were often similar in explaining subject data, so we treated them together in the main text (Supplementary Fig. 1).

We also tested the possibility that subjects exhibited tendencies to alternate or perseverate following free or forced actions. We implemented this using a stickiness parameter that modified the policy as follows:

$$\pi(s, a^1) = \frac{\exp[(Q(s, a^1) + \kappa \cdot C_t(s, a^1))/\tau]}{\exp[(Q(s, a^1) + \kappa \cdot C_t(s, a^1))/\tau] + \exp[(Q(s, a^2) + \kappa \cdot C_t(s, a^2))/\tau]}$$

where the  $\kappa \in (-\text{inf}, +\text{inf})$  parameter represents the subject's tendency to perseverate, and  $C_t(s, a)$  is a binary indicator for which fractal and action was chosen on the previous trial.

We independently combined the free parameters to produce a family of model fits for each subject. We allowed the learning rate ( $\alpha$ ) and softmax temperature ( $\tau$ ) to differ for each of the two stages in a trial. We therefore fitted a total of 48 models (3 estimates of current state-action value [SARSA, Q,  $\beta$ -pessimistic]  $\times$  presence or absence of free choice bonus [ $\rho$ ]  $\times$  2- vs 1-learning rate [ $\alpha$ ]  $\times$  2- vs 1-temperature [ $\tau$ ]  $\times$  presence or absence of

stickiness [ $\kappa$ ]).

**Parameter estimation and model comparison.** We fitted model parameters using maximum a posteriori (MAP) estimation using the following priors:

$$\alpha \sim \text{beta}(\text{shape1}=1.1, \text{shape2}=1.1)$$

$$1/\tau \sim \text{gamma}(\text{shape}=1.2, \text{scale}=5)$$

$$\beta \sim \text{beta}(\text{shape1}=1.1, \text{shape2}=1.1)$$

$$\rho \sim \text{norm}(\text{mean}=0, \text{sd}=1)$$

$$\kappa \sim \text{norm}(\text{mean}=0, \text{sd}=1).$$

We based hyperparameters for  $\alpha$  and  $1/\tau$  on Daw and colleagues (37). We used the same priors and hyperparameters for all models containing a particular parameter. We used limited-memory quasi-Newton algorithm (L-BFGS-B) to numerically compute MAP estimates, with  $\alpha$  and  $\beta$  bounded between 0 and 1 and  $1/\tau$  bounded below at 0. For each model, we selected the best MAP estimate from 10 random parameter initializations.

For each model for each subject, we fitted a single set of parameters to both training and test data across conditions. We initialized state-action values to zero at the beginning of the training phase for each condition. Data from the training phase consisted of 2<sup>nd</sup>-stage actions and rewards, but we also presented subjects with the 1<sup>st</sup>-stage cues corresponding to the condition being trained (forced or free). Therefore, we fitted the TDRL models assuming that the state-action values associated with the 1<sup>st</sup>-stage fractals also underwent learning during the training phase, and that these backups continued into the test phase, where subjects actually made 1<sup>st</sup>-stage decisions. That is, we initialized the state-action values during the test phase with the final state-action values during the training phase.

We used Schwarz weights to compare models, which provides a measure of the strength of evidence in favor of one model over others and can be interpreted as the probability that a model is best in the Bayesian Information Criterion (BIC) sense(62). We calculated weights for each model as:

$$w_i(\text{BIC}) = \frac{\exp(-\Delta_i(\text{BIC})/2)}{\sum_{k=1}^K \exp(-\Delta_k(\text{BIC})/2)}$$

so that  $\sum w_i(\text{BIC}) = 1$ . We selected the model with the maximal Schwarz weight for each subject.

In order to verify that we could discriminate different state-action value estimates and how accurately we could estimate parameters, we performed model and parameter recovery analyses on simulated datasets (Supplementary Fig. 1).

**Statistical analyses.** We used generalized linear mixed models (GLMM) to examine differences in choice behavior. When the model did not include trial-specific information (e.g., reward on the previous trial), we aggregated data to the block level. Otherwise, we used choice data at the trial level. We included random effects by subject for all models (random intercepts and random slopes for the variable manipulated in each experiment; maximal expected value, relative expected value, or incoherence for experiments 1, 2, and 3, respectively). We performed GLMM significance testing using likelihood-ratio tests, and we corrected for multiple comparisons in post-hoc tests using Tukey's method. We used generalized additive mixed models (GAMM) to examine choice behavior as a function of trial within a block. We obtained smooth estimates of choice behavior using penalized regression splines, with penalization that allowed smooths to be reduced to zero effect(63).

We included separate smooths by block. We performed GAMM significance testing using approximate Wald-like tests(64).

## References

1. Leotti LA, Iyengar SS, Ochsner KN. Born to choose: The origins and value of the need for control. *Trends in Cognitive Sciences*. 2010.
2. Suzuki S. Effects of number of alternatives on choice in humans. *Behav Processes*. 1997;
3. Bown NJ, Read D, Summers B. The Lure of Choice. *J Behav Decis Mak*. 2003;
4. Leotti LA, Delgado MR. The inherent reward of choice. *Psychol Sci*. 2011;
5. Cockburn J, Collins AGE, Frank MJ. A Reinforcement Learning Mechanism Responsible for the Valuation of Free Choice. *Neuron*. 2014;
6. Bobadilla-Suarez S, Sunstein CR, Sharot T. The intrinsic value of choice: The propensity to under-delegate in the face of potential gains and losses. *J Risk Uncertain*. 2017;
7. Voss SC, Homzie MJ. Choice as a Value. *Psychol Rep*. 1970;
8. Catania AC. FREEDOM AND KNOWLEDGE: AN EXPERIMENTAL ANALYSIS OF PREFERENCE IN PIGEONS 1 . *J Exp Anal Behav*. 1975;
9. Suzuki S. Selection of forced- and free-choice by monkeys (*Macaca fascicularis*). *Percept Mot Skills*. 1999;
10. Perdue BM, Evans TA, Washburn DA, Rumbaugh DM, Beran MJ. Do monkeys choose to choose? *Learn Behav*. 2014;
11. Beattie J, Baron J, Hershey JC, Spranca MD. Psychological determinants of decision attitude. *J Behav Decis Mak* [Internet]. 1994 Jun 1 [cited 2022 Jun 13];7(2):129–44. Available from: <https://onlinelibrary.wiley.com/doi/full/10.1002/bdm.3960070206>
12. Ly V, Wang KS, Bhanji J, Delgado MR. A reward-based framework of perceived control.

- Front Neurosci. 2019;13(FEB):65.
13. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am Psychol*. 2000;55(1):68–78.
  14. Deci EL, Ryan RM. *Intrinsic Motivation and Self-Determination in Human Behavior*. Intrinsic Motivation and Self-Determination in Human Behavior. 1985.
  15. Bandura A, Freeman WH, Lightsey R. Self-Efficacy: The Exercise of Control. *J Cogn Psychother* [Internet]. 1999 Jan 1 [cited 2022 Jun 13];13(2):158–66. Available from: <https://connect.springerpub.com/content>
  16. Maier SF, Seligman MEP. Learned Helplessness: Theory and Evidence. *J ol Exp Psychol Gen*. 1976;105(1):3–46.
  17. deCharms R. *Personal causation: The internal affective determinants of behavior*. New York: Academic Press; 1968. 1–398 p.
  18. White RW. Motivation reconsidered: The concept of competence. *Psychol Rev* [Internet]. 1959 Sep [cited 2022 Jun 13];66(5):297–333. Available from: /record/1961-04411-001
  19. Ajzen I. Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior. *J Appl Soc Psychol*. 2002;
  20. Hull CL. *Principles of behavior*. New York: Appleton-Century-Crofts; 1943.
  21. Bromberg-Martin ES, Monosov IE. Neural circuitry of information seeking. *Curr Opin Behav Sci*. 2020 Oct 1;35:62–70.
  22. Kidd C, Hayden BY. The Psychology and Neuroscience of Curiosity. *Neuron*. 2015 Nov 4;88(3):449–60.
  23. Thrun SB. *Efficient Exploration In Reinforcement Learning*. Pittsburgh: Carnegie Mellon University; 1992.
  24. Fowler H. *Curiosity and Exploratory Behavior*. New York: Macmillan; 1965.
  25. Brehm JW. Postdecision changes in the desirability of alternatives. *J Abnorm Soc Psychol*. 1956 May;52(3):384–9.
  26. Festinger L. *A Theory of Cognitive Dissonance* [Internet]. Stanford: Stanford UP; 1957 [cited 2022 Jun 13]. Available from: [https://books.google.fr/books?hl=fr&lr=&id=voeQ-8CASacC&oi=fnd&pg=PA1&ots=9z87Msw9uB&sig=YErRLqdxMzgp8ZeMa0i55CPXm3w&redir\\_esc=y#v=onepage&q&f=false](https://books.google.fr/books?hl=fr&lr=&id=voeQ-8CASacC&oi=fnd&pg=PA1&ots=9z87Msw9uB&sig=YErRLqdxMzgp8ZeMa0i55CPXm3w&redir_esc=y#v=onepage&q&f=false)
  27. Sharot T, Velasquez CM, Dolan RJ. Do decisions shape preference? Evidence from blind

- choice. *Psychol Sci* [Internet]. 2010 Sep [cited 2022 Jun 13];21(9):1231. Available from: [/pmc/articles/PMC3196841/](#)
28. Izuma K, Matsumoto M, Murayama K, Samejima K, Sadato N, Matsumoto K. Neural correlates of cognitive dissonance and choice-induced preference change. *Proc Natl Acad Sci*. 2010;107:22014–9.
  29. Smith JE, Winkler RL. The Optimizer’s Curse: Skepticism and Postdecision Surprise in Decision Analysis. *Manage Sci* [Internet]. 2006 Mar 1 [cited 2022 Jun 13];52(3):311–22. Available from: <https://pubsonline.informs.org/doi/abs/10.1287/mnsc.1050.0451>
  30. Hasselt H. Double Q-learning. Vol. 23, *Advances in neural information processing systems*. New York: Macmillan; 2010.
  31. Zorowitz S, Momennejad I, Daw ND. Anxiety, Avoidance, and Sequential Evaluation. *Comput Psychiatry*. 2020;
  32. Gaskett C. Reinforcement learning under circumstances beyond its control. In: *Proceedings of the International Conference on Computational Intelligence for Modelling Control and Automation* [Internet]. 2003 [cited 2022 Jun 13]. Available from: <http://www.his.atr.co.jp/~cgaskett/>
  33. Ayal S, Zakay D. The perceived diversity heuristic: the case of pseudodiversity. *J Pers Soc Psychol* [Internet]. 2009 Mar [cited 2022 Jul 27];96(3):559–73. Available from: <https://pubmed.ncbi.nlm.nih.gov/19254103/>
  34. Schwartenbeck P, Fitzgerald THB, Mathys C, Dolan R, Kronbichler M, Friston K. Evidence for surprise minimization over value maximization in choice behavior. *Sci Rep* [Internet]. 2015 Nov 13 [cited 2022 Jul 27];5. Available from: <https://pubmed.ncbi.nlm.nih.gov/26564686/>
  35. Erev I, Barron G. On adaptation, maximization, and reinforcement learning among cognitive strategies. *Psychol Rev* [Internet]. 2005 Oct [cited 2022 Jul 27];112(4):912–31. Available from: <https://pubmed.ncbi.nlm.nih.gov/16262473/>
  36. Mistry P, Liljeholm M. Instrumental Divergence and the Value of Control. *Sci Reports* 2016 61 [Internet]. 2016 Nov 8 [cited 2022 Jul 26];6(1):1–10. Available from: <https://www.nature.com/articles/srep36295>
  37. Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ. Model-based influences on humans’ choices and striatal prediction errors. *Neuron* [Internet]. 2011/03/26. 2011;69(6):1204–15. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21435563>

38. Sharot T, Martino B De, Dolan RJ. How Choice Reveals and Shapes Expected Hedonic Outcome. *J Neurosci* [Internet]. 2009 Mar 25 [cited 2022 Jun 13];29(12):3760–5. Available from: <https://www.jneurosci.org/content/29/12/3760>
39. Wang KS, Kashyap M, Delgado MR. The Influence of Contextual Factors on the Subjective Value of Control. *Emotion* [Internet]. 2021 [cited 2022 Jul 4];21(4):881–91. Available from: <https://dx.doi.org/10.1037/emo0000760>
40. Chew SH, Ho JL. Hope: An empirical study of attitude toward the timing of uncertainty resolution. *J Risk Uncertain*. 1994;
41. Ahlbrecht M, Weber M. The Resolution of Uncertainty: An Experimental Study. *J Institutional Theor Econ JITE*. 1996;
42. Zheng Z, Oh J, Hessel M, Xu Z, Kroiss M, Van Hasselt H, et al. What can learned intrinsic rewards capture? In: 37th International Conference on Machine Learning, ICML 2020. 2020.
43. Singh S, Lewis RL, Barto AG, Sorg J. Intrinsically Motivated Reinforcement Learning: An Evolutionary Perspective. *IEEE Trans Auton Ment Dev*. 2010;
44. Botvinick M, Ritter S, Wang JX, Kurth-Nelson Z, Blundell C, Hassabis D. Reinforcement Learning, Fast and Slow. *Trends in Cognitive Sciences*. 2019.
45. Maier SF, Seligman MEP. Learned helplessness at fifty: Insights from neuroscience. *Psychol Rev*. 2016 Jul 1;123(4):349–67.
46. Haggard P, Chambon V. Sense of agency. *Curr Biol* [Internet]. 2012 May 22 [cited 2022 Jun 13];22(10). Available from: <https://pubmed.ncbi.nlm.nih.gov/22625851/>
47. Farkas BC, Chambon V, Jacquet PO. Do perceived control and time orientation mediate the effect of early life adversity on reproductive behaviour and health status? Insights from the European Value Study and the European Social Survey. *Humanit Soc Sci Commun* 2022 91 [Internet]. 2022 Feb 14 [cited 2022 Jun 13];9(1):1–14. Available from: <https://www.nature.com/articles/s41599-022-01066-y>
48. Kraus MW, Piff PK, Mendoza-Denton R, Rheinschmidt ML, Keltner D. Social class, solipsism, and contextualism: How the rich are different from the poor. *Psychol Rev*. 2012;119(3):546–72.
49. Liljeholm M. Instrumental Divergence and Goal-Directed Choice. In: *Goal-Directed Decision Making* [Internet]. Academic Press; 2018 [cited 2022 Jul 27]. p. 27–48. Available from: <https://doi.org/10.1016/B978-0-12-812098-9.00002-4>

50. Landau MJ, Kay AC, Whitson JA. Compensatory control and the appeal of a structured world. *Psychol Bull* [Internet]. 2015 May 1 [cited 2022 Jun 13];141(3):694–722. Available from: <https://pubmed.ncbi.nlm.nih.gov/25688696/>
51. Whitson JA, Galinsky AD. Lacking control increases illusory pattern perception. *Science* (80- ) [Internet]. 2008 Oct 3 [cited 2022 Jun 30];322(5898):115–7. Available from: <https://www.science.org/doi/10.1126/science.1159845>
52. Chambon V, Théro H, Vidal M, Vandendriessche H, Haggard P, Palminteri S. Information about action outcomes differentially affects learning from self-determined versus imposed choices. *Nat Hum Behav*. 2020;
53. Dorfman HM, Bhui R, Hughes BL, Gershman SJ. Causal Inference About Good and Bad Outcomes. *Psychol Sci* [Internet]. 2019 Apr 1 [cited 2022 Jun 13];30(4):516–25. Available from: <https://journals.sagepub.com/doi/full/10.1177/0956797619828724>
54. Hulaj R, Nyström MBT, Sörman DE, Backlund C, Röhlcke S, Jonsson B. A Motivational Model Explaining Performance in Video Games. *Front Psychol* [Internet]. 2020 Jul 14 [cited 2022 Jun 13];11. Available from: <https://pubmed.ncbi.nlm.nih.gov/32760321/>
55. Sidarus N, Palminteri S, Chambon V. Cost-benefit trade-offs in decision-making and learning. *PLoS Comput Biol* [Internet]. 2019 [cited 2022 Jun 13];15(9). Available from: <https://pubmed.ncbi.nlm.nih.gov/31490934/>
56. Rotter JB. Generalized expectancies for internal versus external control of reinforcement. *Psychol Monogr* [Internet]. 1966 [cited 2022 Jun 14];80(1):1–28. Available from: </record/2011-19211-001>
57. Abramson LY, Seligman ME, Teasdale JD. Learned helplessness in humans: Critique and reformulation. *J Abnorm Psychol* [Internet]. 1978 Feb [cited 2022 Jun 14];87(1):49–74. Available from: </record/1979-00305-001>
58. Palminteri S, Lefebvre G, Kilford EJ, Blakemore SJ. Confirmation bias in human reinforcement learning: Evidence from counterfactual feedback processing. *PLOS Comput Biol* [Internet]. 2017 Aug 1 [cited 2022 Jun 16];13(8):e1005684. Available from: <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005684>
59. Palminteri S, Khamassi M, Joffily M, Coricelli G. Contextual modulation of value signals in reward and punishment learning. *Nat Commun* [Internet]. 2015 Aug 25 [cited 2022 Jun 16];6. Available from: <https://pubmed.ncbi.nlm.nih.gov/26302782/>
60. Brainard DH. *The Psychophysics Toolbox*. *Spat Vis*. 1997;

61. Kleiner M, Brainard DH, Pelli DG, Broussard C, Wolf T, Niehorster D. What's new in Psychtoolbox-3? Perception. 2007;
62. Wagenmakers EJ, Farrell S. AIC model selection using Akaike weights. Psychon Bull Rev 2004 111 [Internet]. 2004 [cited 2022 Jun 14];11(1):192–6. Available from: <https://link.springer.com/article/10.3758/BF03206482>
63. Wood SN. Generalized Additive Models. An Introduction with R, Second Edition. Chapman and Hall; 2017. 496 p.
64. Wood SN. On p-values for smooth components of an extended generalized additive model. Biometrika [Internet]. 2013 Mar 1 [cited 2022 Jul 7];100(1):221–8. Available from: <https://academic.oup.com/biomet/article/100/1/221/192816>

**Acknowledgements:** J.M. was supported by the Agence Nationale de la Recherche (ANR) grant ANR-19-CE37-0014-01 (ANR PRC) and by the European Commission (H2020-MSCA-IF-2018-#845176). D.B. was supported by a FRM fellowship (FDM201906008526). V.C. was supported by the ANR grants ANR-17-EURE-0017 (Frontiers in Cognition), ANR-10-IDEX-0001-02 PSL (program 'Investissements d'Avenir'), ANR-16-CE37-0012-01 (ANR JCJ) and ANR-19-CE37-0014-01 (ANR PRC). B.L. was supported by the ANR grant ANR-19-CE37-0014-01. The authors of this article are grateful to Karim Ndiaye, operational manager of the PRISME platform at the ICM for his valuable help during participant testing.

**Author Contributions:** J.M., V.C. and B.L. designed the study; J.M., M.R.A., D.B. and A.K. performed the experiments and preliminary analyses V.C.; J.M., and B.L. designed and performed final analyses; J.M., V.C. and B.L. wrote the manuscript.

**Data availability statement:** All data and related metadata underlying the findings reported

will be deposited in Zenodo (DOI: 10.5281/zenodo.7057043) at the time of publication.

**Code reporting:** Code written in support of this publication will be made publicly available in Zenodo (DOI: 10.5281/zenodo.7057080) at the time of publication.

## Supporting Information

### Supporting Information Text

In order to verify that we could discriminate different TDRL models and accurately estimate parameters, we performed model and parameter recovery analyses on simulated datasets. We considered six key models, the three different value estimates (SARSA, Q-learning,  $\beta$ -pessimistic) with and without a free choice bonus ( $\rho$ ). For each of these models, we simulated the performance of 5000 virtual subjects using the same set of conditions that real subjects experienced in experiment 1. We explored a wide parameter space, drawing generative parameters from the following distributions:

$$\alpha \sim \text{beta}(\text{shape1}=1.75, \text{shape2}=2.75)$$

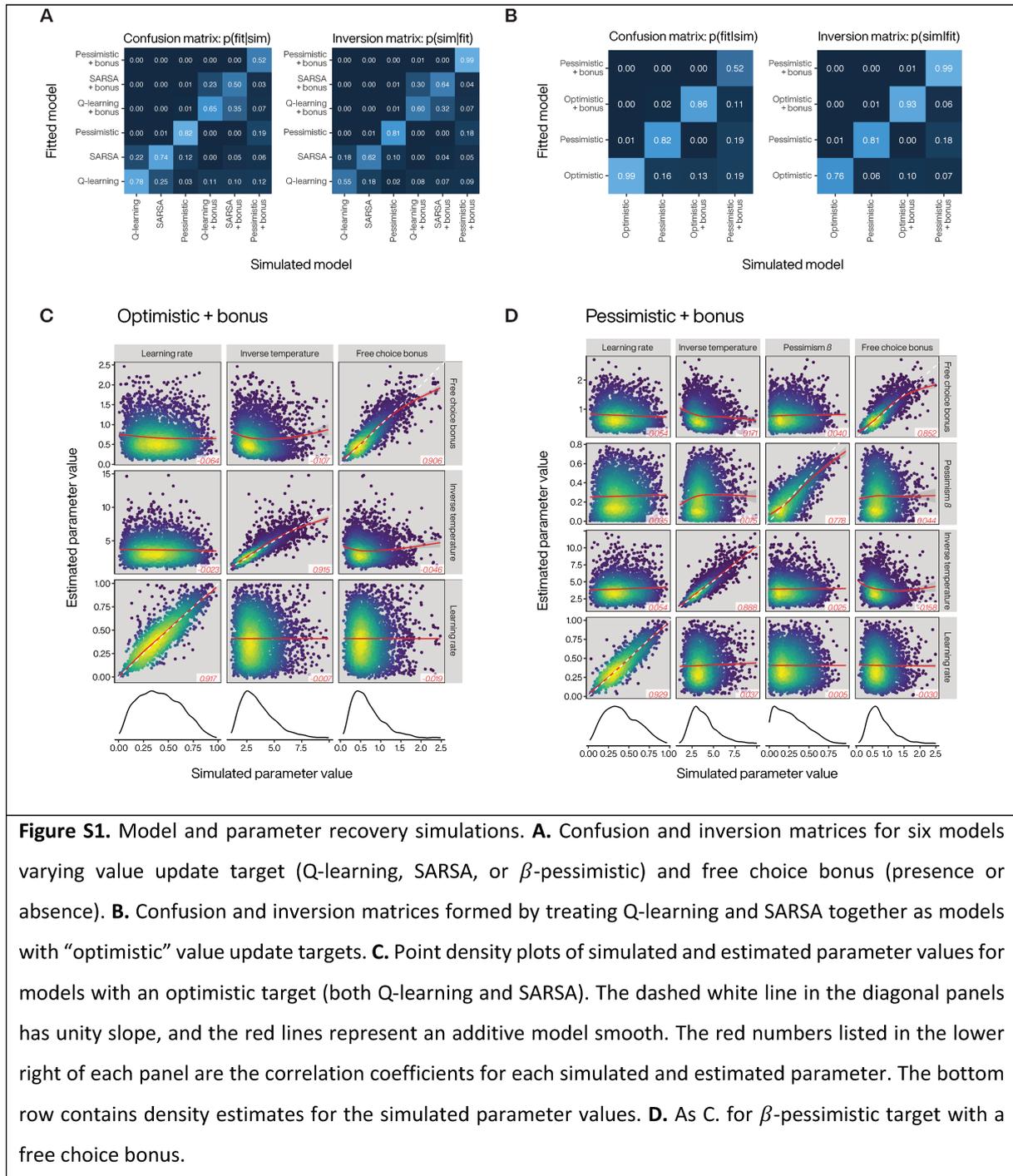
$$1/\tau \sim \text{gamma}(\text{shape}=2.5, \text{scale}=1)$$

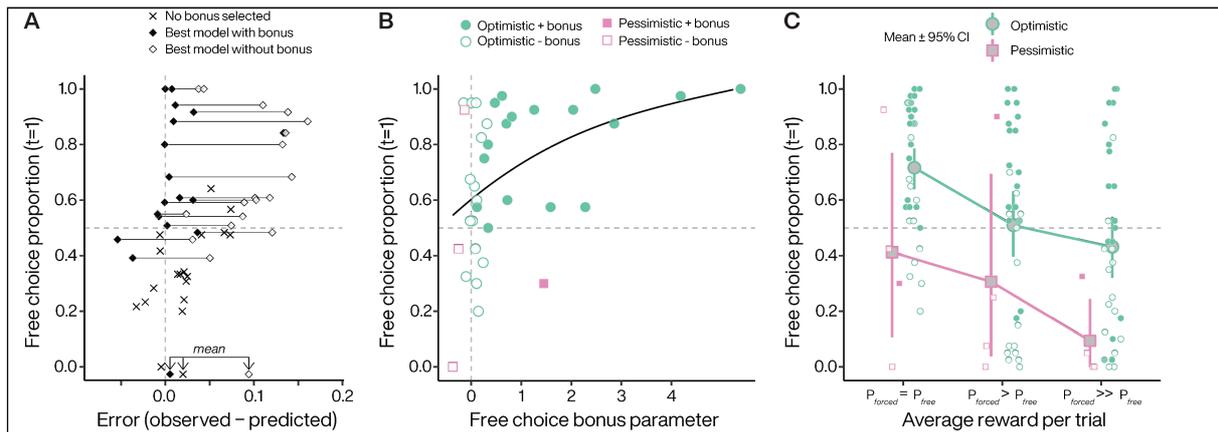
$$\beta \sim \text{beta}(\text{shape1}=1, \text{shape2}=2.25)$$

$$\rho \sim \text{gamma}(\text{shape}=2.5, \text{scale}=0.25)$$

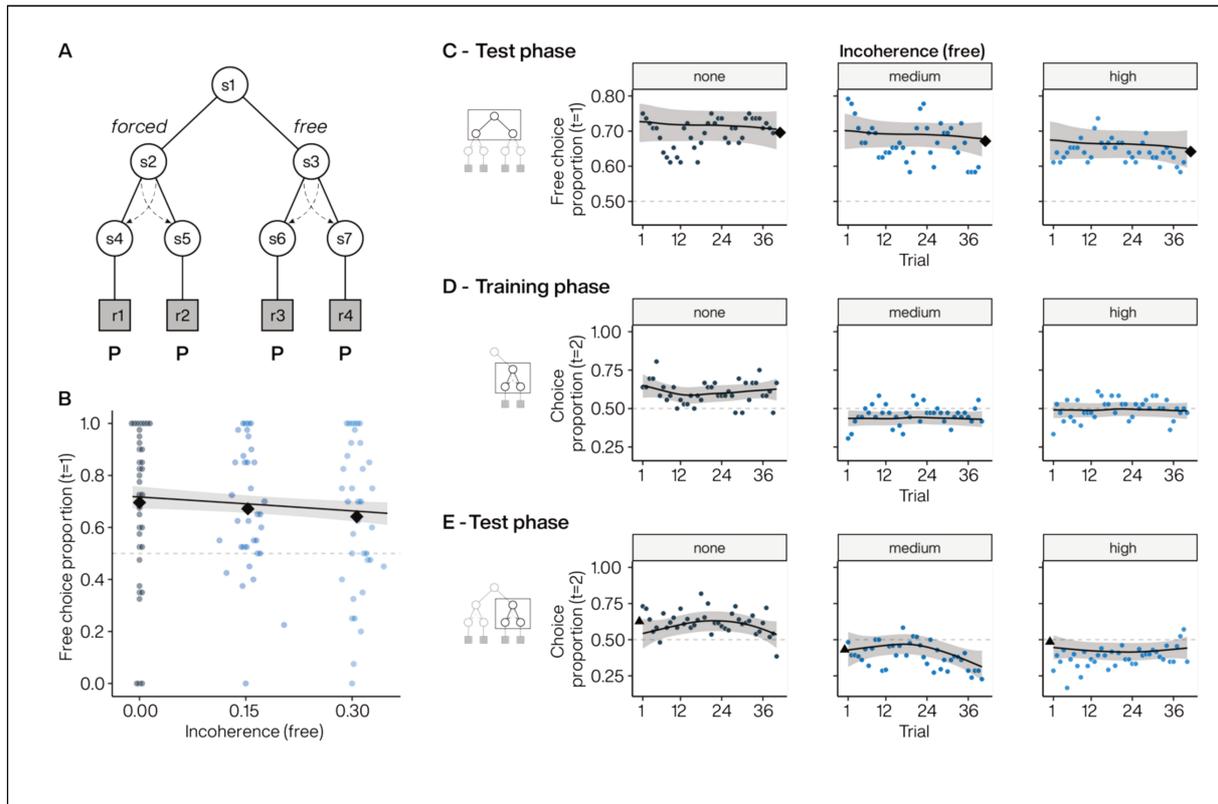
We fitted simulated data and performed model selection using the same procedures as for real data. From this data, we calculated confusion and inversion matrices(1) (Fig. S1A), which give the probability that the data simulated by one model is best fit by another and the probability that one model is more likely generated by another, respectively. We found that model recovery was good, although Q-learning and SARSA algorithms were confused in the parameter ranges used. This is due to the fact that Q-learning is strictly optimistic by considering only the best future state-action value, whereas SARSA can be more or less optimistic depending on the sensitivity of the mapping from state-action value differences to behavioral policy. Thus, our use of a softmax policy means that SARSA approaches Q-learning as  $1/\tau$  increases. Therefore, we treated Q-learning and SARSA algorithms together as “optimistic” in the main text. Pooling these two models together indicates that model recovery is good (Fig. S1B).

We also examined parameter recovery using correlations between estimated and simulated parameters, both with the same parameter as well as across parameters (Fig. S1C-D). We found that parameter recovery was good (diagonal correlations), and parameter confounding was quite weak (off-diagonal correlations).





**Figure S2.** Reinforcement learning models capture individual choice behavior in experiment 2. **A.** Obtained free choice proportion as a function of model error, averaged over all conditions. For subjects where the selected model did not include a free choice bonus, only one symbol (X) is plotted. For subjects where the selected model included a free choice bonus, two symbols are plotted. Filled symbol represents the fit error with the selected model, and the open symbol represents the next best model that did not include a free choice bonus. Lines connect individual subjects. **B.** Bonus coefficients increase as a function of subjects' preference for *free* options irrespectively of the target policy they used when performing the task. Choice preference from equal probability blocks ( $P=0.75$ ). Filled circles indicate that the best model included a free choice bonus parameter. Line illustrates a generalized additive model smooth. **C.** Both optimistic and pessimistic subjects decrease their *free* option preference as the *forced* option value increases. Symbol legend from B applies to the small points representing individual means in C. Error bars for 95% CI.



**Figure S3.** Choice proportion across different action-outcome incoherence level. A. Experiment 3 task design using seven states task design where we manipulated the incoherence in both free and forced options. B. Subject preference for free options during 1<sup>st</sup>-stage. Colored points indicate individual subject mean choice preference per block, plotted against the incoherence level. Black diamonds indicate the average of subject means per block. Line indicates the estimated choice preference from a GLMM, with 95% CI. C. Dynamics of free option preferences across test phase blocks for incoherence set at 0 (i.e., none, left), 0.15 (i.e. medium, middle) and 0.30 (i.e. high, right). Each point represents the average free option preference as a function of trial within a block. Diamonds: as in B. Lines indicate the estimated choice preference from a GAMM, with 95% CI. D to E. Dynamics of the selection of the two 2<sup>nd</sup>-stage targets (equally rewarded) in free options across the blocks for incoherence levels set at 0 (left), 0.15 (middle) and 0.30 (right) during the training (D) and test (E) phases. Triangles represents the final average selection at the end of the training phases. Lines: as in C.

**SI References**

1. Wilson RC, Collins AGE. Ten simple rules for the computational modeling of behavioral data. *Elife*. 2019 Nov 1;8.

## ETUDE 2

**Arguments anatomiques pour une valeur intrinsèque du  
choix chez des patients atteints de la maladie de Parkinson**



## Etude 2 : introduction

Nous avons vu dans la première étude que le choix présentait donc les caractéristiques d'une récompense intrinsèque, quantifiable et permettant aux individus de rétablir un sentiment de contrôle sur leur environnement. Par le biais de modélisations computationnelles, ce comportement pouvait être expliqué par des algorithmes d'apprentissage par renforcement. Plusieurs études ont démontré que les neurones dopaminergiques sous-tendent cet apprentissage en permettant d'encoder une erreur de prédiction chez le singe (Schultz et al., 1997). Ce résultat a été confirmé chez l'humain, aussi bien chez le sujet sain recevant de la dopamine exogène, que chez les patients parkinsoniens dont les capacités d'apprentissage sont altérées sans traitement dopaminergique et rétablies lors de la prise de ce dernier (Palminteri et al., 2009; Pessiglione et al., 2006). L'activité des neurones dopaminergiques est également modulée par le traitement d'une récompense intrinsèque comme l'obtention d'information (Bromberg-Martin & Hikosaka, 2009). Concernant la préférence pour le choix, plusieurs études en IRM fonctionnelle ont relié l'augmentation de débit du striatum à des situations où il était possible de réaliser plusieurs choix librement (Leotti & Delgado, 2014; Sharot et al., 2010; K. S. Wang & Delgado, 2019). Enfin, le NST qui est une structure centrale dans le fonctionnement des ganglions de la base, est impliquée dans la prise de décision. Son activité semble être également modulée par la composante motivationnelle de l'action (Cavanagh et al., 2011; Zénon et al., 2016). Chez les patients parkinsoniens, la stimulation cérébrale profonde du NST permet de diminuer les symptômes moteurs mais fait également émerger des réponses impulsives (Frank et al., 2007).

En prenant comme modèle la maladie de Parkinson, nous avons inclus deux populations différentes. L'une composée des patients prenant un traitement oral dopaminergique, l'autre composée de patients parkinsoniens traités neurochirurgicalement par stimulation du NST. Nous attendons que dans ces deux groupes, un effet aigu des traitements oraux ou chirurgicaux influe la préférence pour le choix. En particulier, que celle-ci soit diminuée lorsque le patient est sans traitement, et qu'elle soit rétablie lorsque le traitement est actif.

La tâche utilisée dans cette étude est identique à celle de l'expérience 1 chez les sujets sains. Les sections suivantes décrivent la méthodologie détaillée ainsi que les principaux résultats provenant du manuscrit de l'article situé à la fin de ce chapitre.

## Etude 2 : méthodologie

### Participants

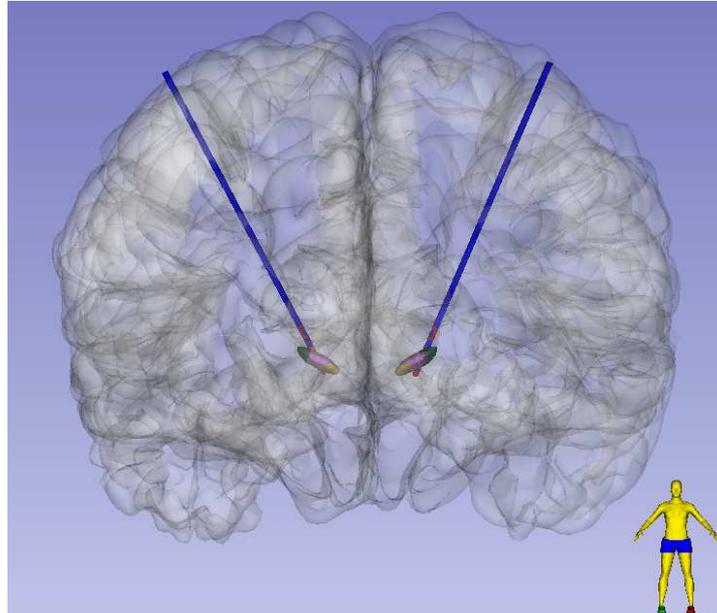
Quarante-cinq patients ont été recrutés dans les services de neurologie et de neurochirurgie de l'hôpital de la Pitié-Salpêtrière (Paris, France). Tous les patients remplissaient les critères de la maladie de Parkinson de la *United Kingdom Parkinson's Disease Society Brain Bank* (Hughes et al., 1992). Les critères d'exclusion étaient la présence d'une déficience cognitive significative (MoCA < ou = 25/30) ou d'une maladie neurologique ou psychiatrique active concomitante.

Pour le premier groupe de patients (groupe DOPA, n=25, 7 femmes), les inclusions ont été effectuées lors d'une hospitalisation pour évaluer l'éligibilité à la chirurgie. Le second groupe est constitué de patients déjà opérés (groupe NST, n=20, 9 femmes), recrutés pour une évaluation postopératoire environ un an après la chirurgie (moyenne : 15 mois). Huit d'entre eux étaient implantés avec les électrodes de SCP Medtronic 3389 (Medtronic Inc, Minneapolis, MN, USA) et les douze autres étaient implantés avec les électrodes directionnelles Boston Vercise Cartesia (Boston Scientific, Marlborough, MA). Parmi ces derniers, cinq avaient un courant directionnel sur au moins une des deux électrodes (une électrode par hémisphère pour tous les patients).

#### Implantation des électrodes

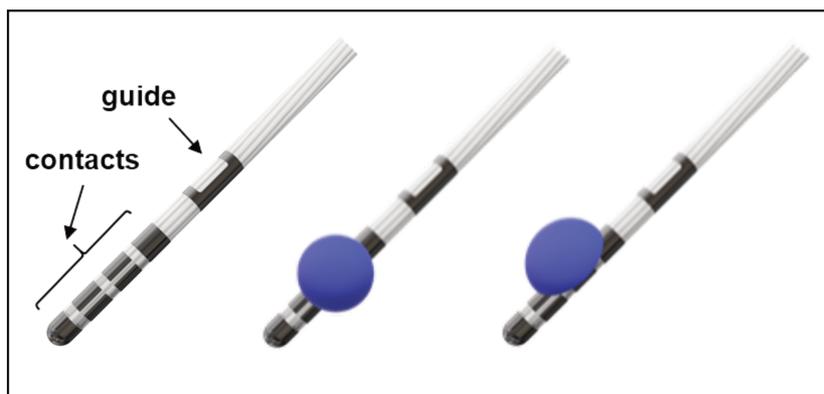
Quelques jours avant la chirurgie, les patients ont bénéficié d'une IRM 3 Tesla avec des séquences FLAIR 3D isotropes (*Fluid-attenuated inversion recovery*), QSM (*Quantitative Susceptibility Mapping*), MP2RAGE (*standard Magnetization Prepared Two Rapid Gradient Echo*) (0,9 mm iso-voxel. Les images FLAIR ont été utilisées pour le ciblage direct du NST qui apparaît en hyposignal sur cette séquence. Le jour de l'intervention une IRM 1,5 Tesla stéréotaxique T1 anatomique avec injection de gadolinium a été acquise pour le ciblage indirect avec l'atlas *Yeb* (Bardinet et al., 2009; Yelnik et al., 2007) et le calcul des trajectoires. Au cours de l'intervention, un enregistrement électrophysiologique peropératoire avec trois microélectrodes permet de confirmer la position du NST de chaque

côté puis une stimulation pendant la phase d'éveil est effectuée afin de choisir la meilleure position au sein du NST. Finalement, deux électrodes ont été implantées et connectées au neurostimulateur. La position des électrodes a été vérifiée à l'aide d'une tomodensitométrie cérébrale (résolution de  $0,392578 \times 0,392578 \times 0,625 \text{ mm}$ ).



**Figure 28.** Exemple de reconstruction des trajectoires des électrodes au sein du NST, avec son territoire limbique (jaune), associatif (rose), et sensorimoteur (vert) défini avec l'atlas YeB (Bardinet et al., 2009).

Les électrodes de stimulation classiques permettent de délivrer le courant électrique de manière sphérique. L'apparition récente de nouvelles électrodes directionnelles permet d'administrer le courant électrique dans une direction donnée afin d'éviter les structures ou les fibres à l'origine des effets indésirables potentiels (figure 2). Ainsi, le champ de stimulation n'est plus distribué de manière homogène autour du contact mais peut être déformé dans le plan désiré (figure 1). Ces électrodes présentent huit contacts répartis sur quatre niveaux, les deux niveaux du milieu étant divisés en trois contacts autour de l'électrode. Un guide radio opaque est également présent au-dessus du premier contact.



**Figure 29.** Représentation schématique d'une électrode directionnelle (gauche), délivrant un courant sphérique (milieu), ou de manière directionnelle (droite ; d'après Boston Scientific, Marlborough, MA).

Lors de l'évaluation, une échelle UPDRS de la sévérité des symptômes de la maladie a été réalisée pour chaque patient en condition de traitement ON et OFF. Le calcul de l'équivalence en L-DOPA des patients a été réalisée au moyen des tables habituelles de conversion des différents traitement antiparkinsoniens (Tomlinson et al., 2010). Les paramètres de SCP ont été colligés individuellement afin d'être modélisé ultérieurement.

## Tâche expérimentale

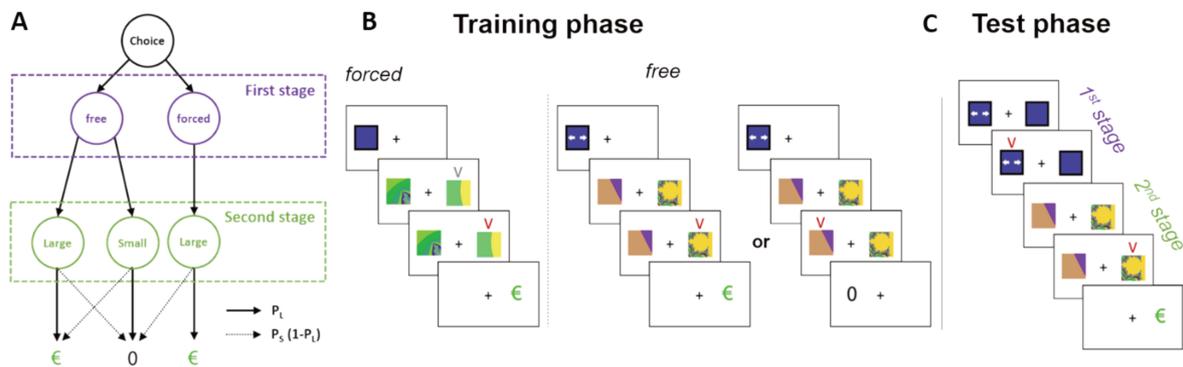
Il s'agit de la même tâche qu'effectuée par les sujets sains dans l'expérience 1.

Les patients étaient assis devant un écran et il leur a été dit que différentes paires d'images leur seraient présentées avec différentes probabilités d'être récompensées. Le but de la tâche était d'accumuler le plus possible d'argent symbolisé par un euro vert (€) sur l'écran. Les participants pouvaient sélectionner ces images de deux manières différentes : soit en choisissant eux-mêmes ("choix libre"), soit en étant guidés par un ordinateur, lorsqu'une seule des deux images devait être sélectionnée ("choix forcé").

La tâche leur était présentée sur support visuel (diaporama) et détaillée par oral. Durant l'explication, le patient a la possibilité de se familiariser avec la structure de la tâche en réalisant un essai simulé. Cet essai était également accompagné d'explications pour s'assurer d'une compréhension optimale. Les patients sélectionnent les images en appuyant sur les touches gauche et droite du clavier de l'ordinateur.

La tâche présente une structure en deux étapes. Lors de la première étape, deux images étaient affichées à l'écran, représentant des options "libres" ou "forcées", et les patients devaient faire leur choix. Lors de la deuxième étape, deux images différentes étaient affichées. Dans l'option "libre", les patients pouvaient choisir n'importe laquelle des deux images, tandis que dans l'option "forcée", ils devaient choisir celle qui avait été présélectionnée par l'ordinateur. Cette image obligatoire était indiquée par un V gris au-dessus de l'image. Notons que l'ordinateur sélectionnait toujours la même image et que la sélection de l'image opposée était désactivée pendant la condition "forcée".

Au cours des deux étapes, après avoir appuyé sur l'une des deux touches, les patients recevaient un retour visuel affiché au-dessus de l'image sélectionnée (V rouge). Après avoir sélectionné l'image de la deuxième étape, le résultat de la récompense était affiché et pouvait être soit le symbole vert euro lorsque l'essai était récompensé, soit un zéro noir lorsque la récompense était absente. (Figure 1).



**Figure 30.** Structure de la tâche en deux étapes. A : diagramme illustrant les différentes voies que les sujets empruntent au cours de la tâche. B : illustration schématique d'un essai pendant l'entraînement. Pour chaque block de probabilité, les sujets s'entraînaient sur deux blocs séparés : un avec choix libre, un avec choix forcé. C : illustration schématique d'un essai pendant le test.

Les contingences de récompense variaient entre trois blocs :

- Dans un bloc, les récompenses étaient stochastiques puisque toutes les images de la deuxième étape étaient associées à une probabilité de récompense de  $P = 0,5$ .
- Dans un autre bloc, les probabilités de récompense pour les images forcées et la cible la plus récompensée en libre étaient fixées à  $P = 0,75$  et à  $1-P$  pour l'image la moins récompensée en libre.

- Dans le dernier bloc, la délivrance des récompenses était déterministe puisque le choix des images en forcé ou de la plus récompensée en libre conduisait toujours à une récompense ( $P=1$ ) alors que le choix de l'autre image en libre conduisait systématiquement à son absence ( $P=0$ ).

Les contingences de récompense étaient donc identiques entre les options libre et forcée, mais les images étaient différentes entre les options ainsi que dans tous les différents blocs. Il est important de noter qu'en option forcée, l'ordinateur a toujours sélectionné la même image pour les trois blocs (qui était également la plus récompensée lorsque  $P > 0,5$ ). Ce faisant, le choix de l'image en "forcé" ou le choix de l'image la plus récompensée en "libre" sont équivalents en termes d'accumulation de récompenses.

Au début de chaque bloc, c'est-à-dire avant la phase de test, les patients ont effectué une phase d'entraînement afin d'apprendre les contingences entre les différentes images et leurs probabilités de récompense associées. Au cours de cette phase, il n'est pas proposé aux patients de choisir entre des options libres ou forcées et ils ont donc effectué successivement des essais libres et forcés obligatoires (par exemple, 48 essais forcés puis 48 essais libres). A noter que la durée de la phase d'entraînement pour  $P=1$  a été divisée par deux en raison de la nature déterministe, et donc facile, de la remise de la récompense. Les patients étaient explicitement informés que les stimuli restaient identiques entre les phases d'entraînement et de test.

La tâche a été écrite en utilisant la Psychophysics Toolbox (<http://psychtoolbox.org/>, Brainard D.H., 1997) dans MATLAB (Natick, Massachusetts : The MathWorks Inc).

## **Déroulement de la passation**

Les patients ont effectué la tâche et tous les blocs deux fois : avec traitement (ON) et sans traitement (OFF) pour les groupes DOPA et NST. Les patients du groupe NST n'étaient pas sous traitement médicamenteux pendant les deux sessions pour tenir compte uniquement

de l'effet de la stimulation. Toutes les images étaient également différentes entre les sessions OFF et ON.

Pour le groupe DOPA, la session OFF a eu lieu selon la procédure locale, c'est-à-dire 12h après l'arrêt du traitement pour la lévodopa et 72h pour les agonistes dopaminergiques. Cette procédure a été mise en place pour permettre une élimination efficace de tous les traitements pharmacologiques antiparkinsoniens. La session ON a eu lieu le même jour, 45 minutes après la prise de la dose équivalente de leur traitement habituel du matin plus 50 mg de lévodopa.

Pour le groupe NST, l'élimination des traitements pharmaceutiques pour les sessions ON et OFF s'est faite de la même manière que pour le groupe DOPA. La session ON (stimulation ON) a eu lieu en premier et la session OFF a commencé au moins une heure après l'arrêt de la stimulation.

## **Analyse statistique**

Les analyses statistiques des données démographiques ont été réalisées au moyen d'une test de Wilcoxon.

En raison de la structure hiérarchique de la tâche, des modèles mixtes ont été réalisés à l'aide du paquet lme4 sur R (Bates et al., 2015, R Core Team, <https://www.R-project.org>). Des tests de Wilcoxon à deux échantillons ont été utilisés pour les statistiques démographiques.

Des modèles linéaires généralisés mixtes (GLMM) ont été utilisés pour les variables binomiales ("choix libre"/"choix forcé" et images de seconde étape). Les effets principaux et les interactions de tous les modèles ont été rapportés en utilisant les statistiques z de Wald et les valeurs p correspondantes. Pour les variables continues (temps de réaction), nous avons réalisé des modèles mixtes linéaires et rapporté les statistiques t et les valeurs p correspondantes pour les effets principaux et les interactions des modèles.

Les patients ont été considérés comme un effet aléatoire pour tous les modèles. Une pente pour chaque effet fixe a été ajoutée à la partie aléatoire du modèle lorsque cela était possible

; les pentes ont été ajoutées à la composante aléatoire de manière hiérarchique jusqu'à la singularité du modèle (Bolker et al., 2009; Harrison et al., 2018).

Avant l'analyse, les effets fixes continus ont été moyennés et centrés sur l'ensemble des données et les temps de réaction ont été filtrés au-dessus d'un écart absolu médian de 2 fois l'ensemble des données. (Lachaud et Renaud, 2011).

Les statistiques post-hoc ont été calculées à l'aide des paquets emmeans sur R, ajustées pour les comparaisons multiples à l'aide de la méthode de Tukey, les valeurs p ont été calculées à l'aide d'un test z tel que fourni par défaut dans le paquet emmeans sur R (Russell, 2021, <https://CRAN.R-project.org/package=emmeans>). Pour les GLMM, les odd-ratios (OR) sont fournis à la place du test z.

## **Analyse des données d'imagerie**

Les images de tous les patients du groupe NST ont été traitées (n=20). Pour chaque patient était présent un volume 3D T1 pré-opératoire et un volume TDM post-opératoire. Pour 18 patients, un volume 3D FLAIR était également disponible. Pour les deux restants, un volume coronal T2 centré sur les NST étaient disponibles. L'utilisation d'images pondérées T2 permet la visualisation du NST.

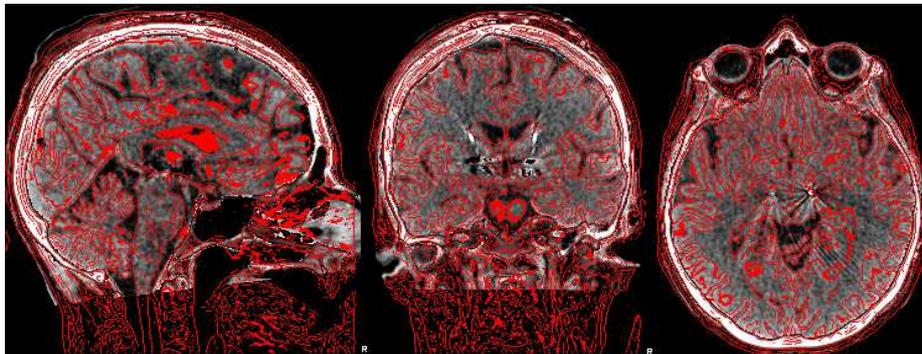
Un pré-traitement a ensuite été réalisé à l'aide du programme Lead-DBS (Horn et al., 2019). Lead-DBS est un outil de traitement d'imagerie de stimulation cérébrale profonde dont l'utilisation croissante est principalement justifiée par la fluidité de son utilisation. Cet outil permet de facilement regrouper les différentes étapes de prétraitement et d'analyse des images de patients en y incluant les programmes de traitement d'image les plus utilisés (comme SPM, ANTs, ou FSL). De plus, son utilisation spécifique dans le champ de la stimulation cérébrale profonde a permis aux programmeurs d'en optimiser les différents paramètres.

## Recalage des images

Pour le recalage intra-sujet, les images TDM ont été co-registrées avec les séquences IRM pondérées en T1 en utilisant une transformation linéaire en deux étapes implémentée dans Advanced Normalization Tools (Avants et al., 2008). La séquence IRM FLAIR a été co-registrée linéairement avec la séquence T1 en utilisant SPM12 (Friston, 2007).

Ensuite, les acquisitions pré et postopératoires ont été recallées spatialement dans l'espace MNI\_ICBM\_2009b NLIN ASYM (Fonov et al., 2011) en utilisant l'approche *SyN* implémentée dans Advanced Normalization Tools (Avants et al., 2008). Une transformation non linéaire a été ajoutée en utilisant *SyN* et se sont concentrés consécutivement sur les ganglions de la base tels que définis par le masque sous-cortical dans Schönecker et al. (2009).

Les résultats du recalage étaient ensuite tous vérifiés visuellement pour s'assurer d'une superposition correcte des structures entre les différentes acquisitions notamment les contours du troisième ventricule et s'assurer de l'absence d'artefacts de micro-mouvements, source d'erreur de recalage.



**Figure 31.** Exemple de vérification visuelle de recalage entre l'IRM (file de fer rouge) et la TDM d'un même sujet dans Lead-DBS. La juste superposition des deux images au niveau des ventricules et du contour du mésencéphale servait particulièrement de gage de réussite du recalage.

### Explication détaillée :

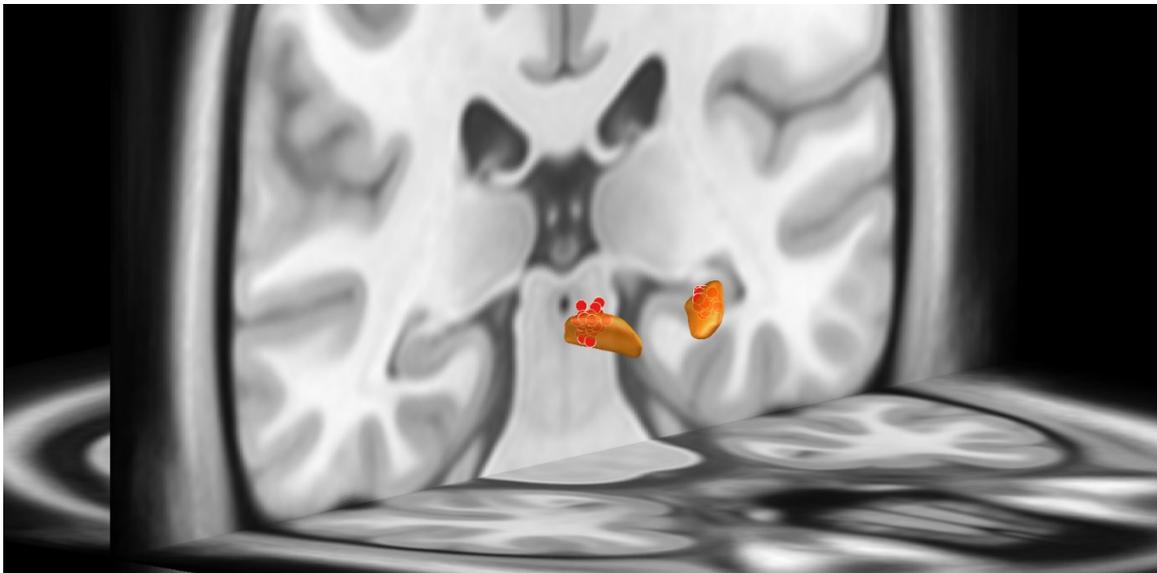
Dans le cas de notre population et de la structure d'intérêt qui est ici le NST, le recalage présente un enjeu de précision accru. Du fait de la dimension du NST à l'échelle du millimètre, toute déformation inadéquate de l'image native peut entraîner des résultats

erronés. Ainsi, la technique suscitée a été validée en comparant plusieurs techniques, dont *Syn*, sur 103 IRM. Un atlas des ganglions de la base était ensuite superposé aux images, dont le résultat a été comparé au traçage manuel du NST réalisé par des experts.. La technique de recalage *Syn* produisait le meilleur rapport de vraisemblance entre les résultats de l'atlas et celui des experts.

### Reconstruction de la position des électrodes de stimulation

Les électrodes ont été modélisées en utilisant la méthode PaCER (Husch et al., 2018). Cette méthode a toutefois échoué pour un patient et a nécessité l'utilisation d'un algorithme alternatif (TRAC/CORE). Après analyse visuelle de l'ensemble des reconstructions sur les deux méthodes, il n'a pas été nécessaire de réaliser d'ajustements manuels

Brièvement, le modèle se sert de l'intensité des voxels contenus dans la TDM postopératoire pour retracer ensuite l'électrode et localiser l'ensemble des contacts.



**Figure 5.** Visualisation de chaque contact activé pour l'ensemble des patients (point rouge). LE NST est ici en orange.

Les coordonnées des contacts actifs pour chaque patient sont reportées dans le tableau 1.

patient	Gauche			Droite		
	x	y	z	x	y	z
1	12,0	9,5	1,8	-12,1	9,8	2,7
2	10,9	10,9	3,9	-11,5	11,0	3,3
3	11,2	10,6	2,0	-10,6	10,5	2,6
4	11,9	9,6	4,8	-11,0	10,4	2,9
5	11,2	9,3	3,2	-11,6	11,1	3,1
6	12,5	11,5	3,4	-11,0	11,9	2,3
7	11,7	9,5	3,2	-10,7	10,1	4,1
8	11,2	9,5	3,0	-12,6	10,8	2,3
9	12,6	9,3	2,0	-14,0	10,5	0,9
10	12,9	10,6	3,3	-11,2	11,6	1,4
11	12,3	11,4	2,7	-12,0	11,4	2,5
12	11,3	12,0	3,0	-12,9	11,4	1,9
13	12,0	10,0	2,8	-11,2	10,5	3,9
14	12,7	10,4	2,5	-12,4	11,0	2,2
15	10,7	9,6	2,3	-12,2	14,0	3,7
16	10,8	9,7	2,7	-11,1	10,0	4,2
17	11,9	9,5	2,8	-10,3	10,3	2,7
18	12,6	10,9	1,6	-10,9	9,6	1,4
19	11,5	8,9	2,0	-12,1	10,3	2,5
20	13,5	11,1	3,7	-12,7	12,1	4,5

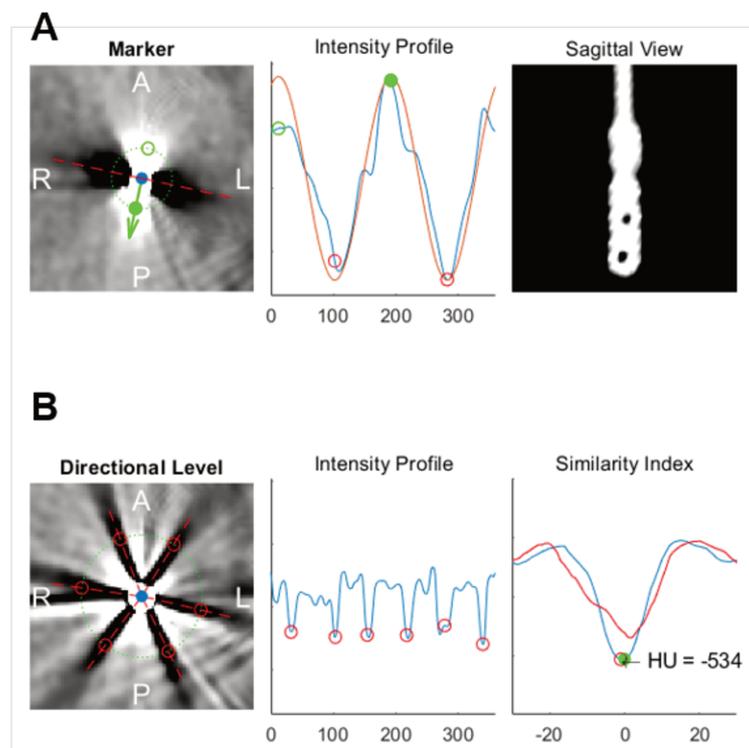
**Tableau 1.** Coordonnées (mm) des contacts dans le référentiel AC-PC (CP l'origine).

### Détection de l'orientation de l'électrode

Pour les patients équipés d'une électrode directionnelle, l'orientation de l'électrode a été effectuée avec le programme DiODE (Dembek et al., 2021).

### Explication détaillée :

Ce programme utilise les artefacts produits par l'électrode sur la TDM. En effet, la TDM étant une technique d'imagerie émettrice de rayons X, l'électrode métallique crée des anomalies sur l'image qu'il est possible d'analyser. En alignant un plan orthogonal à l'axe de l'électrode, ce programme détecte la direction d'intensité maximale au niveau du guide de l'électrode, ce dernier étant asymétrique. Il analyse également l'orientation des artefacts des contacts directionnels (les deux niveaux du milieu) pour compléter la détection de l'orientation.



**Figure 6.** Exemple d'un patient présentant une électrode directionnelle avec illustration des différents artefacts. **A.** le sens de l'artefact est détecté par le programme qui en détermine une première direction. La détection du maximal d'amplitude de l'artefact dans la vue sagittale permet à son tour de détecter sa polarité (vers l'avant ou vers l'arrière sur l'axe). **B.** Enfin, le pattern en étoile réalisé par les contacts et l'espacement entre chacune des "branches" permet de raffiner l'orientation de l'électrode.

## Calcul des volumes de tissu activé (VTA)

L'estimation du volume de tissu activé (VTA) a été réalisée en utilisant la méthode quadri-compartiments implémentée dans LEAD. Pour construire un modèle conducteur volumique de l'électrode DBS et du tissu environnant, un maillage volumique tétraédrique a été généré sur la base des maillages de surface des électrodes DBS et des noyaux sous-corticaux à l'aide de la boîte à outils Iso2Mesh ([http:// iso2mesh.sourceforge.net/](http://iso2mesh.sourceforge.net/)) telle qu'elle est incluse dans LEAD. La distribution du champ électromagnétique de la stimulation a été simulé en utilisant FieldTrip-SimBio dans LEAD (<https://www.mrt.uni-jena.de/simbio/index.php/> ; <http://fieldtriptoolbox.org>). Le gradient de la distribution de potentiel a été calculé par dérivation de la solution de la méthode des éléments finis (FEM). Le gradient a été seuillé pour les magnitudes supérieures à une valeur couramment utilisée de 0,2 V/mm pour définir l'étendue et la forme du volume de tissu activé.

### **Explication détaillée :**

Son principe repose sur des notions de biophysique et bio-ingénierie de l'étude des champs électromagnétiques et leur conductivité dans les milieux. Le courant traversant l'électrode émet un rayonnement à l'extrémité du contact activé, perdant progressivement de son intensité tout en se propageant le long des tissus nerveux. Il est donc possible de modéliser un volume théorique d'action de la stimulation. Plusieurs modèles de modélisation de ce volume sont inclus dans LEAD-DBS; la plupart se basent sur le modèle de McIntyre et al. (2004) et permettent une acquisition rapide. Horn et al. (2017) ont introduit un nouveau modèle plus précis prenant en compte quatre compartiments de conductivité différente que sont la substance grise, la substance blanche, les contacts de l'électrode (en Platine/Iridium), et le matériau isolant l'électrode.

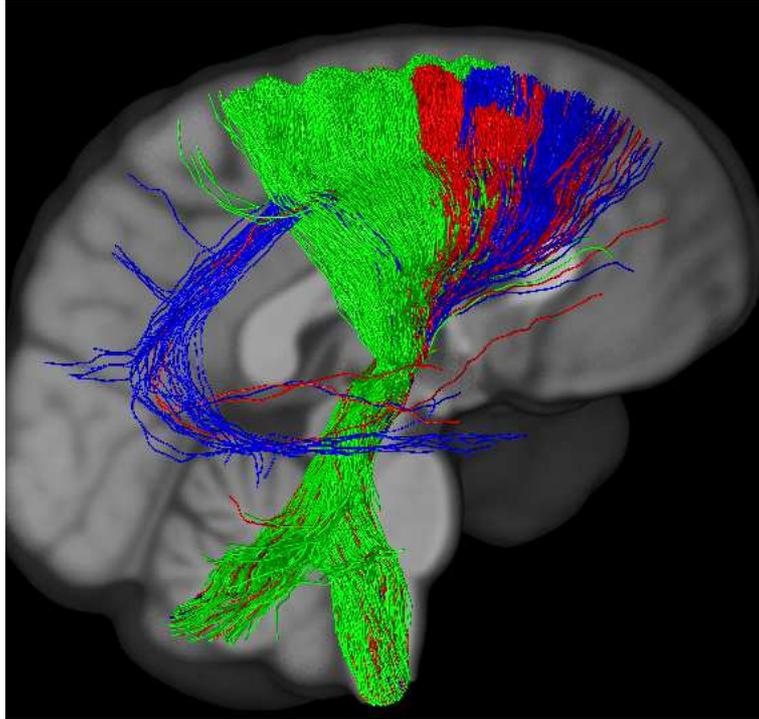
## Connectivité corticale

### Principe général de l'imagerie de diffusion

L'imagerie de diffusion permet d'obtenir une approximation de la connectivité des faisceaux de substance blanche. Il se base sur l'hétérogénéité de diffusion des molécules d'eau en fonction des milieux qu'elle traverse. Par exemple, l'eau pure présente une diffusion stochastique, sans direction préférentielle, elle présente donc une isotropie. A l'inverse, les molécules contenues dans la substance blanche auront tendance à suivre une direction parallèle aux fibres axonales, elle présente donc une anisotropie.

### Utilisation d'un connectome normatif

En l'absence de données individuelles de diffusion, nous avons utilisé un connectome normatif issu des données de 33 patients et réalisé par un membre de l'équipe (Temiz et al., en révision favorable, 2022). Le but de ce tractogramme normatif est de pouvoir analyser la connectivité cortico-subthalamique chez le patient parkinsonien. Bien que les images des patients soient moins résolues car les séquences sont plus rapides, nous avons pu valider ce tractogramme, en particulier la connectivité de la voie cortico-subthalamique, en le comparant aux résultats obtenus par une étude précédente utilisant les données du Human Connectome Project (HCP ; Temiz et al., 2020).

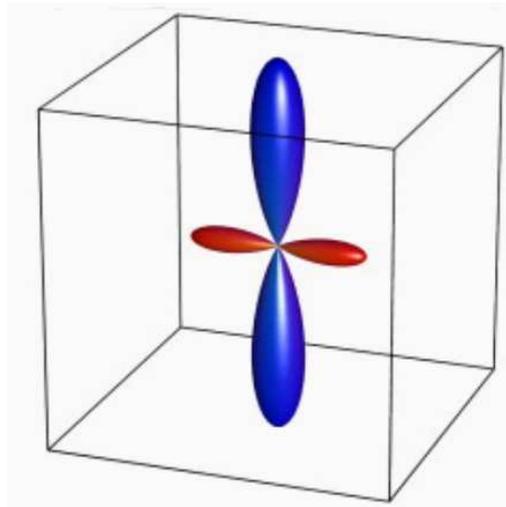


**Figure 7.** Représentation de la connectivité d'un patient parkinsonien à l'aide du connectome normatif en fonction du territoire anatomo-fonctionnel d'origine dans le NST. Vert : territoire sensorimoteur. Rouge : territoire limbique. Bleu : territoire associatif.

#### **Explication détaillée :**

L'utilisation de connectomes normatifs est croissante depuis plusieurs années et, suivant les qualités de résolution des images, a l'avantage de fournir des données difficilement atteignables dans certaines populations de patients pour lesquelles de longues acquisitions d'image ne sont pas réalisables. Cette approche semblerait également donner des résultats équivalents aux acquisitions individuelles dans la maladie de Parkinson (Q. Wang et al., 2021).

Le connectome contient dans chaque voxel une modélisation mathématique de l'anisotropie de l'eau au moyen d'une *Fibre Orientation Distribution* (FOD). Ces FOD présentent une forme géométrique polylobée dont la taille de chaque lobe représente une mesure de l'importance de la diffusion de l'eau dans cette direction (figure 3).



**Figure 8.** Représentation schématique d'une FOD dans un voxel; chaque taille de lobe correspond au poids représenté par l'anisotropie de l'eau dans cette direction (depuis R. E. Smith et al., 2013).

### Tractographie des VTA avec le cortex

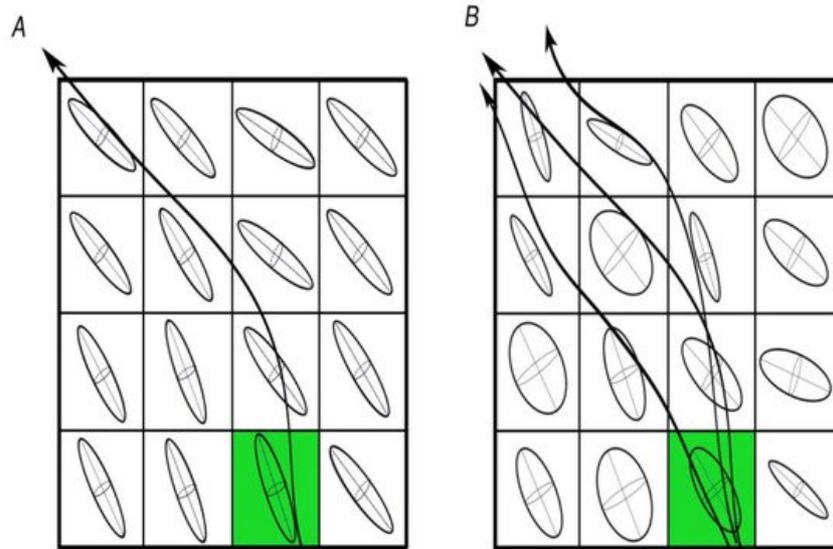
A partir du connectome normatif, la tractographie allant de chaque VTA au cortex était calculée. Chaque VTA était alors recallée dans le même espace que le connectome, et des lignes de champs ou *streamlines* étaient générées par une méthode probabiliste avec pondération de chaque *streamline* utilisant la méthode SIFT2. Cette méthode permet de corriger les faux positifs créés par la méthode probabiliste et d'apporter une valeur quantifiable du nombre estimé de fibres dans chaque *streamline*.

#### Explication détaillée :

Le terme *streamline* se distingue ici de fibres car les dimensions d'un voxel sont différentes d'un facteur 1000 par rapport au diamètre axonal. Une *streamline* représenterait donc plutôt la direction majoritaire d'un ensemble de fibres dans les limites de la résolution de l'image.

La tractographie probabiliste est une méthode permettant d'améliorer la sensibilité de la tractographie. Elle s'oppose à une méthode déterministique. Dans cette dernière, chaque direction de la *streamline* est calculée de proche en proche en suivant la direction principale de la FOD, elle crée donc un nombre limité de *streamlines* au risque de ne pas détecter des faisceaux de fibres ayant un trajet trop complexe. Dans l'approche probabiliste, les

*streamlines* sont calculées avec une orientation pondérée par la probabilité de distribution de chaque voxel environnant (figure 4). Ainsi, le nombre de *streamlines* produites est plus grand et tient plus compte de la complexité anatomique de la substance blanche (comme des fibres croisées ; Maier-Hein et al., 2017).



**Figure 9.** Représentation schématique d'une technique de tractographie déterministe (A) et probabiliste (B), (d'après Rodrigues et al., 2018).

Toutefois, pour éviter le nombre de faux positifs, un filtrage des *streamlines* est nécessaire. Pour cela, il existe des méthodes de pondération permettant a posteriori de calculer la vraisemblance dans chaque voxel que le nombre de *streamlines* soit cohérent avec la direction de chaque lobe d'une FOD. Cela permet de pondérer l'importance de chaque *streamline*, c'est-à-dire une valeur quantifiable estimant la vraisemblance qu'une ou plusieurs fibres axonales y passe réellement (R. E. Smith et al., 2015).

### Obtention d'une carte de densité de fibres

Après avoir tracé la connectivité de chaque patient à partir de leurs VTA, la somme des poids des *streamlines* (équivalent à une densité de fibres) était intégrée à chaque voxel de l'image.

Cela permet alors de générer une estimation quantifiable de la densité de fibres par voxel et rendre possible une analyse inter-sujets.

### Constitution de groupes

Afin d'évaluer l'effet de la stimulation cérébrale profonde sur la préférence pour le choix, nous avons utilisé la différence entre la préférence pour le choix ON moins OFF (« delta »), moyenné sur l'ensemble des blocks de probabilité. Nous avons ensuite divisé notre cohorte en utilisant la médiane, entre ceux ayant un faible effet de la stimulation (delta bas) et ceux ayant un effet élevé de la stimulation (delta élevé).

Chaque patient a ensuite été étiqueté dans l'un des deux groupes correspondants, les patients présentant une faible différence de choix libre ON-OFF étant étiquetés comme groupe de référence.

#### **Explication détaillée :**

Cette catégorisation était justifiée par la distribution asymétrique de la différence ON-OFF, avec un pic autour de zéro et des valeurs plus distribuées au-dessus. Étant donné la ségrégation spatiale substantielle de la connectivité corticale dans le NST, nous avons supposé que la localisation stratégique des VAT et de leur connectivité connexe pouvait être à l'origine de cet effet.

### Analyse statistique voxel à voxel

Pour l'analyse statistique, nous avons a priori qu'une connectivité différente des VTA avec les régions non-motrices expliquait les résultats entre les deux groupes. Pour cela, nous avons utilisé un masque cortical correspondant aux régions limbiques et associatives connectées au NST telles que définies par Temiz et al. (2020).

Une fois la carte de densité correspondant à ce masque obtenue, une analyse voxel à voxel était effectuée en utilisant l'outil PALM (Winkler et al., 2014). Pour cela, à l'échelle d'un voxel, les valeurs du groupe ayant un « delta bas » étaient comparées à celles du groupe ayant un « delta haut ». Du fait de multiples comparaisons, une correction par permutations

( $n=1000$ ) utilisant le *famywise error rate* (FWER) a été utilisée par la méthode threshold-free cluster enhancement (TFCE; S. M. Smith & Nichols, 2009).

### **Explications détaillées**

L'utilisation de permutations permet de tester simplement en observant, après permutation des observations, combien de fois la différence entre les moyennes dépassait la différence trouvée sans permutation. Ainsi, toute erreur liée au hasard (risque alpha) provenant de cette distribution par permutation permet de calculer une valeur  $p$  corrigée par FWER. Cette valeur  $p$  correspond à la probabilité d'avoir au moins une valeur significative liée au hasard (risque alpha) pour ce voxel. Il s'agit donc là d'une méthode robuste de comparaisons multiples mais très conservateur.

La méthode TFCE permet de calculer automatiquement des clusters au sein d'une image donnée. Cela permet d'avoir plus de puissance qu'à l'échelle du voxel, tout en gardant une correction robuste des comparaisons multiples par FWER. Cette dernière s'effectue en fait dans les voxels correspondant aux pics des différents clusters et non à tous les voxels ayant une valeur  $p$  significative. Cette méthode est implémentée de manière courante dans le logiciel PALM et devient largement utilisée.

## Etude 2 : résultats

### Données démographiques

En moyenne, les patients du groupe DOPA étaient légèrement plus âgés que ceux du groupe NST mais cette différence n'était pas significative, tout comme l'efficacité cognitive globale (MoCA), la durée d'évolution, ou la sévérité de la maladie de Parkinson (UPDRS OFF). Les patients du groupe NST avaient un score UPDRS ON plus élevé que celui du groupe DOPA. Comme attendu, la dose quotidienne équivalente de lévodopa (LEDD) était significativement plus faible pour les patients du groupe NST.

	groupe DOPA moyenne (DS)	groupe NST moyenne (DS)	Valeur <i>p</i>
Age (années)	56,8 (8,6)	54,2 (7,8)	0,226
MoCA (/30)	28,2 (1,5)	27,5 (1,5)	0,087
Evolution (années)	10,4 (4,3)	12,3 (3,3)	0,113
UPDRS III (OFF)	43,5 (14,5)	47,7 (16,3)	0,355
UPDRS III (ON)	10 (7,3)	16,6 (9,9)	<b>0,012</b>
LEDD	1355 (525)	404,6 (224)	<b>&lt;0,0001</b>

Tableau 2. Tableau récapitulatif des principales données démographiques.

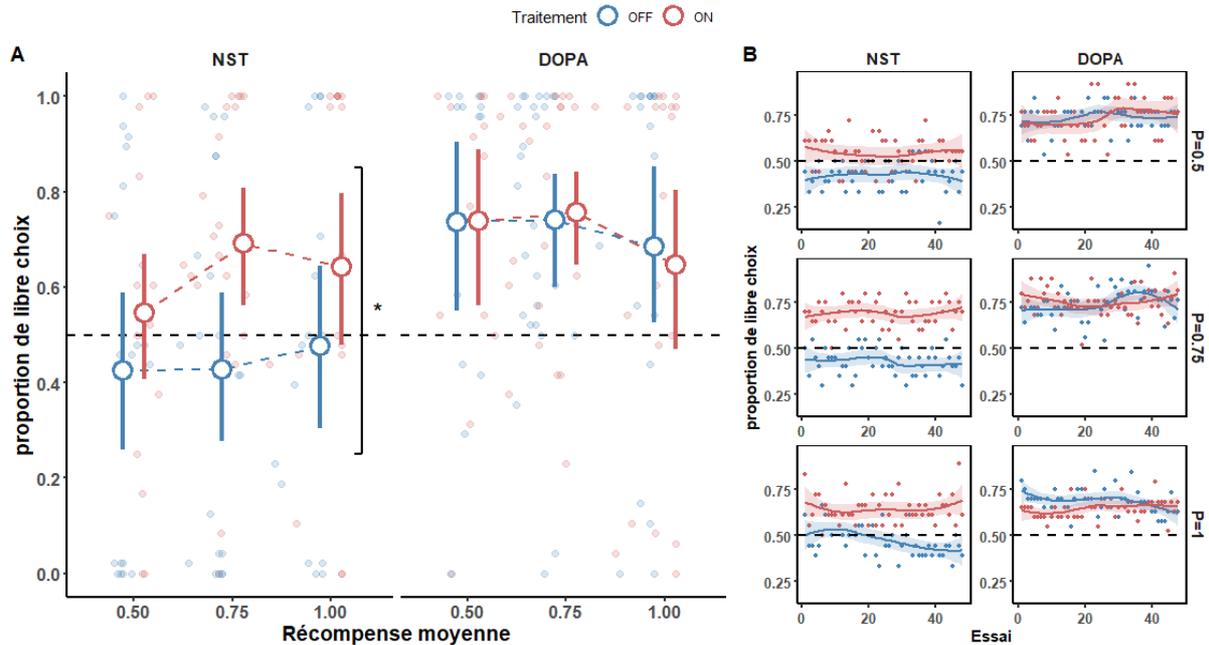
### Première étape : préférence pour le choix

Le modèle incluait le groupe, la condition de traitement, la moyenne des récompenses obtenues pour chaque bloc et leurs interactions. En utilisant la moyenne des récompenses obtenues plutôt que la probabilité de récompense théorique, le modèle prend en compte la variabilité des sujets ayant eu un apprentissage sub-optimal.

Nous avons d'abord regardé les préférences moyennes pour le choix dans la première étape de la tâche (figure 9A). L'interaction traitement x groupe était significative ( $p=0,034$ ). L'interaction traitement x groupe x récompense obtenue n'était pas significative ( $p=0,570$ ). L'effet principal de la récompense obtenue était non significatif ( $p=0,051$ ).

L'analyse post-hoc intra-groupe retrouvait une différence significative dans le groupe NST entre les patients ON (63% de choix libre) et OFF (44%), ( $p=0,011$ ). Par ailleurs, la préférence pour le choix OFF ne différait pas du seuil d'indifférence (44% *vs.* 50%,  $p=0,339$ ). Dans le groupe DOPA, la différence était non significative avec de hauts niveaux de préférence pour le choix aussi bien en ON qu'en OFF (72% : OFF, 71% : ON,  $p=0,996$ ). L'analyse post-hoc inter-groupe a montré que les patients NST-OFF avaient une préférence pour le choix significativement plus basse que les patients DOPA aussi bien en ON ( $p=0,014$ ), qu'en OFF ( $p=0,032$ ).

En visualisant la dynamique des proportions de choix libre pour chaque bloc de probabilité (figure 9B), les patients expriment précocement leur préférence pour le choix dans les deux groupes, aussi bien en ON qu'en OFF, et cette préférence variait peu tout au long de la session.



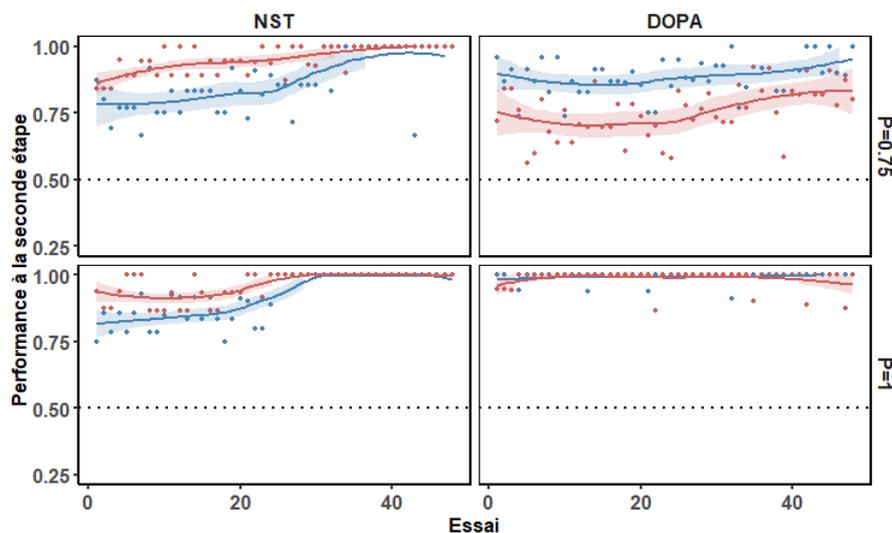
## Seconde étape : choix de l'image la plus récompensée

Nous avons ensuite vérifié si les différences observées à la première étape étaient influencées par l'apprentissage des contingences des récompenses à la seconde étape et si une différence en termes de récompense pouvait expliquer ce résultat. Par exemple, dans le bloc  $P=1$  et  $P=0,75$ , une des deux images était plus récompensée que l'autre. Le modèle comprenait alors le groupe, le traitement et la probabilité de récompense.

L'interaction groupe x traitement était significative ( $p<0,001$ ) alors que groupe x traitement x probabilité de récompense ne l'était pas ( $p=0,185$ ).

L'analyse post-hoc a montré que dans les deux groupes, les sujets choisissaient majoritairement l'image la plus récompensée (DOPA=82%,  $p < 0,0001$  ; NST=81%,  $p < 0,00001$ ), ce qui indique que l'apprentissage était efficace.

Nous avons toutefois constaté une différence significative dans le bloc  $P=0,75$ , les sujets DOPA-ON ayant de moins bonnes performances que les sujets DOPA-OFF (74% vs 88%,  $p=0,046$ ). A l'inverse, les patients NST-OFF étaient moins bon que les patients NST-ON pour ce bloc (83% vs 94%,  $p=0,001$ ). Par ailleurs, la dynamique des courbes d'apprentissage dans chaque groupe et chaque bloc de probabilité était différente de l'expression du choix libre (figure 11).

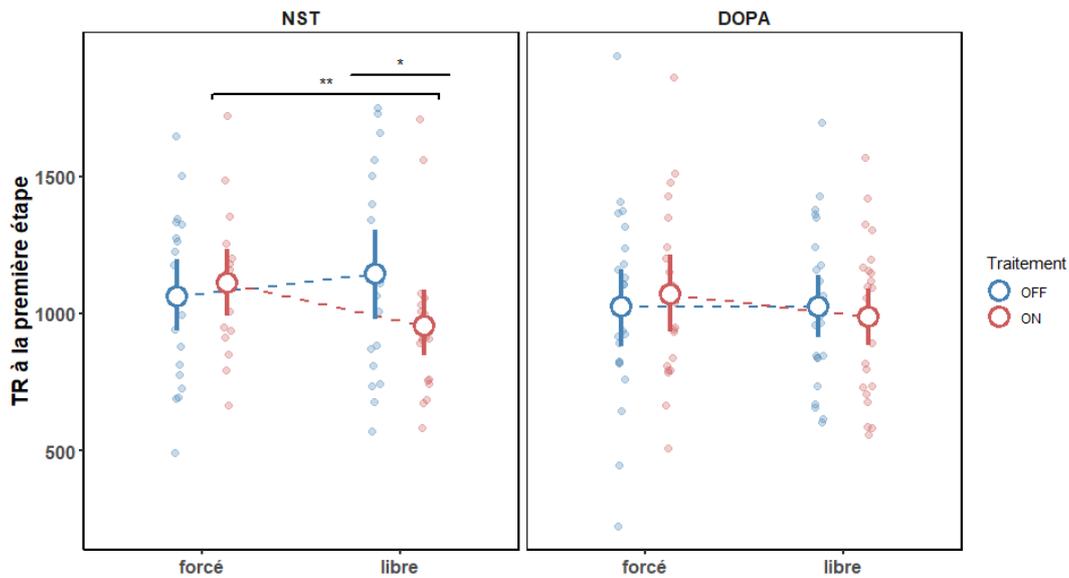


**Figure 32.** Dynamique de sélection de l'image la plus récompensée à travers la tâche par bloc, essai par essai. Les points rouges (ON) et bleus (OFF) représentent la moyenne de la sélection de l'image essai par essai parmi tous les sujets. Les lignes rouges et bleues représentent un lissage type LOESS et leurs erreurs standard respectives.

## **Temps de réactions entre les essais « choix libre » et « choix forcé » à la première étape**

Le modèle comprenait le groupe, le traitement et la sélection réalisée à la première étape « libre » ou « forcé ».

L'interaction groupe x traitement x sélection à la première étape était significative ( $p=0,0006$ ). L'analyse post-hoc a retrouvé que les sujets du groupe NST sélectionnaient plus rapidement le choix libre que le choix forcé en ON (différence :  $-230$  ms,  $p=0,0001$ ) et étaient plus lent en OFF qu'en ON pour sélectionner le choix libre ( $196$  ms,  $p=0,035$  ; figure 11).



**Figure 33.** Temps de réaction lors de la première étape. Les points rouges (ON) et bleus (OFF) représentent les temps de réaction (TR) individuels et les cercles rouges et bleus vides représentent la moyenne des temps de réactions en fonction des essais libres ou forcés pour les patients ON et OFF respectivement. Les barres d'erreurs correspondent à l'intervalle de confiance à 95%. Pour le groupe NST, le crochet avec deux étoiles représente une différence significative entre les patients libres et forcés pendant ON ( $p=0.0001$ ). La ligne horizontale avec une étoile représente la différence significative entre ON et OFF pour le choix libre ( $p=0.035$ ).

## Temps de réactions entre les essais « choix libre » et « choix forcé » à la seconde étape

Le modèle était le même que celui réalisé à la première étape.

L'effet principal traitement ( $p=0.007$ ) et la probabilité de récompense ( $p=0.005$ ) étaient significatifs, sans interaction significative avec le groupe de patients ( $p > 0.176$ ).

Les tests post-hoc ont montré que les temps de réaction étaient plus rapides sous traitement que sans traitement (92 ms NST, -31 ms DOPA,  $p=0,007$ ), et plus rapides lorsque la probabilité de récompense était plus élevée (différence à  $P=0,50$  comparée à  $P=1,-36$  ms DBS, - 41,5 ms DOPA,  $p=0,003$ ).

## **Analyse de la connectivité des VAT avec le cortex cérébral**

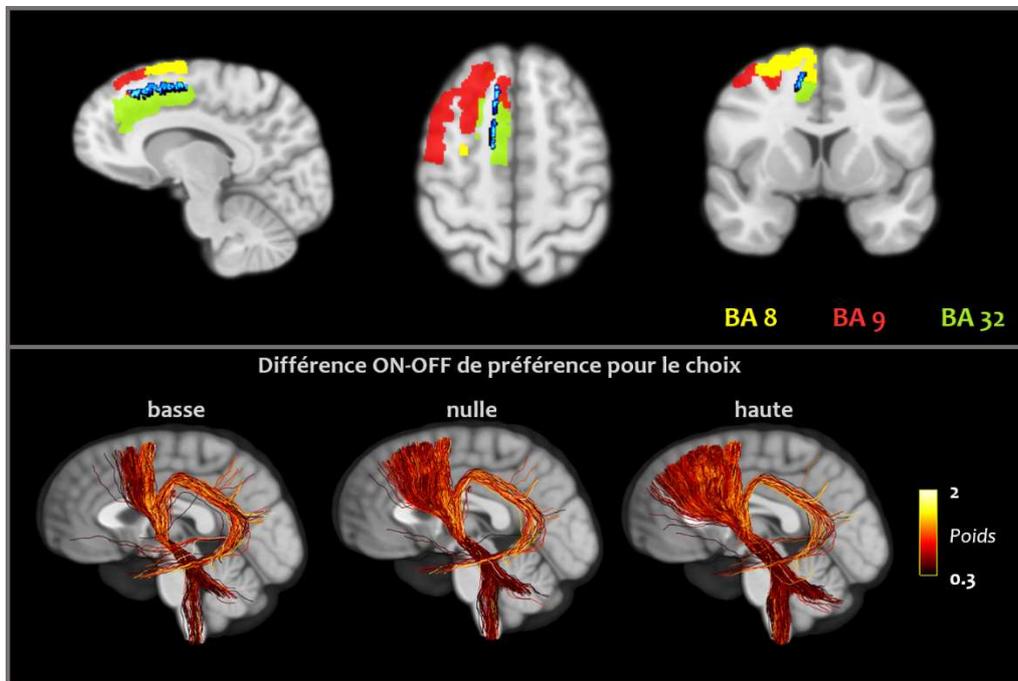
Dans le groupe des patients opérés, il n'y avait pas de différence significative de localisation des contacts actifs dans le NST dans les plans x, y et z.

Nous avons aussi comparé les différences des coordonnées des contacts actifs entre l'hémisphère droit et l'hémisphère gauche. Les contacts actifs de l'hémisphère droit étaient significativement plus antérieurs que ceux de l'hémisphère gauche ( $y_{\text{droit}}=10,8$  mm,  $y_{\text{gauche}}=10,2$  mm par rapport à la CP,  $p=0,004$ ). Il n'y avait pas de différence significative de coordonnées entre les deux hémisphères dans le plan axial ( $p=0,477$ ) et le plan coronal ( $p=0.298$ ).

Nous avons comparé les VAT, dont la différence n'était pas significative entre les deux hémisphères bien que les VAT du côté droit aient un volume plus grand (gauche : 79,2 mm<sup>3</sup>, droite : 94,9 mm<sup>3</sup>,  $p=0.231$ ).

L'analyse voxel à voxel de la connectivité entre les différentes VAT et le cortex préfrontal tel que décrit dans les méthodes a retrouvé une différence de connectivité avec la région située dans le cortex préfrontal médian droit et le cortex cingulaire antérieur dorsal droit (CCA dorsal) de 436 mm<sup>3</sup>. Cette région était significativement recrutée chez les patients du

groupe NST présentant une différence ON-OFF pour la sélection du choix libre ( $p < 0.05$  corrigé FWER-TFCE ; figure 13).



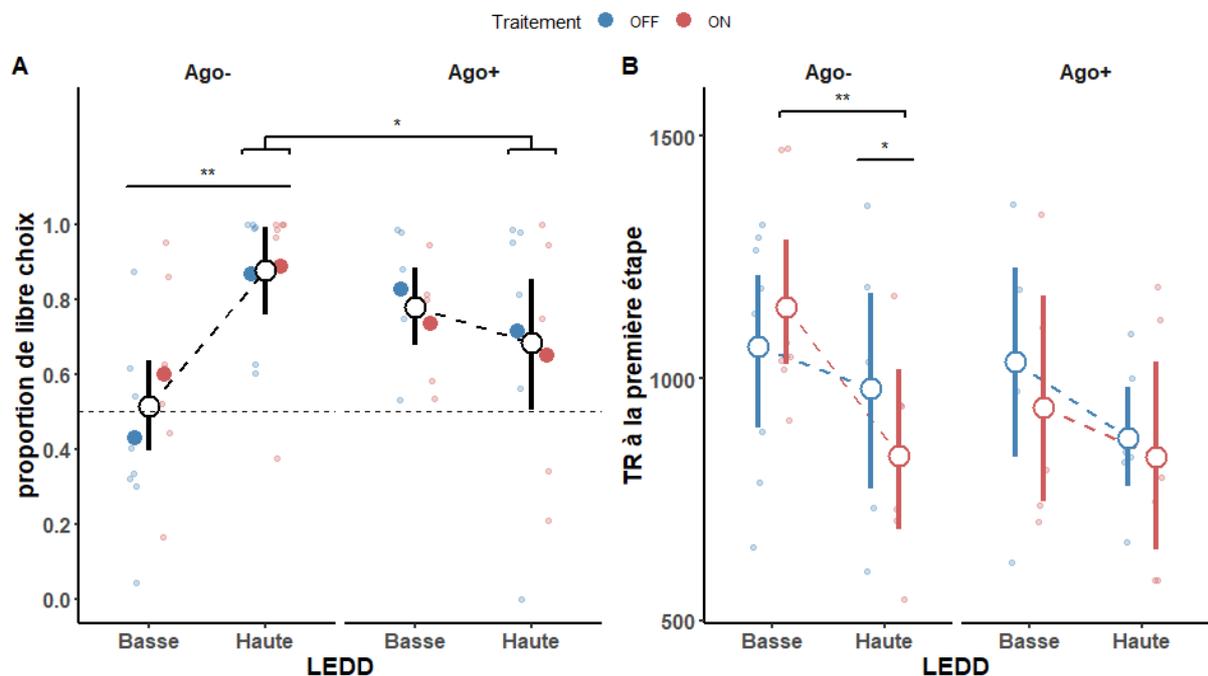
**Figure 34.** Analyse de connectivité entre les patients présentant une différence ON-OFF pour le choix libre élevée et ceux ayant une différence réduite. Panel du haut : représentation en bleu de la carte statistique seuillée à  $p < 0.05$ . Les aires de Brodmann 8 (jaune), 9 (rouge) et 32 (vert) sont également représentées à type indicatif. Panel du bas : projection des *streamlines* pour trois patients ayant respectivement une différence ON-OFF basse (-0,23), nulle (0) et haute (+0,97). L'échelle correspond au poids de la connectivité pour chaque *streamline* calculée avec la méthode SIFT2.

## Analyse en sous-groupes des patients DOPA

Du fait de l'absence d'effet aigu ON/OFF dans la préférence pour le choix, nous avons regardé si les différences de doses de traitement au long cours étaient capables d'influencer la préférence pour le choix libre.

Les patients étaient divisés par la médiane de la LEDD en deux groupes ayant respectivement moins ou plus de 1050 mg de LEDD. La prise de traitement par agonistes dopaminergiques était également prise en compte. Le modèle incluait ainsi le facteur LEDD (supérieur ou inférieur à 1050 mg), le facteur prise d'agonistes (Ago- ou Ago+).

L'interaction LEDD x Ago était significative ( $p=0,01$ ). L'analyse post-hoc a montré que les patients Ago- présentait une préférence pour le choix moindre lorsqu'ils avaient une dose LEDD basse (80% vs. 54%,  $p=0,002$  ; figure 13A). Par ailleurs, parmi les patients ayant une LEDD élevée, les patients Ago- sélectionnaient le choix libre plus souvent que les patients Ago+ (80% vs. 69%,  $p=0,036$ ). Par ailleurs, le traitement aigu (ON/OFF) n'interagissait avec aucun autre facteur ( $p>0.216$ ) et l'effet de la probabilité de récompense était non significatif ( $p=0.55$ ).



**Figure 35.** A : préférence pour le libre choix en fonction de la LEDD (>1050 ou <1050g). Les petits points rouges (ON) et bleus (OFF) représentent la préférence moyenne pour le libre choix au niveau individuel. Les gros points rouges et bleus représentent la moyenne de tous les sujets, respectivement pour les conditions ON et OFF. Les cercles vides entourés de noir représentent la performance moyenne en matière de libre choix pour tous les patients en fonction de la LEDD et la prise d'agoniste. Les barres d'erreurs correspondent à l'intervalle de confiance à 95%. La ligne horizontale en pointillés représente le niveau d'indifférence. Les crochets surmontés d'une étoile représentent une différence significative à  $p=0,036$  entre les patients Ago- et Ago+ chez ceux ayant une LEDD haute. La barre horizontale surmontée de deux étoiles représente une différence significative à  $p=0,002$  entre les patients ayant une LEDD basse et haute chez les Ago-. B : temps de réaction à la première étape en fonction de la LEDD. Le code couleur est identique à la figure 13A. Les crochets avec deux étoiles représentent une différence significative en ON entre les patients ayant une LEDD basse et ceux ayant une LEDD haute ( $p=0,010$ ). La barre horizontale surmontée d'une étoile représente une différence significative entre les patients ON et OFF chez les Ago- ayant une LEDD haute ( $p=0,0439$ ).

Finalement, nous avons examiné les temps de réaction à la première étape pour les sous-groupes de patients DOPA.

L'interaction LEDD x traitement était significative ( $p=0,01$ ) et l'interaction Ago x LEDD x traitement présentait une valeur  $p$  marginale ( $p=0,059$ ).

Les analyses post-hoc ont montré que les patients sans agonistes et avec une LEDD élevée étaient plus rapides en ON qu'en OFF ( $p=0,044$ ). Les patients ON sans agonistes étaient également plus rapides lorsqu'ils avaient une LEDD élevée comparés à ceux ayant une LEDD basse ( $p=0,010$ ).

## Conclusion de l'étude 2

Au moyen du même paradigme que celui de l'étude 1, nous avons mis en évidence l'implication du système dopaminergique et du NST dans le traitement de récompenses intrinsèques. Premièrement, l'hyperactivité du NST, comme c'est le cas chez les patients stimulés en OFF, altère la proportion de choix qui se trouve au seuil d'indifférence. Cette altération est ensuite rétablie par la stimulation chez les patients ON, qui de plus accélèrent leur réponse à la première étape. De plus, un réseau entre le site de stimulation des patients et le CCA dorsal semble être impliqué dans la préférence pour le choix. Deuxièmement, les patients traités oralement présentent une haute préférence pour le choix aussi bien en ON qu'en OFF mais la prise chronique de L-DOPA semble influencer les performances des patients. Ainsi, lorsque le traitement dopaminergique quotidien est bas, les patients présentent également une préférence pour le choix proche du seuil d'indifférence, ce qui est rétabli par la prise de hautes doses de traitement sur le long terme. Ces hautes doses de traitement accélèrent également les réponses des sujets, indépendamment de leur sélection à la première étape. Ces différents résultats sont discutés dans le chapitre suivant.

## Article de l'étude 2

### Deep brain stimulation of the subthalamic nucleus and dopamine medication enhance intrinsic motivation for free choice in Parkinson's disease patients

#### Authors

David Bendetowicz<sup>1,2</sup>, Elodie Hainque<sup>2</sup>, Virginie Czernecki<sup>1,2</sup>, Gizem Temiz<sup>1</sup>, Carine Karachi<sup>1,3,\*</sup>, Brian Lau<sup>1,\*</sup>, Jérôme Munuera<sup>1,4,\*</sup>

1. Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Inserm, CNRS, APHP, Paris, France

2. AP-HP, Hôpital de la Pitié Salpêtrière, Neurology department, Paris, France

3. AP-HP, Hôpital de la Pitié Salpêtrière, Neurosurgery department, Paris, France

4. Institut Jean Nicod, Département d'études cognitives, École normale supérieure, EHESS, CNRS, PSL University, Paris, France

\*Corresponding authors: [carine.karachi@icm-institute.org](mailto:carine.karachi@icm-institute.org); [brian.lau@upmc.fr](mailto:brian.lau@upmc.fr); [jerome.munuera@icm-institute.org](mailto:jerome.munuera@icm-institute.org)

**Keywords:** motivation, decision-making, dopamine, deep brain stimulation, Parkinson

**Abstract:**

Human decisions are frequently explained as the balancing of potential rewards and punishments, such as food, money, or lost time. However, rational models of decision making based on this idea often fail to accurately predict human behavior. Humans and other animals appear to frequently base decisions on value attribution that is often not clearly linked to extrinsic outcomes. For example, humans prefer opportunities to choose, even when these have no positive impact on future rewards or punishments. Thus, access to free choice may be intrinsically rewarding, a form of motivation that invigorates an agent to perform behaviors in and of themselves. There is limited evidence about the neural mechanisms underlying intrinsic motivation, and it is unclear how deficits in intrinsic motivation manifest themselves following impairment of brain networks involved in processing extrinsic reward. Here we seek to understand how extrinsic and intrinsic motivations act together during decision making. In order to investigate the neural mechanisms underlying this integration, we studied Parkinson's disease (PD) patients on and off treatment, which allowed us to manipulate the dopaminergic system and prefrontal cortices that are involved in processing of extrinsic reward. In one group, treatment consisted of acute dopamine replacement therapy, and in second group treatment consisted of deep brain stimulation (DBS) of the subthalamic nucleus (STN). We found that PD patients on treatment preferred opportunities to choose even though these did not afford more extrinsic rewards. However, patients lose this preference in conditions that diminish the quantity of dopamine in their brains or by reducing the activation of medial prefrontal cortical areas via deep brain stimulation. These results advance our understanding of the neurocognitive determinants of intrinsic reward and its interplay with general motivational processes by unmasking motivational mechanisms associated with behavioral impairments observed in PD and other neuropsychiatric disorders involving the dopaminergic system.

## Introduction

Although Parkinson's disease (PD) has traditionally been associated with motor symptoms, non-motor symptoms are increasingly recognized as an important component of the disease. These include neuropsychiatric manifestations such as depression, anxiety, apathy, psychosis and cognitive impairment, which can be aggravated or even triggered by treatments such as dopamine replacement therapy and deep brain stimulation (DBS; Limousin & Foltynie, 2019; Voon et al., 2006; Weintraub & Mamikonyan, 2019). While PD motor symptoms are directly linked to the death of midbrain dopamine (DA) neurons (Damier et al., 1999; Fearnley & Lees, 1991). These neurons are also important in the processing of extrinsic rewards and punishments (H. Matsumoto et al., 2016; M. Matsumoto & Hikosaka, 2009; Schultz, 1998). The discovery that DA neurons encode a reward prediction error has led to the development of a powerful reinforcement learning framework for understanding how DA and basal ganglia dysfunction produce learning and decision-making deficits in PD (Schultz, 1998; Sutton & Barto, 2018).

Acute administration of dopaminergic medication in PD patients can enhance learning from positive rewards and reduce learning from negative outcomes (Frank et al., 2004; Palminteri et al., 2009; Rutledge et al., 2009). On the other hand, deep brain stimulation (DBS) of the subthalamic nucleus (STN) in PD patients can lead to impulsive choices (Ballanger et al., 2009; Cavanagh et al., 2011; Frank et al., 2007). While these manipulations highlight the role of DA and associated basal ganglia circuits in outcome-based decision-making, they focus exclusively on extrinsic outcomes (i.e., feedback indicating correctness or explicit outcomes such as money or points earned). Decisions may also be *intrinsically motivated*, that is, driven by an inherent satisfaction distinct from extrinsic rewards or punishments (Ryan & Deci, 2000). For example, studies in humans and other animals show that subjects prefer access to opportunities to choose, even when this has no impact on immediate extrinsic outcomes (Catania & Sagvolden, 1980; Munuera et al., 2022; Suzuki, 1999; Voss & Homzie, 1970). Although choice preference appears economically irrational, it is proposed that opportunities to choose are intrinsically rewarding because they offer potential control over ones' environment (Munuera et al., 2022; Ryan & Deci, 2000).

Personal control is central in numerous psychological theories, where constructs such as autonomy (Deci & Ryan, 1985; Ryan & Deci, 2000), controllability (Bandura et al., 1999; Maier & Seligman, 1976), personal causation (deCharms, 1968), effectance (White, 1959), perceived behavioral control (Ajzen, 2002) or self-efficacy (Bandura et al., 1999) are key for motivating behaviors that are not economically rational or easily explained as satisfying basic drives such as hunger, thirst, sex, or pain avoidance (Hull, 1943).

The DA system may be important in the processing of both extrinsic and intrinsic rewards and punishments. This is supported by the observation that midbrain DA neurons encode the inferred value of visual cues that are informative, without changing the rate of extrinsic reward (Bromberg-Martin & Hikosaka, 2009). Striatal activation is enhanced when subjects can choose freely (compared to forced selections of equivalent value), or even simply when subjects know that free choice is available (Leotti & Delgado, 2014; Tricomi et al., 2004). In addition, humans expressing more DARPP-32, a gene linked to striatal dopaminergic plasticity, show stronger preference for choosing items they were free to explore compared to items they were forced to sample (Cockburn et al., 2014). It remains unclear whether intrinsic rewards such as the opportunity to choose are altered by dysfunction of the DA system or through causal manipulation of associated basal ganglia structures.

In order to investigate the role of DA and basal ganglia networks in the intrinsic motivation for choice opportunities, we designed a task where subjects revealed their preference by selecting between free- and forced-choice options. We addressed the question of the involvement of the dopaminergic and basal ganglia systems in free choice preference of PD patients. We tested two separate population of patients: with dopamine replacement therapy or DBS of the STN. We tested both groups with and without treatment and found that STN-DBS patients with strong sub-cortico-cortical connectivity dramatically increased their free choice premium, while the amount of dopamine replacement therapy differentially modulates free choice preference.

## Methods

## Participants:

We recruited 45 participants through the neurology and neurosurgery services at the Pitié-Salpêtrière Hospital (Paris, France). Parkinson's disease was diagnosed following the United Kingdom Parkinson's Disease Society Brain Bank clinical diagnostic criteria (Hughes et al., 1992). Exclusion criteria included the presence of a significant cognitive impairment (MoCA  $\leq$  25-30) or concomitant significant neurological or psychiatric disease. For our first group of patients (DOPA group, n=25, 7 females), inclusions were assessed during a visit at the hospital for evaluation of eligibility for DBS surgery. Patients who already underwent STN-DBS surgery constitute our second group (STN-DBS group, n=20, 9 females) and were recruited (delay after surgery: 15 months  $\pm$  5.3 SD) during a routine follow-up. Eight were implanted with Medtronic 3389 electrode (Medtronic Inc., Minneapolis, MN, USA) and the remaining twelve were implanted with Boston Vercise Cartesia directional electrode (Boston Scientific, Marlborough, MA). Among those, five of them had steering current on at least one of the two electrodes (one electrode per hemisphere for all patients). Details of stimulation parameters are provided in Table S1. All the subjects were fully informed of all the aspects of the study and gave consent in accordance with the declaration of Helsinki.

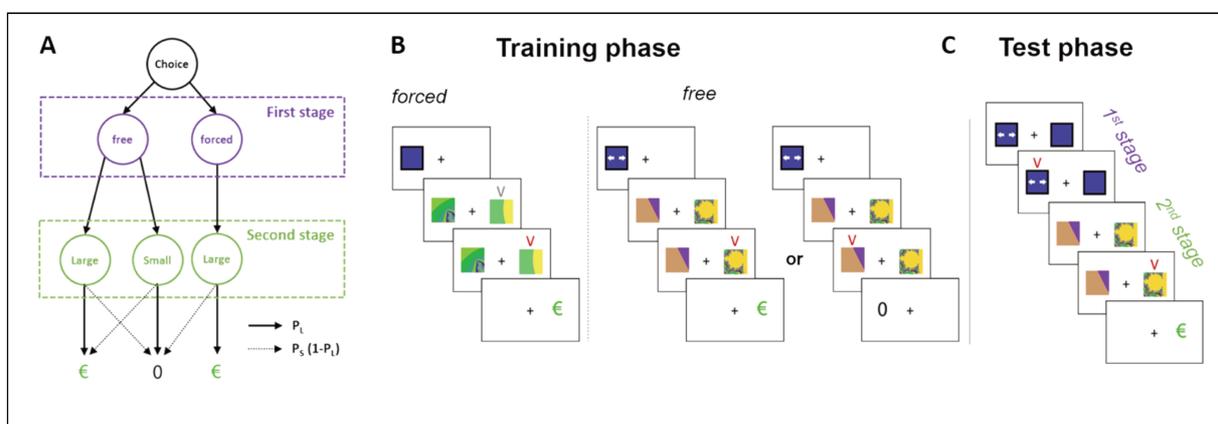
## Task description:

Patients were seated in front of an LCD screen (15,6 inches, resolution 1920 x 1080) and were told that they would be presented with different pairs of fractal images (targets), which were associated to different probabilities of being rewarded. The goal of the task was to accumulate the maximal amount of reward symbolized by a green euro (€) on the screen. Patients could select these fractal targets in two different ways: either by choosing themselves (*free*) or guided by the computer (*forced*). They received oral and written instructions and before starting the task they performed a dozen trials to familiarize

themselves with the task structure. Patients selected targets using the left and right key press on the computer keyboard.

The task consisted of a series of trials each with a two-stage structure. For each trial, during the first stage, two image targets were displayed on the screen representing *free* or *forced* options and patients had to make an initial selection. On the second stage, two different targets were displayed. In the *free* option, patients could choose any of the two second-stage targets, while in the *forced* option they must select the one that was preselected by the computer. This mandatory second-stage forced target was indicated by a grey V-shape above the image. During this *forced* condition, the computer always selected the same target and selection of the opposite image was disabled. During both stages, after pressing one of the two keys, patients received a visual cue displayed above their selected target (red V-shape) that validated their selection. After selecting the second-stage target, the reward outcome was displayed and could be either the green euro symbol when the trial was rewarded or a black zero when the reward was absent (Figure 1).

Importantly, before each testing phase, patients performed a training phase in order to learn the contingencies between the different targets and their associated reward probabilities. During this phase patients did not choose between *free* or *forced* options and they therefore performed mandatory *free* or *forced* trials in succession (e.g., 48 forced then 48 free trials or 24 trials in free and forced when the reward delivery were deterministic, see below). Patients were explicitly informed that the stimuli remained identical between training and testing phases.



**Figure 1. Two-stage task structure.** **A.** Diagram illustrating the different pathways that subjects could go through during the task. At the second stage, “Large” and “Small” represent the most rewarded and the less rewarded targets.  $P$  represent the probabilities of reward delivery for the large (L) and small (S) second-stage targets. When extrinsic

rewards were delivered a green euro was displayed at the end of the trial. A black zero was associated to the non-rewarded trial. **B.** Illustration of typical trials during the training phase. For each probability block, subjects learned the reward contingencies associated to the second-stage targets in forced and free option. When training for forced choice, subjects had to select the target indicated by a V-shape grey symbol. When training for free choice, subjects could freely select one of the two targets. V-shape red symbol indicated subject's selection after keypress. **C:** schematic illustration of a trial during test phase. At first-stage, subjects had to select if they wanted to perform the trial either the free or forced option.

Reward contingencies varied between three blocks, a block being defined as one training follow by one testing phase (48 trials) with the same reward probability. In one block the reward delivery was stochastic and all the second-stage targets were associated with a reward probability of  $P = 0.5$ . In another block, the probabilities of reward for the forced second-stage target and the most rewarded second-stage target in free were set at  $P = 0.75$ , and at  $1-P$  for the less rewarded target in free. In the last block, reward delivery was deterministic and selecting the target in forced or the most rewarded in free always led to a reward ( $P=1$ ) while choosing the other target in free led systematically to its absence ( $P=0$ ). Thus, reward contingencies were identical between the free and forced options, but the image targets were different between options as well as in all the different blocks. Because, in forced option, the computer always selected the most rewarded target (when  $P > 0.5$ ), choosing the target in *forced* or choosing the most rewarded target in *free* was equivalent in term of reward accumulation. The task was written using the Psychophysics Toolbox (<http://psycho toolbox.org/>) running MATLAB (Natick, Massachusetts: The MathWorks Inc).

## Procedure:

Patient performed the task and all blocks twice, ON and OFF treatment, for both the DOPA and STN-DBS groups. The patients in the STN-DBS group were OFF medication for both sessions to only explore the effect of stimulation. All the image targets were also different between the OFF and ON sessions.

For the DOPA group, OFF session occurred following the local procedure that is after 12h after discontinuation of the treatment for levodopa and 72 h for dopamine agonists. This procedure was implemented to allow wash-out of all antiparkinsonian pharmacological

treatments. ON session occurred the same day, 45 minutes after taking the levodopa equivalent dose of their usual morning treatment plus 50 milligrams (mg) of levodopa. For the STN-DBS group, the wash-out of pharmacological treatments for both ON and OFF sessions were the same procedure as the DOPA group. ON session (stimulation ON) occurred first and OFF session start at least an hour after stimulation was stopped. The acute dopaminergic challenge and ON/OFF testing of DBS patients were part standard clinical follow-up.

### Statistical analyses:

We used mixed effect models to account for the nested nature of the data (lme4 package, R Core Team, <https://www.R-project.org>). Generalized linear mixed models (GLMMs) were used for binary variables (first- and second-stage choices). Main effects and interactions of the models were reported using Wald's z statistics and corresponding p-values. For reaction times, we used linear mixed models and reported t-statistics and corresponding p-values for main effects and interactions. Patients were treated as a random effect for all models. One slope for each fixed effect was added to the random part of the model when possible; slopes were added to the random component hierarchically until singularity of the model (Bolker et al., 2009; Harrison et al., 2018). Before analysis, continuous fixed effects were scaled across the whole data and reaction times were filtered above 2-fold median absolute deviation across the entire dataset. (Lachaud & Renaud, 2011). We included in the model the actual obtained reward probability that was calculated as the mean of all the rewarded trials during training and testing, for every block of probability. By doing so, we consider the variability between subjects that could have had suboptimal learning performances. Post-hoc statistics were calculated using the *emmeans* packages on R adjusted for multiple comparisons using Tukey method, p-values were calculated using a z-statistic as provided by default in the *emmeans* package in R (<https://CRAN.R-project.org/package=emmeans>). We used two-sample Wilcoxon tests for demographic statistics.

Subgroup analyses: We distinguished between the use of dopaminergic agonists and the total amount of remaining treatment (Levodopa Equivalent Daily Dose, LEDD). We used the median LEDD to dichotomize patients between those who received low ( $\leq 1050$  mg,  $n=13$ ) or high ( $>1050$  mg,  $n=12$ ) amount of LEDD. We used this transformation instead of continuous LEDD because the quantity of medication was very skewed and followed a bimodal distribution with one peak of LEDD at 971 and another at 1550 mg (Supplementary Figure S1). This categorization approach was also performed in a large scale study assessing behavioral changes in Parkinson's disease (Weintraub et al., 2010). Similarly, we considered the use of dopaminergic agonists as a categorical variable because of a quasi-equivalence between users (Ago+,  $n=11$ ) and non-users (Ago-,  $n=14$ ). Mean LEDD in patient with high LEDD was 1646 and 887 for those with low LEDD. One sample Wilcoxon tests for each group compared to the respective peaks of the bimodal distribution showed no statistical difference (high-LEDD:  $W=52$ ,  $p=0.675$ , low-LEDD:  $W=19$ ,  $p=0.126$ ).

### Image processing:

Preoperative MRI volumetric T1-weighted, T2-weighted FLAIR sequences and postoperative CT scans were used for the 20 patients of the STN-DBS group. Pre-processing pipelines was computed with LEAD-DBS software (Horn et al., 2019) using the default presets. We visually checked the outcome at every step of the procedure. CT images was co-registered to MRI T1-weighted sequences using a two-stage linear registration as implemented in Advanced Normalization Tools (Avants et al., 2008) and MRI FLAIR sequence was linearly co-registered to T1 using SPM12 (Penny et al., 2011). Pre- and post-operative acquisitions were spatially normalized into MNI\_ICBM\_2009b\_NLIN\_ASYM space (Fonov et al., 2011) using the SyN registration approach as implemented in Advanced Normalization Tools (Avants et al., 2008), adding supplementary two nonlinear SyN-registrations that consecutively focused on the basal ganglia as defined by subcortical masks in Schönecker et al. (2009).

DBS electrodes localizations were corrected for brainshift in postoperative acquisitions by applying a refined affine transform calculated between pre- and post-operative acquisitions

that were restricted to a subcortical area of interest as implemented in the brainshift-correction module of Lead-DBS software (Horn & Kühn, 2015). Electrodes were automatically pre-localized using the PaCER algorithm (Husch et al., 2018). For one patient, the PaCER algorithm failed and TRAC/CORE was used instead (Horn & Kühn, 2015). Visual verification of the reconstruction and manual correction were performed if necessary. For patients with directional leads, orientation was determined using the implemented algorithm (DiODE) in Lead-DBS (Dembek et al., 2021). Coordinates were then converted from the MNI space into AC/PC coordinates using LEAD-DBS (Horn, Kühn, et al., 2017). After normalization and localization of electrode trajectories, volume of activated tissue (VAT) for each electrode were estimated using a finite element approach based on a four-compartment tetrahedral mesh as detailed in Horn et al. (2017). The resulting estimated E-field was thresholded above 0.2 V/mm in order to obtain finite volumes.

### Connectivity analysis:

The use of normative connectome has been used in various studies for assessing connectivity in PD patients (Horn, Reich, et al., 2017; Wang et al., 2021). Here we used a high-definition normative connectome issued from a local cohort of 33 patients with PD (Temiz et al., under revision, 2022). Following standard procedure describe in Temiz et al (2022), high-resolution multishell DWI data were acquired for each subject. Then fiber orientation diffusions (FOD) were calculated at the individual level and aligned to create an FOD template. Whole brain tractography using a probabilistic approach were used on the FOD template and Brodmann areas were used for segmentation of areas of interest. The use of a Brodmann segmentation is justified by a previous work on the cortico-subthalamic pathway parcellation that related all the Brodman areas to one of the three anatomic-functional territories of the STN (limbic, associative and sensorimotor). By doing so, cortical mask for each of the three territories could be provided (Temiz et al., 2020). A weighting factor for each streamline were computed using the Spherical informed filtering techniques (SIFT2) and applied to the whole brain tractography. Assigning weights to the streamlines

using SIFT2 is an useful solution to provide a reliable and quantitative estimation of connectivity (R. E. Smith et al., 2015).

Cortical connectivity was extracted from the whole brain tractography for each VTA after their spatial co-registration in the normative space of the template. For statistical analysis, track-density maps containing at every voxel the SIFT2 values were extracted from the whole brain tractography using a cortical mask corresponding to the conjunction of associative and limbic territories (or non-motor regions). These density maps are 3D images that reflect the number of streamlines terminations within that voxel. Those cortical track-density maps for each patient were then merged in a 4D volume for statistical analysis.

We performed voxelwise analysis using PALM from FSL (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/PALM>) to compute statistical maps using permutation methods. Two-sample t-tests were performed for each voxel of the volume, with 1000 permutations. P-values were corrected using familywise error (FWE) with the threshold-free cluster enhancement approach (TFCE) that could provide reliable correction at the voxel level (S. M. Smith & Nichols, 2009). The TFCE approach allows data driven clustering without a priori but also permits robust FWE correction at the voxel level.

To assess the effect of DBS on preference for choice, we used the difference between preference for choice ON minus OFF treatment, averaged across the reward probabilities. We then split our cohort using the median, between those with low DBS effect (low ON-OFF difference) and those with a high DBS effect. This categorization was justified by the skewed distribution of the ON-OFF difference, with a peak around zero and more distributed values above it. Given the substantial spatial segregation of the cortical connectivity into the STN, we assumed that strategical localization of the VATs and their related connectivity could drive this effect. We treated patients in the low ON-OFF free choice difference group as the reference group. The patient data were then included in the design matrix for PALM analysis. The contrast matrix provided two series of statistical maps: one which contains voxels with positive t-values associated with higher connectivity in the high ON-OFF group and the other associated to the opposite, i.e., positive t-values were associated with higher connectivity in the low ON-OFF group.

## Results

### Demographic data

We studied choice preference in two groups of PD patients, one group with STN-DBS electrodes tested OFF and ON DBS ( $n = 20$ ), and a second group of patients tested OFF and ON acute oral levodopa treatment ( $n = 25$ ). Patient demographics are listed in Table 1. LEDD was significantly lower for DBS patients since DBS treatment typically reduces pharmacological medication. UPDRS-III ON in the STN-DBS group was significantly higher than in the DOPA group. We did not find any other significant differences between patient groups, although DOPA patients were on average older, displayed less severe motor symptoms (UPDRS-III OFF), and had a shorter disease duration.

	DOPA		STN-DBS		W	p-value
	mean	(SD)	mean	(SD)		
Age (years)	56.8	(8.6)	54.2	(7.8)	303.5	0.226
MoCA (/30)	28.2	(1.4)	27.5	(1.5)	324	0.087
Duration (years)	10.4	(4.3)	12.3	(3.3)	180.5	0.113
UPDRS III (OFF)	43.5	(14.5)	47.7	(16.3)	209	0.355
UPDRS III (ON)	10	(7.3)	16.6	(9.9)	139.5	0.012
LEDD (mg)	1355.2	(525)	404.6	(224.3)	494	<0.0001

**Table 1.** Demographic data for the two groups of patients. Mean and standard deviation (SD) are provided for each variable. W for two-sample Wilcoxon's test.

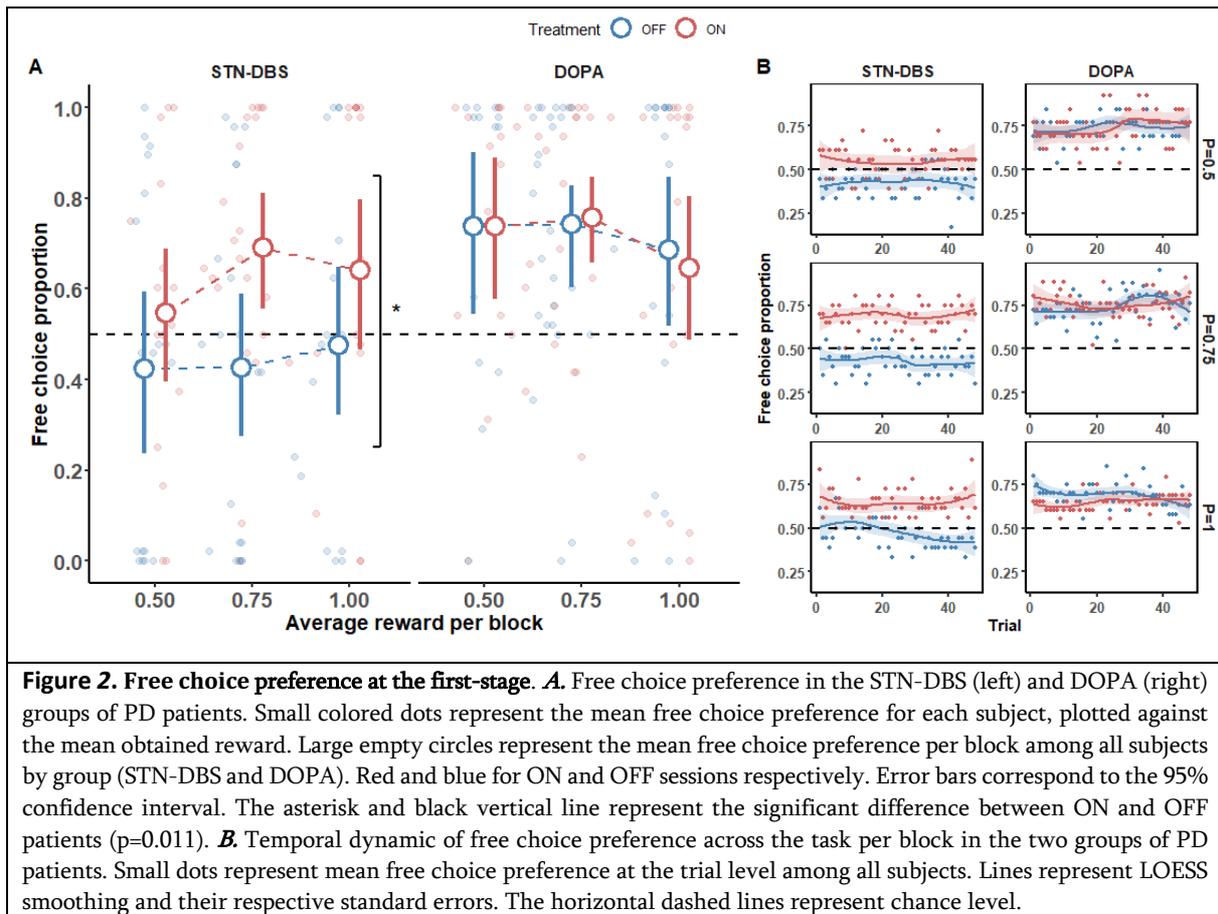
We asked these patients to perform series of trials in a two-stage behavioral task. They had to take a first decision between the opportunity to choose from two terminal actions (*free*) or to perform a mandatory computer-selected terminal action (*forced*), which they then actualized in the second stage (Figure 1).

### First-stage preference for free choice

We examined first-stage choice preference by looking at how often patients selected the *free* option in the first stage of the task (Figure 2A). We used a generalized linear mixed

model (GLMM) to test for an interaction between patient group (STN-DBS or DOPA), treatment (ON or OFF), and reward probability (calculated using obtained rewards over each block). We found a significant two-way interaction (group x treatment,  $z=-2.12$ ,  $p=0.034$ ), while the three-way interaction (group x treatment x reward probability) was not significant ( $z=-0.57$ ,  $p=0.570$ ). The main effect for reward probability was not significant ( $z=1.95$ ,  $p=0.051$ ). Post-hoc tests revealed that STN-DBS patients preferred the *free* option more when ON-DBS compared to OFF-DBS, (63% vs. 44%,  $z=3.07$ ,  $p=0.011$ ). Indeed, the average choice preference OFF-DBS was not significantly different from indifference (44% vs. 50%,  $z=-0.95$ ,  $p=0.339$ ). By contrast, the free choice preference in the DOPA patient group was significantly greater than 50% for all reward probabilities ( $z=3.736$ ,  $p=0.004$ ), and acute levodopa challenge did not affect this preference (72 % OFF vs. 71% ON,  $z=0.227$ ,  $p=0.996$ ). Comparison between patient groups revealed that free choice preference was significantly greater OFF-DOPA compared to OFF-DBS ( $z= 2.78$ ,  $p=0.032$ ) and ON-DOPA compared to OFF-DBS ( $z= 3.00$ ,  $p=0.014$ ), but was not significantly different between ON-DOPA compared to ON-DBS ( $z= 0.89$ ,  $p=0.810$ ).

We next investigated the evolution of first choice behavior over the course of the different blocks over which the programmed reward probability was constant. For each block, the average percentage of *free* selections was approximately constant for both the STN-DBS and DOPA groups and did not vary over time with treatment (Figure 2B). Notably, when present, free choice preference was immediately expressed from the beginning of the testing phase (Figure 2B).



## Second-stage target selection during free choice

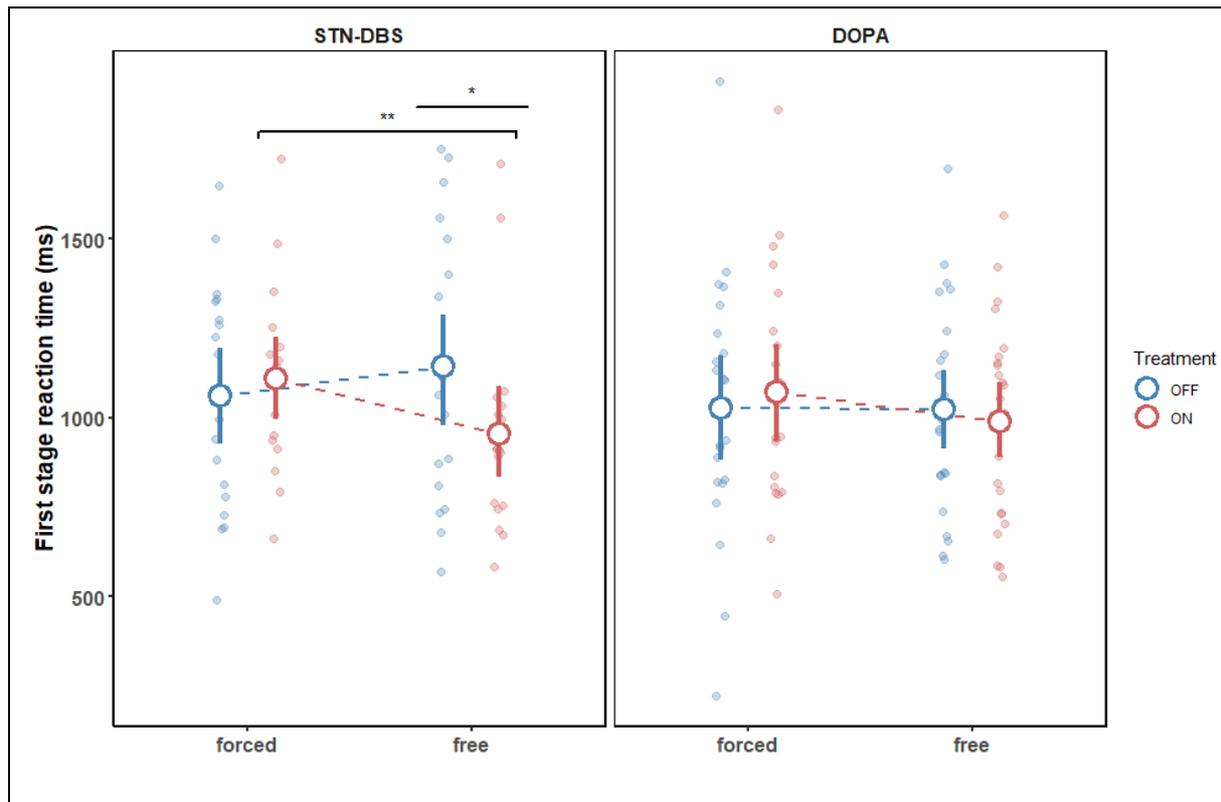
We next investigated behavior at the second-stage of the task to determine whether differences in obtained rewards could explain free choice preference. For example, for two blocks ( $P=0.75$  and  $P=1$ ) one of the second-stage targets was more highly rewarded in the free options, and patients may have exhibited choice avoidance because they failed to learn the reward contingencies and therefore earned fewer rewards compared to the computer-selected action (which always led the higher-rewarded target).

We assessed understanding of reward contingencies at  $P=0.75$  and  $P=1$  by looking at how frequently patients selected the more highly rewarded target in the second-stage during the test phase. We used a GLMM to test for an interaction between patient group (STN-DBS or DOPA), treatment (ON or OFF), and reward probability (obtained reward over block). We found a significant two-way interaction (group  $\times$  treatment,  $z=-3.98$ ,  $p<0.001$ ), and the three-way interaction (group  $\times$  treatment  $\times$  reward probability) was not significant ( $z=1.326$ ,  $p=0.185$ ). Post-hoc analysis revealed that both patient groups preferred the more

highly rewarded target (DOPA=82%,  $z=7.32$ ,  $p<0.0001$ ; DBS=81%,  $z=6.91$ ,  $p<0.00001$ ), indicating that they learned the reward contingencies. Looking at specific blocks by treatment condition, we observed that the DOPA patient group performed worse ON-DOPA compared to OFF-DOPA during the P=0.75 block (74% vs 88%,  $z=-1.992$ ,  $p=0.046$ ). By contrast, patients OFF-DBS performed worse than when they were ON-DBS at P=0.75 only (83% vs 94%,  $z=3.37$ ,  $p=0.001$ ). We also examined how selection of the most rewarded target in free option changed across the blocks where the programmed reward probability was held constant at P=0.75 and P=1. The temporal dynamic of second-stage choices shows that while patients in the STN-DBS group increased selection of the highly rewarded targets over the block, this did not appear to be associated with first-stage choice preference (Supplemental Figure S2).

### First-stage reaction time differences while choosing free and forced options

We examined whether patients were faster or slower for selecting the *free* or *forced* options at the first stage. We used an LMM to test for an interaction between patient group (STN-DBS or DOPA), treatment (ON or OFF), and first-stage choice (*free* or *forced*). We found a significant three-way interaction (group x treatment x first-stage choice,  $t=3.46$ ,  $p=0.0006$ ). Post-hoc tests revealed that DBS patients were faster when selecting *free* compared to *forced* when they were ON-DBS (-230ms,  $z=4.37$ ,  $p=0.0001$ ), and were slower OFF compared to ON when they were selecting *free* (196 ms,  $z=2.70$ ,  $p=0.035$ ; Figure 3). There was no main effect for group ( $t=-0.210$ ,  $p=0.834$ ) or group x treatment ( $t=-0.434$ ,  $p=0.666$ ), indicating that DBS patients were not generally faster at making a first-stage choice compared to DOPA patients.



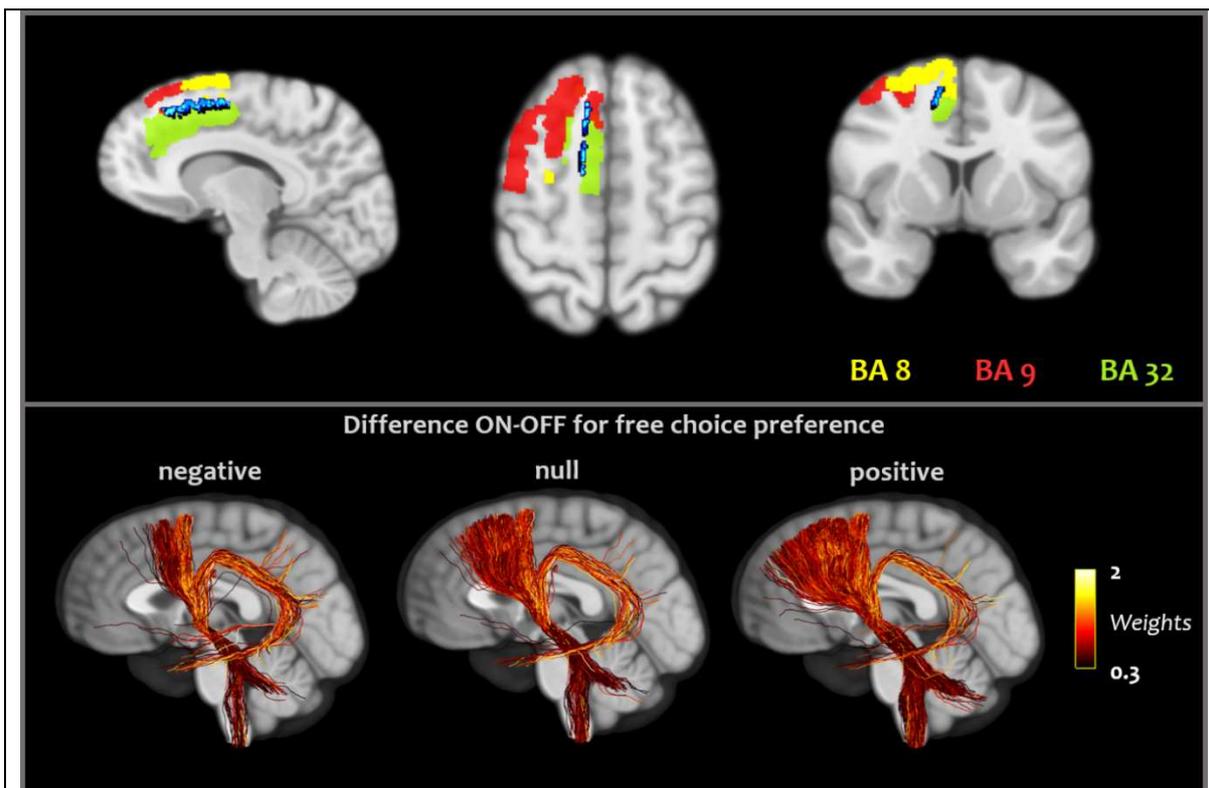
**Figure 3. Reaction times (RT) during first-stage selection.** Red (ON) and blue (OFF) dots represent the mean individual RT of STN-DBS (left) and DOPA (right) PD patients while selecting forced or free option. Large empty circles represent the mean RT. Errorbars correspond to 95% confidence interval. For the STN-DBS group, the bracket with two asterisks on top represents a significant difference between free and forced options for ON patients ( $p=0.0001$ ). The horizontal line with one asterisk represents the significant difference between ON and OFF sessions for STN-DBS patients selecting free option ( $p=0.035$ ).

## Second-stage reaction times during target selection

We used the same LMM to examine second-stage reaction times. We found significant main effects for treatment ( $t=-2.788$ ,  $p=0.007$ ) and reward probability ( $t=-2.993$ ,  $p=0.005$ ), with no significant interactions with patient group ( $p > 0.176$ ). Post-hoc tests revealed that reaction times were faster ON treatment compared to OFF treatment (-92 ms DBS, -31 ms DOPA), and faster when reward probability was higher (-36 ms DBS for  $P=0.50$  vs.  $P=1$ , -41.5 ms DOPA).

## Connectivity analysis

In the DBS patient group, we investigated whether sub-cortico-cortical connectivity could explain the strong increase in choice preference when ON-DBS. We used a normative tractogram to quantify connectivity between cortex and the volume of activated tissue (VAT) by each DBS electrode (see Methods). Voxelwise analysis highlighted a subregion of the right dorsomedial prefrontal cortex and anterior cingulate cortex ( $436 \text{ mm}^3$ ) where connectivity with DBS VATs was significantly higher in patients who selected the *free* option more ON-DBS compared to OFF-DBS (Figure 4, upper panels). Importantly, we did not observe any significant voxels with the inverse contrast searching for voxels with higher connectivity in patients who selected the *forced* option more ON-DBS compared to OFF-DBS. The bottom panels in Figure 4 illustrate connectivity in three patients; one who decreased his *free* preference by 23% (negative), a second who did not change his choice preference (null), and a third who increased his choice preference by 97% (positive).

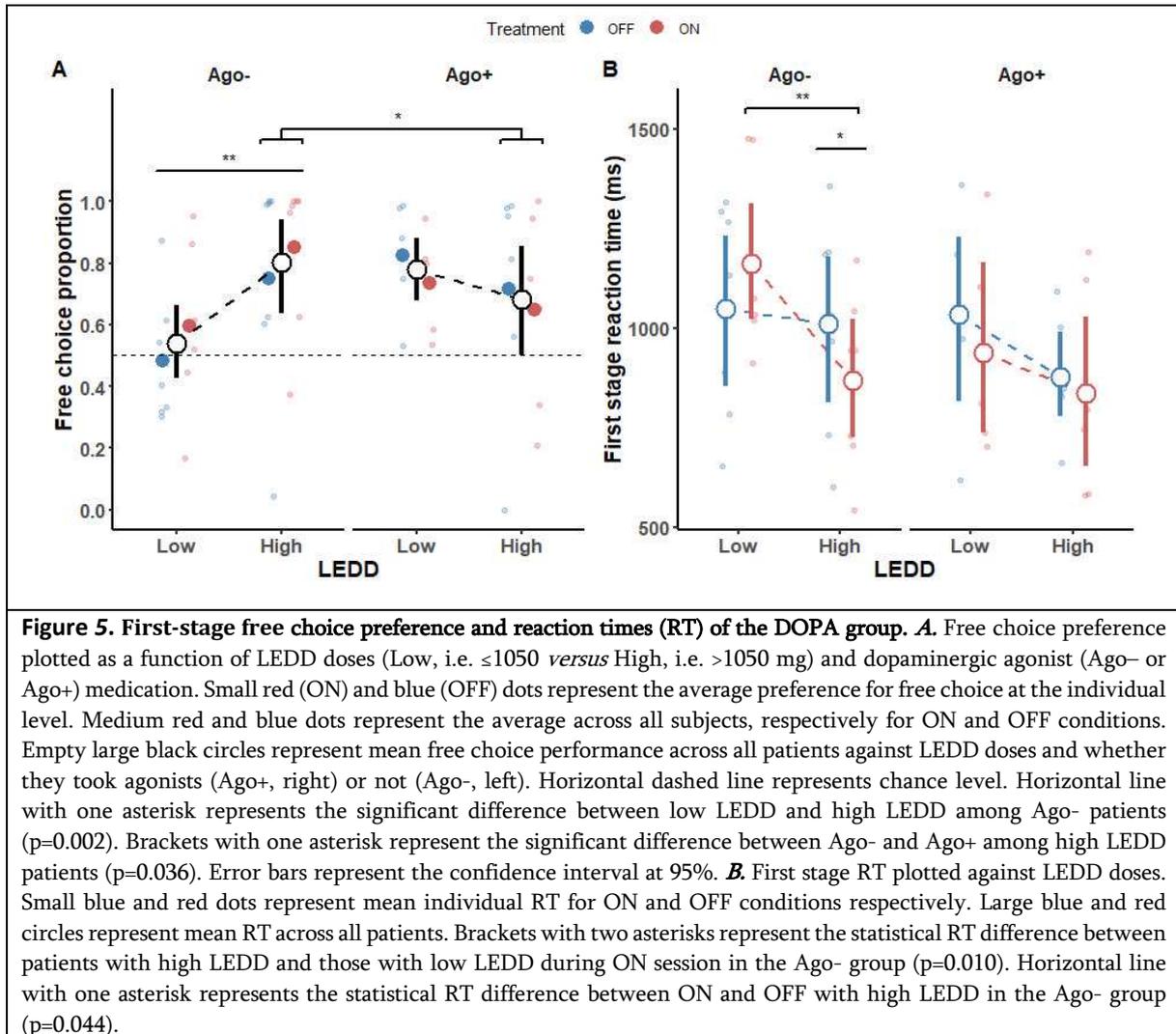


**Figure 4.** Voxelwise analysis between low and high ON-OFF free choice preference in the STN-DBS group. *Top:* the blue region corresponds to the statistical map with a threshold of  $p < 0.05$  (FWER-TFCE) overlaid on a T1 volume of the template. Yellow, red and green regions correspond to the Brodmann area 8, 9 and 32 respectively. *Bottom:* From left to right, streamlines projections of individual patients with low (-0.23), null and positive (0.97) effect of the stimulation on free choice preference. The images are displayed using the radiological orientation.

We also investigated the location of active stimulation contacts in the STN. Coordinates of active electrodes were not different between left and right hemispheres (two-sample:  $W=771$ ,  $p=0.477$ ) and in the dorso-ventral axis ( $W=802$ ,  $p=0.298$ ). However, we found a significant difference in the antero-posterior axis with right electrodes being more anterior than left (left:  $y=10.2$ ; right:  $y=10.8$ ;  $W=431$ ,  $p=0.004$ ); Left and right VAT volumes were not statistically different, although right VATs were bigger on average (left:  $74 \text{ mm}^3$ , right  $95 \text{ mm}^3$ ,  $W=155$ ,  $p=0.231$ ).

### Subgroup analyses of the DOPA patients

We next investigated whether heterogeneity in pharmacological therapy could influence choice preference in the DOPA patient group. We split patients in the DOPA group by dosages of dopamine agonist and other dopamine replacement therapy (see Methods). We then used a GLMM to model choice preference, including factors for levels of long-term dopaminergic medication and their interaction (high or low, dopamine agonist [Ago] or other dopamine replacement [LEDD]). We found a significant interaction of LEDD x Ago ( $z=-2.53$ ,  $p=0.01$ ). Post-hoc tests revealed that for patients without agonists (Ago-), those with high LEDD selected the *free* option significantly more than those with low LEDD (80% vs. 54%,  $z=3.55$   $p=0.002$ ; Figure 5A). Furthermore, among patients with high LEDD, Ago- patients selected the *free* option significantly more than Ago+ patients (80% vs. 69%,  $z=-2.69$ ,  $p=0.036$ ; Figure 5A). Interestingly, acute levodopa treatment did not interact with any terms of the model ( $p>0.216$ ), nor did we find a significant main effect of the reward probability ( $z=-0.594$ ,  $p=0.55$ ).



We also examined first-stage reaction times by DOPA patient subgroups, finding significant two-way interaction (LEDD x treatment,  $t=-2.71$   $p=0.01$ ), and a marginal three-way interaction (Ago x LEDD x treatment,  $t=1.94$ ,  $p=0.059$ ). Post-hoc analyses revealed that among patients without dopamine agonists (Ago-), patients with high LEDD and were faster ON compared to OFF acute levodopa treatment ( $z=2.06$ ,  $p=0.044$ ). Also, ON patients were faster if they received high LEDD compared to low LEDD ( $z=2.572$ ,  $p=0.010$ , Figure 5B). Although patients with high LEDD were faster than those with low LEDD this effect was not significant (876 ms vs 1003 ms,  $t=0.270$ ,  $p=0.789$ )

## Discussion

Intrinsic motivation is a construct used to explain behavior that an agent apparently performs independently of extrinsic outcomes (Ryan & Deci, 2000). Neural mechanisms underlying intrinsic motivation remain unclear, and no direct evidence exists regarding whether such behaviors are impaired in neuropsychiatric disorders like PD. Because the behavioral preference for choice opportunities is well established in humans, we used it as a type of intrinsic reward to investigate these questions. We used a paradigm that allowed subjects to choose between the opportunity to choose again or to make a computer-selected action, and we characterized how intrinsic motivation is impaired in PD.

Free choice preference was strongly modulated by STN-DBS. OFF-DBS PD patients were indifferent between the *free* and *forced* options, while STN-DBS induced an increase in choice preference to a level similar to that previously observed in healthy humans (Leotti & Delgado, 2011; Munuera et al., 2022). Using a normative tractogram, we found that the degree of connectivity between medial frontal cortical (MFC) and the VATs estimated for each patient was associated with the increase in choice preference induced by STN-DBS. The cortical areas involved included Brodmann's areas 8, 9, and 32, which is consistent with the current knowledge about the connectivity of the STN with limbic and associative regions of the cortex (Haynes & Haber, 2013; Temiz et al., 2020). These cortical areas have been classically related to the processing of conflicting information or response inhibition in stimulated PD patients. For example, the decrease in cerebral blood flow in this region was correlated with a higher amount of impulsive responses leading to commission errors (Ballanger et al., 2009; Thobois et al., 2007). Computational modeling of such impairment in PD patients revealed that stimulated patients had a more permissive decision threshold during conflict (Frank et al., 2007; Green et al., 2013), which was also related to the MFC in an EEG study (Cavanagh et al., 2011).

Medial frontal cortex including the dorsal ACC also appears important during value-based decision making (Cai & Padoa-Schioppa, 2012; Shenhav et al., 2016; Vassena et al., 2020). Although these studies focused on extrinsic rewards (e.g., money or fluids), it is possible that MFC is involved in encoding intrinsic value, consistent with work indicating that this

region is important for the sense of agency associated with voluntary decisions (Haggard, 2008; Tanaka et al., 2008). It might be the case that DBS antidromically activates projections from the MFC to the STN, reestablishing the sense of agency associated with voluntary actions leading to positive consequences. An improved sense of agency, and not better performance per se, is supported by the observation that patients correctly learned the reward contingencies both OFF- and ON-DBS.

DBS may have additional effects in addition to modulation of cortico-STN fibers. One possibility is modulation of neurons within the STN itself. Recordings in the STN show subpopulations of neurons responding to distinct rewards, encoding their aversive or appetitive character as well as motivational information (Breyse et al., 2015; Espinosa-Parrilla et al., 2013; Nougaret et al., 2022). Similar results have been obtained in PD patients during neuronal recordings and analysis of local field potentials showing that STN activity is correlated with motivation to obtain a monetary reward as well as the subjective value of rewarded stimuli (Justin Rossi et al., 2017; Zénon et al., 2016). It has also been shown that STN-DBS promotes energization of reward-based actions (Atkinson-Clement et al., 2019; Zénon et al., 2016). This raises the possibility that a pro-motivational effect could come from activation of widespread dopaminergic axons projecting to the STN (Augood et al., 2000; Francois et al., 1985; McIntyre et al., 2004), consistent with well-established role of the dopaminergic system in extrinsic reward processing (Schultz, 2000; Schultz et al., 1997) as well as the intrinsically rewarding property of information (Bromberg-Martin & Hikosaka, 2009). Thus, STN-DBS in PD could reestablish free choice preference by activating the dopaminergic terminals found in this structure. This motivational effect is consistent with STN stimulation leading to faster reaction times specifically when patients selected the *free* option. A role for dopamine is further suggested by the observation that DBS patients OFF stimulation were indifferent between the *free* and *forced* options. This could be due to the substantial reduction of dopaminergic medication in STN-DBS PD patients, resulting in these patients being in a hypodopaminergic state when OFF stimulation.

To further investigate the role of dopamine in intrinsic motivation, we asked whether dopamine replacement therapy modulates choice preference in a second group of PD patients. We found that on average patients in the DOPA group preferred choice at a level similar to healthy humans (Leotti & Delgado, 2011; Munuera et al., 2022), independent of acute levodopa treatment. However, it is known that chronic medication type can influence behavioral (dys)functions in distinct manners (Voon et al., 2017; Weintraub & Mamikonyan, 2019). For example, both dopaminergic agonists and high doses of levodopa could lead to different forms of impulse control behaviors (Corvol et al., 2018; Evans et al., 2009, 2010; Weintraub et al., 2010), which may be associated with upregulation of dopaminergic receptors in the output structure of the midbrain dopaminergic area (Anderson & Nutt, 2011; Voon et al., 2017). Therefore, we asked how different chronic pharmacological therapies affected choice preference. We found that patients with the lowest amount of dopaminergic treatment (low LEDD and Ago-) were indifferent between the *free* and *forced* options. By contrast, patients with increased dopaminergic treatment, be it through levodopa or dopamine agonists, preferred free choice at a level similar to healthy humans. These findings in the DOPA group rejoin those of the STN-DBS group since in both groups a lower level of chronic dopaminergic treatment yielded indifference between free and forced choice. This suggests that dopaminergic depletion is involved in motivation for intrinsic reward although acute dopaminergic treatment was not sufficient to reestablish this preference. As mentioned above this could be a consequence of an absence of dopaminergic upregulation of downstream structure. Additionally, Ago- low LEDD patients were the subjects with the slowest first-stage reaction time compared to the other DOPA subgroups which also endorse a lack of vigor associated to an impairment of motivational system.

The sustained choice preference in DOPA group patients is surprising since they did not always select the most rewarded target in the second-stage (although they were well above chance level). This is economically irrational since it decreases the expected value of the *free* option since the computer always selected the most rewarded target in following selection of the *forced* option. This is strictly true during the deterministic block (P=1)

where the forced second-stage target was always rewarded, and the clear choice preference in this block rules out the possibility that patients incorrectly undervalued the *forced* option.

Altogether our results support the idea that a hypodopaminergic state can reduce intrinsic motivation. This lack of motivation can however be reestablished with different therapeutic strategies such as the stimulation of specific sub-cortical-prefrontal network or by high doses of levodopa or DA agonists. Also, our results suggest that, in this experimental context, value of intrinsic reward was not as high as the value of extrinsic reward. Indeed second-stage selection of the most rewarded target was quite similar ON and OFF treatments whereas first-stage preference varied. This might be the case because the intensity of extrinsic rewards was sufficiently high to be efficiently processed either by the remaining dopaminergic system or compensated by other mechanisms.

More generally our results also suggest that depletion of DA and associated impairment of the sub-cortico-cortical system decrease intrinsic motivation for choice opportunities. Further investigation should address whether some general motivational deficits and depression-like symptoms in subjects could be better explained by a lack of intrinsic motivation in addition to impaired motivation for extrinsic rewards.

## References

- Ajzen, I. (2002). Perceived Behavioral Control, Self-Efficacy, Locus of Control, and the Theory of Planned Behavior <sup>1</sup>. *Journal of Applied Social Psychology*, *32*(4), 665-683. <https://doi.org/10.1111/j.1559-1816.2002.tb00236.x>
- Anderson, E., & Nutt, J. (2011). The long-duration response to levodopa : Phenomenology, potential mechanisms and clinical implications. *Parkinsonism & Related Disorders*, *17*(8), 587-592. <https://doi.org/10.1016/j.parkreldis.2011.03.014>
- Atkinson-Clement, C., Cavazzini, É., Zénon, A., Witjas, T., Fluchère, F., Azulay, J.-P., Baunez, C., & Eusebio, A. (2019). Effects of subthalamic nucleus stimulation and levodopa on decision-making in Parkinson's disease. *Movement Disorders: Official Journal of the Movement Disorder Society*, *34*(3), 377-385. <https://doi.org/10.1002/mds.27625>
- Augood, S. J., Hollingsworth, Z. R., Standaert, D. G., Emson, P. C., & Penney Jr., J. B. (2000). Localization of dopaminergic markers in the human subthalamic nucleus. *Journal of Comparative Neurology*, *421*(2), 247-255. [https://doi.org/10.1002/\(SICI\)1096-9861\(20000529\)421:2<247::AID-CNE9>3.0.CO;2-F](https://doi.org/10.1002/(SICI)1096-9861(20000529)421:2<247::AID-CNE9>3.0.CO;2-F)
- Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric Diffeomorphic Image Registration with Cross-Correlation : Evaluating Automated Labeling of Elderly and Neurodegenerative Brain. *Medical image analysis*, *12*(1), 26-41. <https://doi.org/10.1016/j.media.2007.06.004>
- Ballanger, B., van Eimeren, T., Moro, E., Lozano, A. M., Hamani, C., Boulinguez, P., Pellecchia, G., Houle, S., Poon, Y. Y., Lang, A. E., & Strafella, A. P. (2009). Stimulation of the subthalamic nucleus and impulsivity : Release your horses. *Annals of Neurology*, *66*(6), 817-824. <https://doi.org/10.1002/ana.21795>
- Bandura, A., Freeman, W. H., & Lightsey, R. (1999). Self-Efficacy : The Exercise of Control. *Journal of Cognitive Psychotherapy*, *13*(2), 158-166. <https://doi.org/10.1891/0889-8391.13.2.158>
- Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models : A practical guide for ecology and evolution. *Trends in Ecology & Evolution*, *24*(3), 127-135. <https://doi.org/10.1016/j.tree.2008.10.008>
- Breyse, E., Pelloux, Y., & Baunez, C. (2015). The Good and Bad Differentially Encoded within the Subthalamic Nucleus in Rats. *eNeuro*, *2*(5), ENEURO.0014-15.2015. <https://doi.org/10.1523/ENEURO.0014-15.2015>
- Bromberg-Martin, E. S., & Hikosaka, O. (2009). Midbrain Dopamine Neurons Signal Preference for Advance Information about Upcoming Rewards. *Neuron*, *63*(1), 119-126. <https://doi.org/10.1016/j.neuron.2009.06.009>
- Cai, X., & Padoa-Schioppa, C. (2012). Neuronal Encoding of Subjective Value in Dorsal and Ventral Anterior Cingulate Cortex. *Journal of Neuroscience*, *32*(11), 3791-3808. <https://doi.org/10.1523/JNEUROSCI.3864-11.2012>
- Catania, A. C., & Sagvolden, T. (1980). Preference for Free Choice Over Forced Choice in Pigeons. *Journal of the Experimental Analysis of Behavior*, *34*(1), 77-86. <https://doi.org/10.1901/jeab.1980.34-77>

- Cavanagh, J. F., Wiecki, T. V., Cohen, M. X., Figueroa, C. M., Samanta, J., Sherman, S. J., & Frank, M. J. (2011). Subthalamic nucleus stimulation reverses mediofrontal influence over decision threshold. *Nature Neuroscience*, *14*(11), 1462-1467. <https://doi.org/10.1038/nn.2925>
- Cockburn, J., Collins, A. G. E., & Frank, M. J. (2014). A Reinforcement Learning Mechanism Responsible for the Valuation of Free Choice. *Neuron*, *83*(3), 551-557. <https://doi.org/10.1016/j.neuron.2014.06.035>
- Corvol, J.-C., Artaud, F., Cormier-Dequaire, F., Rascol, O., Durif, F., Derkinderen, P., Marques, A.-R., Bourdain, F., Brandel, J.-P., Pico, F., Lacomblez, L., Bonnet, C., Brefel-Courbon, C., Ory-Magne, F., Grabli, D., Klebe, S., Mangone, G., You, H., Mesnage, V., ... DIGPD Study Group. (2018). Longitudinal analysis of impulse control disorders in Parkinson disease. *Neurology*, *91*(3), e189-e201. <https://doi.org/10.1212/WNL.0000000000005816>
- Damier, P., Hirsch, E. C., Agid, Y., & Graybiel, A. M. (1999). The substantia nigra of the human brain. II. Patterns of loss of dopamine-containing neurons in Parkinson's disease. *Brain: A Journal of Neurology*, *122* ( Pt 8), 1437-1448. <https://doi.org/10.1093/brain/122.8.1437>
- deCharms, R. (1968). *Personal causation : The internal affective determinants of behavior*. New York: Academic Press. <https://doi.org/10.4324/9781315825632>
- Deci, E. L., & Ryan, R. M. (1985). Intrinsic Motivation and Self-Determination in Human Behavior. In *Intrinsic Motivation and Self-Determination in Human Behavior*. <https://doi.org/10.1007/978-1-4899-2271-7>
- Dembek, T. A., Hellerbach, A., Jergas, H., Eichner, M., Wirths, J., Dafsari, H. S., Barbe, M. T., Hunsche, S., Visser-Vandewalle, V., & Treuer, H. (2021). DiODe v2 : Unambiguous and Fully-Automated Detection of Directional DBS Lead Orientation. *Brain Sciences*, *11*(11), 1450. <https://doi.org/10.3390/brainsci11111450>
- Espinosa-Parrilla, J.-F., Baunez, C., & Apicella, P. (2013). Linking reward processing to behavioral output : Motor and motivational integration in the primate subthalamic nucleus. *Frontiers in Computational Neuroscience*, *7*, 175. <https://doi.org/10.3389/fncom.2013.00175>
- Evans, A. H., Lawrence, A. D., Cresswell, S. A., Katzenschlager, R., & Lees, A. J. (2010). Compulsive use of dopaminergic drug therapy in Parkinson's disease : Reward and anti-reward. *Movement Disorders*, *25*(7), 867-876. <https://doi.org/10.1002/mds.22898>
- Evans, A. H., Strafella, A. P., Weintraub, D., & Stacy, M. (2009). Impulsive and compulsive behaviors in Parkinson's disease. *Movement Disorders: Official Journal of the Movement Disorder Society*, *24*(11), 1561-1570. <https://doi.org/10.1002/mds.22505>
- Fearnley, J. M., & Lees, A. J. (1991). Ageing and Parkinson's disease : Substantia nigra regional selectivity. *Brain: A Journal of Neurology*, *114* ( Pt 5), 2283-2301. <https://doi.org/10.1093/brain/114.5.2283>
- Fonov, V., Evans, A. C., Botteron, K., Almli, C. R., McKinstry, R. C., & Collins, D. L. (2011). Unbiased Average Age-Appropriate Atlases for Pediatric Studies. *NeuroImage*, *54*(1), 313-327. <https://doi.org/10.1016/j.neuroimage.2010.07.033>

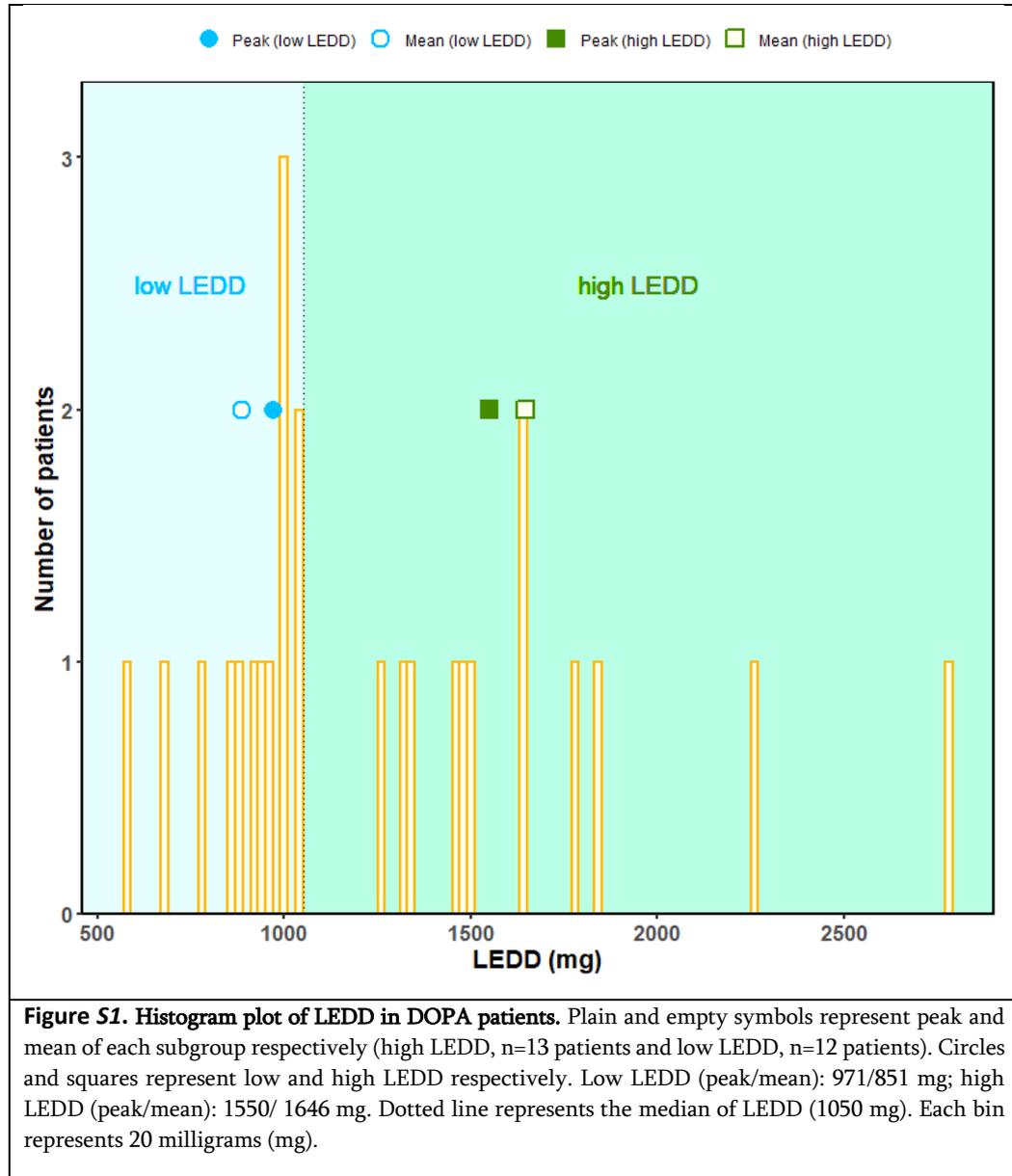
- Francois, C., Percheron, G., Yelnik, J., & Heyner, S. (1985). A histological atlas of the macaque (*Macaca mulatta*) substantia nigra in ventricular coordinates. *Brain Research Bulletin*, *14*(4), 349-367. [https://doi.org/10.1016/0361-9230\(85\)90196-0](https://doi.org/10.1016/0361-9230(85)90196-0)
- Frank, M. J., Samanta, J., Moustafa, A. A., & Sherman, S. J. (2007). Hold Your Horses : Impulsivity, Deep Brain Stimulation, and Medication in Parkinsonism. *Science*, *318*(5854), 1309-1312. <https://doi.org/10.1126/science.1146157>
- Frank, M. J., Seeberger, L. C., & O'Reilly, R. C. (2004). By Carrot or by Stick : Cognitive Reinforcement Learning in Parkinsonism. *Science*, *306*(5703), 1940-1943. JSTOR. <https://doi.org/10.1126/science.1102941>
- Green, N., Bogacz, R., Huebl, J., Beyer, A.-K., Kühn, A. A., & Heekeren, H. R. (2013). Reduction of Influence of Task Difficulty on Perceptual Decision Making by STN Deep Brain Stimulation. *Current Biology*, *23*(17), 1681-1684. <https://doi.org/10.1016/j.cub.2013.07.001>
- Haggard, P. (2008). Human volition : Towards a neuroscience of will. *Nature Reviews Neuroscience*, *9*(12), 934-946. <https://doi.org/10.1038/nrn2497>
- Harrison, X. A., Donaldson, L., Correa-Cano, M. E., Evans, J., Fisher, D. N., Goodwin, C. E. D., Robinson, B. S., Hodgson, D. J., & Inger, R. (2018). A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ*, *6*, e4794. <https://doi.org/10.7717/peerj.4794>
- Haynes, W. I. A., & Haber, S. N. (2013). The organization of prefrontal-subthalamic inputs in primates provides an anatomical substrate for both functional specificity and integration : Implications for Basal Ganglia models and deep brain stimulation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(11), 4804-4814. <https://doi.org/10.1523/jneurosci.4674-12.2013>
- Horn, A., & Kühn, A. A. (2015). Lead-DBS : A toolbox for deep brain stimulation electrode localizations and visualizations. *NeuroImage*, *107*, 127-135. <https://doi.org/10.1016/j.neuroimage.2014.12.002>
- Horn, A., Kühn, A. A., Merkl, A., Shih, L., Alterman, R., & Fox, M. (2017). Probabilistic conversion of neurosurgical DBS electrode coordinates into MNI space. *NeuroImage*, *150*, 395-404. <https://doi.org/10.1016/j.neuroimage.2017.02.004>
- Horn, A., Li, N., Dembek, T. A., Kappel, A., Boulay, C., Ewert, S., Tietze, A., Husch, A., Perera, T., Neumann, W.-J., Reisert, M., Si, H., Oostenveld, R., Rorden, C., Yeh, F.-C., Fang, Q., Herrington, T. M., Vorwerk, J., & Kühn, A. A. (2019). Lead-DBS v2 : Towards a comprehensive pipeline for deep brain stimulation imaging. *NeuroImage*, *184*, 293-316. <https://doi.org/10.1016/j.neuroimage.2018.08.068>
- Horn, A., Reich, M., Vorwerk, J., Li, N., Wenzel, G., Fang, Q., Schmitz-Hübsch, T., Nickl, R., Kupsch, A., & Volkmann, J. (2017). Connectivity predicts deep brain stimulation outcome in Parkinson disease. *Annals of neurology*, *82*(1), 67-78.
- Hughes, A. J., Daniel, S. E., Kilford, L., & Lees, A. J. (1992). Accuracy of clinical diagnosis of idiopathic Parkinson's disease : A clinico-pathological study of 100 cases. *Journal of Neurology, Neurosurgery, and Psychiatry*, *55*(3), 181-184. <https://doi.org/10.1136/jnnp.55.3.181>
- Hull, C. L. (1943). *Principles of behavior*. Appleton-Century-Crofts.

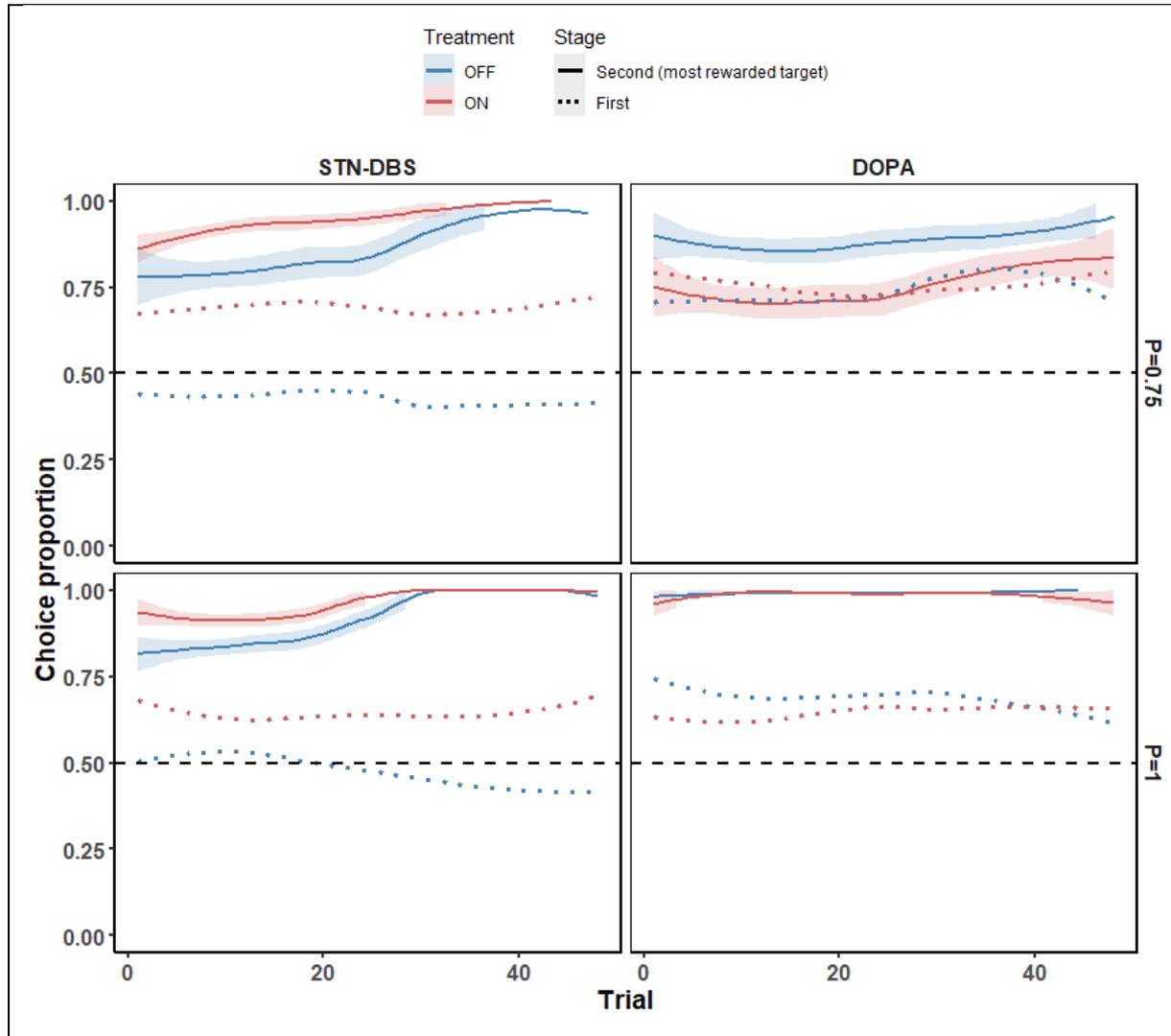
- Husch, A., V. Petersen, M., Gemmar, P., Goncalves, J., & Hertel, F. (2018). PaCER - A fully automated method for electrode trajectory and contact reconstruction in deep brain stimulation. *NeuroImage: Clinical*, *17*, 80-89. <https://doi.org/10.1016/j.nicl.2017.10.004>
- Justin Rossi, P., Peden, C., Castellanos, O., Foote, K. D., Gunduz, A., & Okun, M. S. (2017). The human subthalamic nucleus and globus pallidus internus differentially encode reward during action control. *Human Brain Mapping*, *38*(4), 1952-1964. PubMed. <https://doi.org/10.1002/hbm.23496>
- Lachaud, C. M., & Renaud, O. (2011). A tutorial for analyzing human reaction times : How to filter data, manage missing values, and choose a statistical model. *Applied Psycholinguistics*, *32*(2), 389-416. <https://doi.org/10.1017/S0142716410000457>
- Leotti, L. A., & Delgado, M. R. (2011). The Inherent Reward of Choice. *Psychological Science*, *22*(10), 1310-1318. <https://doi.org/10.1177/0956797611417005>
- Leotti, L. A., & Delgado, M. R. (2014). The Value of Exercising Control Over Monetary Gains and Losses. *Psychological Science*, *25*(2), 596-604. <https://doi.org/10.1177/0956797613514589>
- Limousin, P., & Foltynie, T. (2019). Long-term outcomes of deep brain stimulation in Parkinson disease. *Nature Reviews Neurology*, *15*(4), 234-242. <https://doi.org/10.1038/s41582-019-0145-9>
- Maier, S. F., & Seligman, M. E. P. (1976). Learned Helplessness : Theory and Evidence. *Journal of Experimental Psychology: General*, *105*(1), 3-46.
- Matsumoto, H., Tian, J., Uchida, N., & Watabe-Uchida, M. (2016). Midbrain dopamine neurons signal aversion in a reward-context-dependent manner. *ELife*, *5*, e17328. PubMed. <https://doi.org/10.7554/eLife.17328>
- Matsumoto, M., & Hikosaka, O. (2009). Two types of dopamine neuron distinctly convey positive and negative motivational signals. *Nature*, *459*(7248), 837-841. <https://doi.org/10.1038/nature08028>
- McIntyre, C. C., Grill, W. M., Sherman, D. L., & Thakor, N. V. (2004). Cellular effects of deep brain stimulation : Model-based analysis of activation and inhibition. *Journal of neurophysiology*, *91*(4), 1457-1469.
- Munuera, J., Agost, M. R., Bendetowicz, D., Kerebel, A., Chambon, V., & Lau, B. (2022). *Choice seeking is motivated by the intrinsic need for personal control* (p. 2022.09.20.508669). bioRxiv. <https://doi.org/10.1101/2022.09.20.508669>
- Nougaret, S., Baunez, C., & Ravel, S. (2022). Neurons in the Monkey's Subthalamic Nucleus Differentially Encode Motivation and Effort. *Journal of Neuroscience*, *42*(12), 2539-2551. <https://doi.org/10.1523/JNEUROSCI.0281-21.2021>
- Palminteri, S., Lebreton, M., Worbe, Y., Grabli, D., Hartmann, A., & Pessiglione, M. (2009). Pharmacological modulation of subliminal learning in Parkinson's and Tourette's syndromes. *Proceedings of the National Academy of Sciences*, *106*(45), 19179-19184. <https://doi.org/10.1073/pnas.0904035106>
- Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., & Nichols, T. E. (2011). *Statistical Parametric Mapping : The Analysis of Functional Brain Images*. Elsevier.
- Rutledge, R. B., Lazzaro, S. C., Lau, B., Myers, C. E., Gluck, M. A., & Glimcher, P. W. (2009). Dopaminergic Drugs Modulate Learning Rates and Perseveration in Parkinson's

- Patients in a Dynamic Foraging Task. *Journal of Neuroscience*, 29(48), 15104-15114. <https://doi.org/10.1523/jneurosci.3524-09.2009>
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68-78. <https://doi.org/10.1037/0003-066X.55.1.68>
- Schönecker, T., Kupsch, A., Kühn, A., Schneider, G.-H., & Hoffmann, K.-T. (2009). Automated Optimization of Subcortical Cerebral MR Imaging-Atlas Coregistration for Improved Postoperative Electrode Localization in Deep Brain Stimulation. *AJNR. American journal of neuroradiology*, 30, 1914-1921. <https://doi.org/10.3174/ajnr.A1741>
- Schultz, W. (1998). Predictive Reward Signal of Dopamine Neurons. *Journal of Neurophysiology*, 80(1), 1-27. <https://doi.org/10.1152/jn.1998.80.1.1>
- Schultz, W. (2000). Multiple reward signals in the brain. *Nature Reviews. Neuroscience*, 1(3), 199-207. <https://doi.org/10.1038/35044563>
- Schultz, W., Dayan, P., & Montague, P. R. (1997). A Neural Substrate of Prediction and Reward. *Science*, 275(5306), 1593-1599. <https://doi.org/10.1126/science.275.5306.1593>
- Shenhav, A., Cohen, J. D., & Botvinick, M. M. (2016). Dorsal anterior cingulate cortex and the value of control. *Nature Neuroscience*, 19(10), 1286-1291. <https://doi.org/10.1038/nn.4384>
- Smith, R. E., Tournier, J.-D., Calamante, F., & Connelly, A. (2015). SIFT2: Enabling dense quantitative assessment of brain white matter connectivity using streamlines tractography. *NeuroImage*, 119, 338-351. <https://doi.org/10.1016/j.neuroimage.2015.06.092>
- Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, 44(1), 83-98. <https://doi.org/10.1016/j.neuroimage.2008.03.061>
- Sutton, R. S., & Barto, A. (2018). *Reinforcement learning: An introduction* (Second edition). The MIT Press.
- Suzuki, S. (1999). Selection of Forced- and Free-Choice by Monkeys (Macaca Fascicularis). *Perceptual and Motor Skills*, 88(1), 242-250. <https://doi.org/10.2466/pms.1999.88.1.242>
- Tanaka, S. C., Balleine, B. W., & O'Doherty, J. P. (2008). Calculating Consequences: Brain Systems That Encode the Causal Effects of Actions. *The Journal of Neuroscience*, 28(26), 6750-6755. <https://doi.org/10.1523/JNEUROSCI.1808-08.2008>
- Temiz, G., Santin, M.-N., Olivier, C., Collomb-Clerc, A., Fernandez-Vidal, S., Hainque, E., Bardinet, E., Lau, B., François, C., Karachi, C., & Welter, M.-L. (2022). Freezing of gait depends on the cortico-subthalamic networks recruitment after STN-DBS in PD patients. *under revision*.
- Temiz, G., Sébille, S. B., Francois, C., Bardinet, E., & Karachi, C. (2020). The anatomofunctional organization of the hyperdirect cortical pathway to the subthalamic area using in vivo structural connectivity imaging in humans. *Brain Structure and Function*, 225(2), 551-565. <https://doi.org/10.1007/s00429-019-02012-6>

- Thobois, S., Hotton, G. R., Pinto, S., Wilkinson, L., Limousin-Dowsey, P., Brooks, D. J., & Jahanshahi, M. (2007). STN Stimulation Alters Pallidal—Frontal Coupling during Response Selection under Competition. *Journal of Cerebral Blood Flow & Metabolism*, *27*(6), 1173-1184. <https://doi.org/10.1038/sj.jcbfm.9600425>
- Tricomi, E. M., Delgado, M. R., & Fiez, J. A. (2004). Modulation of Caudate Activity by Action Contingency. *Neuron*, *41*(2), 281-292. [https://doi.org/10.1016/s0896-6273\(03\)00848-1](https://doi.org/10.1016/s0896-6273(03)00848-1)
- Vassena, E., Deraeve, J., & Alexander, W. H. (2020). Surprise, value and control in anterior cingulate cortex during speeded decision-making. *Nature Human Behaviour*, *4*(4), 412-422. <https://doi.org/10.1038/s41562-019-0801-5>
- Voon, V., Kubu, C., Krack, P., Houeto, J.-L., & Tröster, A. I. (2006). Deep brain stimulation : Neuropsychological and neuropsychiatric issues. *Movement Disorders*, *21*(S14), S305-S327. <https://doi.org/10.1002/mds.20963>
- Voon, V., Napier, T. C., Frank, M. J., Sgambato-Faure, V., Grace, A. A., Rodriguez-Oroz, M., Obeso, J., Bezard, E., & Fernagut, P.-O. (2017). Impulse control disorders and levodopa-induced dyskinesias in Parkinson's disease: An update. *The Lancet. Neurology*, *16*(3), 238-250. [https://doi.org/10.1016/s1474-4422\(17\)30004-2](https://doi.org/10.1016/s1474-4422(17)30004-2)
- Voss, S. C., & Homzie, M. J. (1970). Choice as a Value. *Psychological Reports*, *26*(3), 912-914. <https://doi.org/10.2466/pr0.1970.26.3.912>
- Wang, Q., Akram, H., Muthuraman, M., Gonzalez-Escamilla, G., Sheth, S. A., Oxenford, S., Yeh, F.-C., Groppa, S., Vanegas-Aroyave, N., Zrinzo, L., Li, N., Kühn, A., & Horn, A. (2021). Normative vs. Patient-specific brain connectivity in deep brain stimulation. *NeuroImage*, *224*, 117307. <https://doi.org/10.1016/j.neuroimage.2020.117307>
- Weintraub, D., Koester, J., Potenza, M. N., Siderowf, A. D., Stacy, M., Voon, V., Whetteckey, J., Wunderlich, G. R., & Lang, A. E. (2010). Impulse control disorders in Parkinson disease: A cross-sectional study of 3090 patients. *Archives of Neurology*, *67*(5), 589-595. <https://doi.org/10.1001/archneurol.2010.65>
- Weintraub, D., & Mamikonyan, E. (2019). The Neuropsychiatry of Parkinson Disease : A Perfect Storm. *The American Journal of Geriatric Psychiatry*, *27*(9), 998-1018. <https://doi.org/10.1016/j.jagp.2019.03.002>
- White, R. W. (1959). Motivation reconsidered : The concept of competence. *Psychological Review*, *66*(5), 297-333. <https://doi.org/10.1037/H0040934>
- Zénon, A., Duclos, Y., Carron, R., Witjas, T., Baunez, C., Régis, J., Azulay, J.-P., Brown, P., & Eusebio, A. (2016). The human subthalamic nucleus encodes the subjective value of reward and the cost of effort during decision-making. *Brain: A Journal of Neurology*, *139*(Pt 6), 1830-1843. <https://doi.org/10.1093/brain/aww075>

## Supplementary information





**Figure S2. Temporal dynamic of the selection of the most rewarded target at the second stage across the task per block.**

Solid lines represent LOESS smoothing of subjects' choices of the most rewarded second stage targets across blocks and their respective standard error. Top and bottom panels for  $P=0.75$  and  $P=1$  blocks respectively. Blue and red for OFF and ON therapy sessions respectively. Dashed line represents the chance level (0.5). Dotted lines represent free choice proportion at the first stage as already shown in figure 2B.

ID	Manufacturer	Left hemisphere				Right hemisphere			
		active contact(s)	Power (mA or V)	pulse duration ( $\mu$ s)	Frequency (Hz)	active contact(s)	Power (mA or V)	pulse duration ( $\mu$ s)	Frequency (Hz)
1	Boston	5(75%)-7(25%)	2,70	60	130	4	4,00	60	130
2	Boston	5	3,50	100	130	6	2,90	100	130
3	Medtronic	2	2,60	60	90	2	3,10	60	90
4	Medtronic	2	2,00	60	130	1	2,20	60	130
5	Medtronic	1	3,40	60	130	2	3,10	60	130
6	Boston	5 6 7	2,90	60	130	5	2,80	60	130
7	Medtronic	1	1,80	60	130	2	3,00	60	130
8	Boston	5 6 7	3,80	60	130	2 3 4	3,40	60	130
9	Boston	2 3 4	3,60	60	130	2 3 4	2,60	60	130
10	Boston	2 3 4	2,80	60	130	2 3 4	3,80	60	130
11	Boston	5(75%)-6(25%)	3,60	60	130	5(2%)-6(8%)-3(68%)-2(22%)	5,00	90	130
12	Boston	5 6 7	1,40	60	130	2 3 4	2,10	70	130
13	Boston	5 6 7	2,60	60	130	2 3 4	2,40	60	130
14	Boston	3	1,70	60	130	4	2,10	60	112
15	Medtronic	2	2,35	60	130	3	3,35	60	200
16	Medtronic	2	2,70	60	130	1	3,40	60	130
17	Medtronic	2	2,40	60	130	2	3,20	60	130
18	Boston	5 6 7	3,80	60	179	8 (20%) 6(40%) 5(40%)	3,90	60	179
19	Medtronic	2	3,70	60	130	2	2,80	60	130
20	Medtronic	2	2,70	60	205	2	1,85	60	195

**Table S1. Stimulation parameters settings for each patient.** Power of stimulation are given in milliamperes (mA) for Boston Vercise Cartesia and volts (V) for Medtronic 3389 electrodes. Percentages between brackets correspond to the proportion of current delivered for each active contact.

## DISCUSSION



Dans ce travail de thèse et grâce aux résultats obtenus par ces deux études, nous avons pu préciser les déterminants comportementaux permettant aux individus de rechercher les situations avec plus de choix et identifier des structures cérébrales clés impliqués dans cette préférence. Le premier objectif était de confirmer que, si le choix permettait de percevoir plus de contrôle, il devait être intrinsèquement motivant (Ryan & Deci, 2000). Le second était de confirmer que la modulation dopaminergique influençait l'encodage du choix en tant que récompense intrinsèque. Troisièmement, nous avons tenté de déterminer comment le NST, une structure majeure des ganglions de la base, intervenait dans la sélection des décisions favorisant le choix.

### **Le choix permet-il d'augmenter la perception de contrôle ?**

Les travaux antérieurs et les théories sur le libre choix mettent en avant son rôle pour augmenter la perception de contrôle. Ainsi, les joueurs de loterie dans l'expérience de Langer (1975) et les sujets de l'expérience sur le choix de la boîte de nuit de Bown et al. (2003) préfèrent des situations de libre choix alors même que ce dernier est en fait fictif : ils se donnent « l'illusion » d'un choix pour sentir plus de contrôle. Ce scénario a été modélisé dans l'expérience 3 des sujets sains (étude 1). Lorsqu'il existe un degré d'incohérence entre les réponses du sujet et le résultat final, ces derniers tentent de regagner une perception de contrôle en basculant plus fréquemment du côté choix forcé. La notion de perception est essentielle. En effet, dans cette tâche d'incohérence, l'ordinateur sélectionnait à la seconde étape le contraire de ce que le sujet avait choisi. Alors qu'il n'y avait aucune possibilité de restaurer de la cohérence par les sujets eux-mêmes, ces derniers ont tout de même augmenté leur proportion de choix libre. Ce comportement ne visait donc pas à restaurer du contrôle mais à en augmenter la perception par les sujets.

Par sa dimension perceptive, le libre choix pourrait participer à renforcer le sentiment d'agentivité. Cette notion est définie par le sentiment qu'un sujet peut avoir sur ses propres actions en contrôlant son environnement. De la même manière que pour le libre choix, le simple « sentiment » d'agentivité suffit pour qu'un sujet se sente responsable de ses actions,

et non l'action en elle-même. Une autre manière de poser la question, est de préciser à quel moment un sujet se sent responsable de ces actions. Il peut le faire de manière rétrospective, en se rendant compte *a posteriori* des conséquences de ses actions. Ou le sujet peut également se sentir agent de ses actions de manière prospective, c'est-à-dire au moment où la décision est prise de réaliser l'action et avant d'en voir le résultat. Wenke et al. (2010) ont mis en évidence cette dimension prospective au moyen d'une tâche intercalant un stimulus subliminal au début de chaque essai. Des flèches orientant à gauche ou à droite étaient exposées au sujet qui devaient indiquer leur sens. Avant que cette flèche n'apparaisse, une autre flèche allant soit dans le même sens (essais cohérent), soit dans le sens opposé (essai incohérent) était affichée quelques millisecondes avant le choix. Il était également demandé aux sujets, en fonction des essais réalisés, d'évaluer sur une échelle à quel point ils avaient l'impression d'avoir le contrôle de la situation. Il s'est avéré que les essais cohérents amenaient un sentiment de contrôle plus important que les essais incohérents, alors que les deux types d'essais étaient indissociables consciemment. Par analogie avec les procédés mis en place dans nos expériences, puisque les sujets sont sensibles à la perception de contrôle plus qu'au contrôle lui-même, il est possible que les sujets s'engagent dans des situations de choix libre afin d'augmenter leur sentiment d'agentivité.

### **Moyen ou fin : le choix permet-il d'être mieux récompensé ou est-il la récompense ?**

Cette question centrale laisse paraître l'idée que choisir librement ne serait finalement gratifiant que pour obtenir ce qui a été choisi plutôt que l'acte de choisir. Par exemple, préférer des situations de libre choix permet finalement de mieux apprendre de son environnement. Ainsi dans les situations d'incertitude, choisir serait utile en vue de pouvoir explorer et ce comportement devrait disparaître quand l'environnement est suffisamment maîtrisé. C'est d'ailleurs une des limitations du labyrinthe de Voss & Homzie (1970) où les rats, dans un environnement incertain, auraient peut-être préféré le libre choix dans un but exploratoire. Les résultats obtenus dans le groupe de sujets sains (étude 1) et dans le groupe

des patients parkinsoniens sous traitement dopaminergique (DOPA ON, étude 2) vont à l'encontre de cette hypothèse. En effet, avant de réaliser la tâche en deux étapes, nos sujets se sont tous entraînés préalablement pour apprendre les différentes contingences de récompense extrinsèque. Or, dans ce contexte certain où tous les sujets ont appris les différentes contingences à plus de 70%, cette préférence pour le libre choix s'est exprimée de manière nette, toujours au-dessus du seuil d'indifférence. De plus, pour certains sujets chez qui l'apprentissage aurait été sub-optimal ou alors tout simplement par envie d'explorer un peu plus, il aurait été attendu une décroissance progressive de la proportion de choix libre tout au long des essais. Ce résultat aurait alors été le témoin d'une exploration ou d'un apprentissage satisfaisant, ne nécessitant plus de libre choix. Mais cet argument ne peut pas être retenu car, en observant la dynamique du libre choix essai par essai, la préférence pour le libre choix persiste de manière stable et invariable dans le temps.

Nos études redéfinissent également la place du « paradigme du libre choix » qui, malgré son nom ne confère pas une valeur au choix mais plutôt aux options découlant de ce choix, ceci pour réduire une éventuelle dissonance cognitive. Pour rappel, ce paradigme suggère que les conséquences de nos choix sont surévaluées a posteriori lorsque ce dernier est fait librement, c'est-à-dire en induisant un biais de confirmation. C'est ainsi que dans les études de Cockburn et al. (2014) et Chambon et al. (2020), la valeur des options choisies provenant de modèles computationnels étaient plus élevées que celles des options non choisies. A la différence de ces deux dernières études, la structure de notre tâche en deux étapes nous permet de dissocier la décision de faire un choix pour lui-même et l'impact du choix sur les récompenses terminales. Nous avons ainsi montré que cette préférence du choix ne peut pas être entièrement expliquée par ce biais de confirmation. Premièrement, les proportions de choix libres ne variaient pas avec les contingences de récompense extrinsèque, ce qui permet d'écarter l'influence de la récompense finale sur le choix. Deuxièmement, cette préférence du choix libre persiste alors que les récompenses extrinsèques étaient distribuées de manière déterministe ( $P=1/1-P=0$ ). Il n'existe alors aucune dissonance cognitive. Troisièmement, ce biais de confirmation s'explique de manière computationnelle par des

modèles « optimistes », ce qui n'est que partiellement le cas pour les sujets sains (étude 1). Dans cette étude, la modélisation computationnelle a par ailleurs amené un argument supplémentaire pour une valeur intrinsèque du choix libre. En effet, l'ajout d'un « bonus » est nécessaire pour que la plupart des sujets ajuste mieux le modèle. Il a donc été possible de titrer le « coût » du choix libre. Ainsi, les sujets choisissent en proportion équivalente le choix libre du choix forcé lorsque ce dernier amène à des récompenses extrinsèques plus importantes que le premier, ce qui renforce la notion de « bonus » surajouté du choix libre comme récompense.

Au final, le libre choix semblerait par lui-même comporter un caractère gratifiant, indépendamment du contexte dans lequel il est exprimé. Toutefois, l'hypothèse de l'influence des récompenses extrinsèques n'est pas formellement exclue, même si leur influence a été redéfinie dans le cadre de nos deux études. Ces résultats questionnent alors sur les déterminants qui pourraient atténuer la valeur du choix libre.

### **Le choix pourrait-il être aversif ?**

Une certaine proportion de sujets, aussi bien sains que parkinsoniens, sélectionnaient le choix forcé plus fréquemment et même parfois de manière exclusive. Se pose alors la question d'un potentiel effet aversif du choix libre chez certains. Ce phénomène aversif pour le choix avait été évoqué par Iyengar & Lepper (2000) lors de plusieurs expériences en milieu écologique faisant varier la quantité de choix possible. Par exemple, une des études avait installé un stand alimentaire dans un commerce proposant à différentes heures de la journée, soit six choix possibles, soit 24. L'ensemble de ces études ont montré que même si l'abondance favorisait les comportements d'approche, le choix final était facilité lorsque le nombre d'option était plus limité (par exemple, les clients du commerce achetaient plus souvent des produits dans le stand de choix limité que celui avec le choix étendu). Un premier niveau de lecture peut être formulé facilement au vu de ces observations. En accord avec les théories d'apprentissage par renforcement, l'erreur de prédiction doit être

suffisamment ample et l'évènement suffisamment reproduit pour pouvoir dissocier les événements méritant « d'y revenir pour en avoir plus » selon la définition de la récompense par W. Schultz (2007). Dans le cas de multiples options à évaluer, l'estimation de la valeur subjective de chacune de ces options n'est suffisamment pas discriminante, amenant le sujet à restreindre le nombre d'options évaluables.

Le deuxième niveau d'explication, évoqué par les auteurs eux-mêmes, invoque le potentiel regret d'avoir écarté une ou plusieurs options qui auraient pu être meilleures. Cette notion se rapproche de l'ajustement particulier de certains des sujets sains à un modèle pessimiste. Pour rappel, ce modèle évalue à la fois la meilleure et la pire option pour calculer la valeur des actions dans une situation donnée. Une étude récente réalisée par Zorowitz et al. (2020) a utilisé un paradigme du choix en deux étapes pour simuler un modèle d'anxiété dit « pessimiste », similaire à celui de notre étude. En faisant varier un des paramètres, il était possible de rendre le modèle plus ou moins « pessimiste », c'est-à-dire qu'il apprenne plus de la pire option que de la meilleure. Ils ont pu montrer que plus le modèle est pessimiste, plus il aura tendance à éviter le choix, et ce probablement pour éviter de penser avoir fait le mauvais. Ce modèle est toutefois à nuancer puisqu'il s'agit de données simulées. De même, notre étude n'était pas conçue pour évaluer des aspects psychologiques et psychopathologiques mais l'hypothèse de structures de pensée distinctes pouvant influencer sur la préférence pour le choix libre mériterait d'être évaluée sur d'autres modèles pathologiques.

### **La préférence pour le choix reflète-t-elle des réponses impulsives chez les patients stimulés ?**

Dans notre deuxième étude chez les patients souffrant de maladie de Parkinson, le groupe des patients neuro-stimulés choisissent plus souvent et plus rapidement le choix libre que le choix forcé. Ce résultat comportemental interroge sur l'existence d'un certain degré d'impulsivité. L'utilisation du terme « impulsivité » suggérerait que ces patients prennent des décisions trop rapides, peu réfléchies et donc sub-optimales. Par exemple, dans une

tâche d'inhibition comme le Go-NoGo, les patients stimulés réalisent plus d'erreurs de commission ON que OFF (Ballanger et al., 2009). Cette tendance se retrouve également dans des tâches de cohérence de déplacement comportant des informations conflictuelles, et dans lesquelles les patients stimulés ON répondent plus rapidement et commettent plus d'erreur. Dans ce type de tâche, le sujet doit alors identifier dans un nuage de points se déplaçant de manière aléatoire, lesquels se dirigent dans une direction donnée (Green et al., 2013; Pote et al., 2016). Cette différence de performance entre patients neuro-stimulés ON et OFF est moins claire dans des tâches probabilistes comportant également du conflit, où deux stimulus sont récompensés avec une probabilité proche (Frank et al., 2007). On retrouve toutefois une accélération des réponses erronées des patients neuro-stimulés ON, y compris par rapport à des sujets sains. Ces trois exemples ont comme différence d'être traités par des systèmes cérébraux différents, intégrant des processus visuo-spatiaux dans le second exemple ou impliqué dans l'apprentissage par renforcement dans le dernier (Britten et al., 1992; Romo & Schultz, 1990). Dans le cas de la maladie de Parkinson, ces processus pourraient être altérés de manière hétérogène. La tâche de cohérence de mouvement requiert par exemple des ressources attentionnelles qui peuvent être déficitaires dans la maladie de Parkinson, conduisant à plus d'erreurs de la part des sujets. Celle de notre étude, d'apprentissage des valeurs, peut-être elle aussi altérée par un mauvais encodage de la valeur des récompenses. En revanche, tous ont en commun d'activer une voie rapide de sélection de l'action, fixant un seuil de décision en fonction des informations qu'elle reçoit des systèmes en amont. L'hypothèse que le NST soit un des centres appliquant ce seuil a été vérifié expérimentalement et modélisé, par exemple au moyen du *Drift Diffusion Model* dont les paramètres en lien avec ce seuil sont corrélés à l'activité du NST (Cavanagh et al., 2011; Green et al., 2013). Au final, il existerait un système inhibiteur au sein duquel le NST joue un rôle central, permettant la sélection d'actions appropriées, notamment en présence de conflit, et ceci dans des situations diverses. Dans la maladie de Parkinson, ce système serait plus conservateur du fait de l'hyperactivité du NST. En fonction du contexte, la stimulation n'exacerberait alors pas un comportement pathologique mais rétablirait un comportement plus physiologique.

Cette hypothèse est d'ailleurs renforcée par les performances des patients lorsque le traitement par stimulation est éteint. En effet, on observe chez les patients stimulés OFF une diminution de la préférence pour le choix libre autour du seuil d'indifférence (44%) et un ralentissement des réponses pour le choix libre comparé au choix forcé. Notons également que les patients stimulés reçoivent une quantité de traitement médicamenteux deux fois moins importante que les patients du groupe DOPA. Ces constatations peuvent être rapprochées des troubles amotivationnels bien connus observés au décours de la chirurgie chez certains patients (Czernecki et al., 2008; Lhommée et al., 2012; Thobois et al., 2010). Une des explications avancée serait la révélation de lésions mésolimbiques provoquée par la baisse importante du traitement dopaminergique après la chirurgie (Pagonabarraga et al., 2015; Thobois et al., 2010). La diminution de la préférence pour le choix chez les patients neuro-stimulés OFF dépendrait alors de deux mécanismes : tout d'abord une représentation altérée de la valeur subjective des stimulus « choix libre » et « choix forcé » par la déplétion dopaminergique, puis une majoration du seuil d'évidence permettant de discriminer ces deux stimulus médiés par l'hyperactivité du NST. En réajustant ce seuil décisionnel, la stimulation du NST permettrait de palier le message de conflit induit par une représentation sous-évaluée des deux stimulus.

### **Quel est le rôle du réseau fronto-subthalamique ?**

L'effet du NST sur le CCA dorsal a déjà été mis en évidence lors d'études en IRM fonctionnelle et scintigraphie, montrant une diminution de la perfusion de cette aire chez les patients stimulés (Ballanger et al., 2009; Thobois et al., 2007). Notre étude réalisée chez les patients parkinsoniens a mis en évidence que ceux ayant une plus grande différence de préférence pour le choix entre ON et OFF présentaient une VAT connectée de façon plus importante avec la région dorsale du cortex cingulaire antérieur (CCA, aire de Brodmann 32). Ce résultat anatomo-fonctionnel est concordant avec les connaissances actuelles sur la

connectivité du NST avec les régions corticales, en particulier le CCA (Haynes & Haber, 2013; Temiz et al., 2020). En effet, dans notre étude, les régions du NST préférentiellement connectées avec le CCA correspondaient aux sites de stimulation les plus antérieurs au sein du noyau. Comme cela avait été observé dans les études histologiques chez le primate non humain, la partie antérieure du NST présente une connectivité préférentielle avec les territoires corticaux en lien avec les fonctions limbiques et exécutives. La région d'intérêt provenant de l'étude de connectivité se situe d'ailleurs entre plusieurs régions fonctionnellement distinctes, à savoir l'aire de Brodmann (BA) 32, ainsi qu'une partie de BA 8 et 9. Puisque ces aires sont respectivement associées aux territoires limbiques et associatifs au sein du NST, cette observation pourrait rendre compte de l'absence de ségrégation nette entre les différents territoires en ce qui concerne la voie hyperdirecte. Cette constatation est toutefois à mettre en perspective avec la méthodologie utilisée dans notre étude. En effet, l'origine des fibres provenant du NST vers le cortex cérébral a été déterminée non pas en utilisant les coordonnées de localisation du site de stimulation mais le volume de tissu activé autour de l'électrode, ce dernier pouvant se superposer sur deux territoires fonctionnels distincts.

L'existence d'une connectivité fonctionnelle entre le cortex préfrontal médian et le NST a plusieurs fois été mis en évidence en particulier chez les patients parkinsoniens et semble être encore une fois rattaché au traitement d'informations conflictuelles. Cavanagh et al. (2011) a mis en évidence une activité thêta corrélée à l'apparition de conflit dans une région du cortex préfrontal médian. La région mise en évidence dans cette étude est toutefois plus postérieure que la nôtre correspondant au gyrus frontal supérieur et l'aire prémotrice (Brodmann 6) (Koessler et al., 2009). On notera toutefois que cette activité EEG apparaît de manière différée par rapport à l'apparition du stimulus de haut conflit, suggérant un mécanisme de préparation de l'action plutôt qu'un mécanisme d'évaluation du stimulus. Fleming et al. (2010) a mis en avant cette étape d'évaluation. Il a proposé à des sujets sains de réaliser une tâche dans laquelle ils devaient aller dans le sens ou contre l'avis d'un arbitre fictif de tennis annonçant une balle fautive. Le niveau de difficulté pour identifier où se

trouve la balle par rapport à la ligne variait selon les essais, ce qui s'apparente à une tâche de conflit. Cette étude est cependant considérée comme engageant les sujets dans du choix, car celui-ci doit exprimer sa propre opinion et parfois aller contre l'avis de l'arbitre (Leotti et al., 2010). L'analyse du signal en IRM fonctionnelle a montré qu'il existait une augmentation du signal BOLD au sein d'une région du CCA, similaire à la nôtre, ainsi que dans le NST, corrélée à une plus grande propension à exprimer soi-même son choix.

Ce résultat s'inscrit dans un corpus d'études précédentes s'étant intéressé au rôle du CCA dorsal dans le contrôle de l'action volontaire. Lors d'une tâche invoquant des stimulus récompensés, le signal BOLD du CCA dorsal est corrélée à la dévaluation de la quantité de récompense, incitant le sujet à changer de stratégie (Bush et al., 2002). Cette même région est également impliquée dans la décision volontaire des sujets à restreindre une action (appuyer sur une touche de clavier) de manière aléatoire, lorsqu'ils le décidaient (Haggard, 2008). Des régions plus antérieures du CCA dorsal ont été impliquées à la causalité de l'action associant tâche motrice et récompense délivrée après un nombre variable d'essais, se rapprochant alors de la notion d'agentivité. Les sujets pouvaient ainsi indiquer sur une échelle numérique à quel degré leur action avait été déterminante pour eux pour gagner la récompense (Tanaka et al., 2008). Certains auteurs évoquent aussi un rôle du CCA dorsal dans la valeur du contrôle, mettant en avant la connectivité de cette région avec les territoires limbiques du striatum, mais également avec les neurones de l'aire tegmentale ventrale (Shenhav et al., 2016; Williams & Goldman-Rakic, 1998). Ainsi, la nature du conflit évoqué au paragraphe précédent provoqué par la déplétion dopaminergique serait assimilable à une valuation imparfaite du contrôle que l'individu exerce sur son environnement.

Bien-sûr ce résultat a été obtenu suivant un a priori anatomique favorisant la mise en évidence d'une connectivité avec les territoires non-moteurs de la voie hyperdirecte. En effet, le masque délimitant la région d'analyse était issu d'une étude de la connectivité de la voie cortico-subthalamique chez l'homme (Temiz et al., 2020). Toutefois, cette étude a été

réalisée en utilisant des données provenant du Human Connectome Project chez des sujets sains. Dans notre étude, nous avons utilisé un connectome normatif issu d'une population parkinsonienne similaire à la nôtre car sélectionnée pour la chirurgie d'implantation d'électrodes profonde. Au total, le CCA et la voie CCA-NST seraient donc impliquée dans des mécanismes de d'évaluation du contrôle volontaire à des degrés divers de complexité et d'abstraction, allant du choix entre deux stimulus de valeurs proches, jusqu'à la détection d'un lien de causalité entre une action délibérée et ses conséquences.

### **Le NST est-il impliqué dans des mécanismes motivationnels ?**

Depuis plusieurs années, de nombreux travaux ont mis en évidence que le NST n'a pas un rôle inhibiteur mais promoteur de l'action à des degrés divers selon le type de récompenses. Ainsi, des lésions du NST chez le rat entraînent un effet différentiel lors de la présence de nourriture qui est plus consommée, et de cocaïne, qui elle est moins consommée (Baunez et al., 2005), de même pour la stimulation à haute fréquence du NST (Rouaud et al., 2010). Cette action serait sous-tendue par des populations de neurones pouvant encoder l'information apportée par des récompenses distinctes. Plusieurs enregistrements cellulaires du NST chez le singe ont mis en évidence une population de neurones dont l'activité est électivement associée à la présence de récompenses distinctes, encodant également leur caractère aversif ou appétitif, mais également au degré d'effort nécessaire pour obtenir ces récompenses, assimilable alors à un encodage motivationnel (Breysse et al., 2015; Espinosa-Parrilla et al., 2013; Nougaret et al., 2022). Des résultats similaires ont été obtenus chez l'humain lors d'enregistrements des potentiels de champs locaux du NST chez des patients parkinsoniens, montrant notamment qu'une activité dans des bandes de basse fréquence était corrélée au degré de motivation des sujets à obtenir une récompense monétaire (Zénon et al., 2016). Lors d'une étude avec enregistrement unitaire du NST chez l'homme, Justin Rossi et al. (2017) ont également mis en évidence des neurones encodant la valeur de stimulus récompensés dans le territoire limbique mais aussi dans le territoire sensorimoteur

du NST. Ce résultat apporte un élément supplémentaire pour un certain degré de superposition entre les différents territoires anatomo-fonctionnels. La caractérisation exacte de ces neurones reste par contre encore incomplète. Une des hypothèses serait une innervation par les neurones dopaminergiques permettant de relayer ce signal motivationnel. Des travaux histologiques chez le rat ont effectivement montré une arborisation axonale des neurones dopaminergiques parfois profuse de l'ensemble du NST mais chez le singe et chez l'homme, l'innervation dopaminergique du NST est faible et localisée à sa partie antérieure et médiane même s'il existe potentiellement une transmission volumique de la dopamine permettant de couvrir plus largement la structure cible (Agnati & Fuxe, 2000; Augood et al., 2000). Ces dernières constatations anatomiques restent toutefois des preuves indirectes pour expliquer de façon exhaustive l'hypothèse de neurones encodant une information motivationnelle dans le NST dans le cadre de notre étude.

### **Les traitements dopaminergiques ont-ils une durée d'action prolongée ?**

Nous avons montré que bien que les patients ayant un traitement dopaminergique ne présentaient pas de variation de leur préférence pour le choix lors d'un test aigu d'arrêt puis de réintroduction du traitement, la dose totale de traitement par L-DOPA (LEDD) influence la préférence pour le choix. En effet, les patients présentant une LEDD plus élevée avaient une préférence pour le choix élevée, tandis que ceux ayant une LEDD plus basse présentaient une préférence pour le choix proche du seuil d'indifférence (50%). Un effet aigu semble également se distinguer mais du fait du faible effectif dans cette analyse en sous-groupes, ce résultat n'était pas significatif ; il a été observé en revanche une accélération des réponses chez les patients ON, indépendamment de la cible choisie, suggérant un effet promotionnel. L'existence d'une réponse de longue durée à la L-DOPA chez les patients parkinsoniens est un phénomène connu depuis plusieurs décennies, notamment en ce qui concerne l'aspect moteur (Anderson & Nutt, 2011; Nutt et al., 1992). Ainsi, sur de larges

populations, il a été montré que des patients atteints de maladie de Parkinson traités présentaient après sevrage une diminution des symptômes moteurs jusqu'à 30% plus basse que ceux n'ayant jamais reçu de traitement à sévérité et durée égale de la maladie (Cilia et al., 2020). L'effet de la prise chronique de traitement dopaminergique est également connu pour faire émerger des comportements impulsifs et compulsifs chez un certain nombre de patients parkinsoniens. Ainsi, une grande étude regroupant plus de 3000 patients a mis en évidence que des patients présentant un trouble du contrôle des impulsions présentaient une LEDD plus élevée (Weintraub et al., 2010). En revanche, une étude longitudinale sur 5 ans portant sur des patients présentant une maladie de Parkinson *de novo* (moins de 5 ans d'évolution) n'a pas montré cette association. Il s'agissait de patients présentant des doses réduites de traitement (autour de 250 mg de L-DOPA), contrastant avec les caractéristiques de notre population qui ont une maladie plus sévère, et de plus hautes doses de traitement. D'autres comportements associés à la dose de L-DOPA ont également été décrits comme des comportements addictifs ou du *punding*. Ce dernier se définit comme l'engagement intense et excessif dans un comportement complexe et non orienté vers un but (par exemple, démonter des appareils électroniques ou collectionner des objets insignifiants ; Evans et al., 2004). Ce comportement automatique peut parfois altérer significativement le rythme quotidien du sujet, aussi bien en déprivation d'alimentation que de sommeil. Il est possible de réaliser un parallèle avec le comportement des singes dans l'étude de Harlow, ce dernier ayant suggéré qu'une « impulsion manipulatoire » peut être intrinsèquement motivant, le *punding* pourrait être alors le reflet pathologique d'un comportement intrinsèquement motivant.

Les mécanismes pharmacologiques suggérés pour expliquer cette réponse prolongée seraient un stockage prolongé de la L-DOPA dans des compartiments périphériques comme le tissu musculaire, une majoration de la transduction du signal promue par les récepteurs dopaminergiques exposés de façon répétée et prolongée à la L-DOPA, ou encore un effet direct sur l'activité des neurones dopaminergiques (Anderson & Nutt, 2011). Chez des rats traités au 6-OHDA, l'administration prolongée de L-DOPA pendant 4 semaines produit une

augmentation de l'activité spontanée des neurones dopaminergiques. Le mécanisme sous-jacent reposerait sur l'internalisation d'auto-récepteurs D2 inhibiteurs à la surface des cellules (Harden & Grace, 1995). Chez l'humain, au moyen d'études scintigraphiques utilisant des radio-ligands spécifiques de la dopamine, il apparaît une augmentation de la captation dopaminergique chez les patients présentant un trouble du contrôle des impulsions (Evans et al., 2006; Hammes et al., 2019).

Tant au niveau du striatum que de l'activité des neurones dopaminergiques, la prise prolongée de traitement par L-DOPA semblerait avoir un effet activateur de la voie mésolimbique. Cette hypothèse pourrait donc expliquer l'effet chronique sur la préférence pour le choix observée chez nos patients traités oralement. Bien que l'effet de la prise chronique d'agonistes dopaminergiques soit reliée avec l'altération des voies mésolimbiques et la survenue d'un trouble du contrôle des impulsions, elle n'a pas été mise en évidence dans notre étude probablement du fait du faible effectif.

## Conclusion

Au moyen d'un paradigme expérimental permettant d'analyser directement la valeur du choix, nous avons pu mettre en évidence que les individus préféraient majoritairement l'opportunité de pouvoir choisir lorsque cela était possible. Ce comportement permettrait d'augmenter le sentiment de contrôle sur les évènements. En utilisant comme modèle la maladie de Parkinson, nous avons montré que la stimulation à haute fréquence du NST permet d'augmenter la préférence pour le choix, en modulant directement le réseau entre le CCA dorsal et le NST, ce qui faciliterait la prise de décision. Nous avons aussi observé que cette préférence pour choisir faisait intervenir le système dopaminergique, dont la stimulation chronique par le traitement oral majorait cette préférence pour le choix. Ces résultats vont donc dans le sens que le choix est intrinsèquement motivant, au même titre que d'autres stimulus associés à des récompenses extrinsèques.

## Limites et perspectives

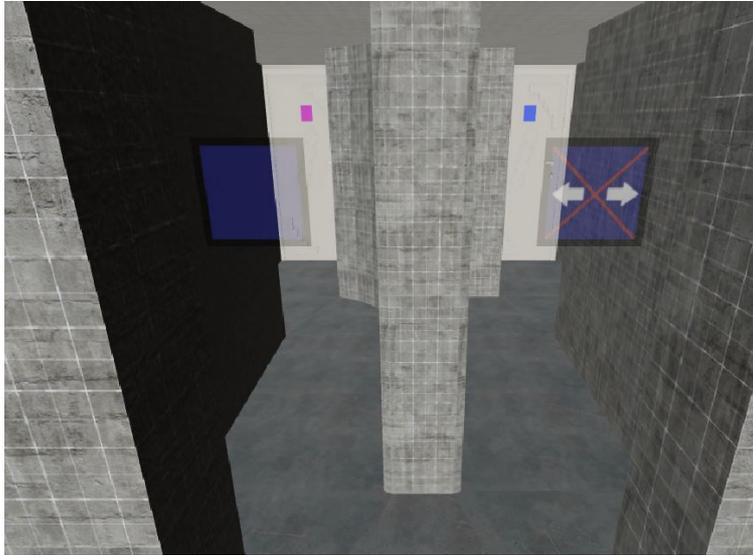
Plusieurs limites existent dans ce travail. Premièrement, le recrutement de patients qui souffrent d'une pathologie invalidante a conduit à restreindre le nombre de sessions afin d'adapter la tâche au handicap. Alors qu'il était également prévu de réaliser la tâche de contrôlabilité, sa longueur et sa complexité nous a conduit à réduire le temps de l'expérience. Dans le groupe de patients stimulés, il était également prévu de réaliser la tâche dans les quatre conditions de traitement, à savoir ON et OFF stimulation et ON et OFF traitement médicamenteux. Pour les mêmes raisons qu'évoquées précédemment, la fatigabilité des patients et le risque d'un désengagement attentionnel nous a conduit à restreindre l'expérience privilégiant l'effet de la stimulation. Deuxièmement, dans le groupe des patients traités oralement, nous avons mis en évidence un effet chronique de la quantité de traitement. Cette analyse en sous-groupe a été réalisée sur de petits effectifs limitant la puissance de cette analyse. Bien que nous ayons trouvé un effet significatif de la L-DOPA, cela n'a pas été le cas pour les patients sous agonistes dopaminergiques. Troisièmement, les groupes de patients présentaient également un manque de caractérisation sur le plan comportemental. En particulier, l'évaluation par l'échelle d'Ardouin des troubles comportementaux hyper et hypo-dopaminergiques était reportée de façon incomplète dans notre cohorte, rendant leur utilisation caduque (Ardouin et al., 2009). Quatrièmement, la mise en évidence des réseaux neuronaux impliqués a été faite avec le calcul probabiliste des VTA et à l'aide de données normatives. Bien que l'utilisation d'un connectome normatif soit une pratique courante et ayant mis en avant des résultats valides chez les patients parkinsoniens (Horn et al., 2017), les conclusions qui en émanent doivent se faire différemment que lors de l'utilisation de données individuelles. Ces résultats doivent se comprendre comme une généralisation de ce que devrait être la connectivité de sujets similaires à notre cohorte. Nous pensons que cette généralisation est valide compte tenu de l'origine des données du connectome utilisé ici, à savoir des patients parkinsoniens

candidats à la stimulation du NST. Enfin, l'étude précise de l'activité du NST lors de la réalisation de la tâche n'a pas pu encore être effectuée, par exemple au moyen de l'enregistrement des PCL.

Afin de poursuivre l'exploration des corrélats neuronaux sur la préférence pour le choix et répondre à certaines limites de notre étude, deux projets sont actuellement en cours. Le premier concerne la passation de notre tâche à une population de patients présentant une lésion focale du cerveau et en particulier des régions préfrontales. Pour cela, des patients présentant une tumeur cérébrale infiltrante (gliale) réalisent la tâche avant et après la chirurgie. Il est prévu alors de segmenter les lésions et de réaliser une analyse similaire à celle effectuée chez les patients stimulés pour obtenir une carte statistique des régions cérébrales impliquées dans les performances des patients. Cette étude permettra de vérifier l'hypothèse de l'implication de régions du cortex préfrontal, aussi bien médial que latéral. En effet, ces régions ont été mises en évidence dans des comportements similaires comme la perception du contrôle ou le sentiment d'agentivité (Haggard, 2017).

Le second projet implique des patients parkinsoniens neuro-stimulés qui réaliseront en réalité virtuelle le même parcours que celui du labyrinthe inventé par Voss & Homzie en 1970 mais modifié de façon à bien séparer les étapes comme dans notre étude. Ce projet sera également complété par un enregistrement des PCL en temps réel de manière non invasive (technologie Percept). Ainsi, il sera possible de corrélérer les événements de la tâche, notamment lors de la première étape, à l'activité neuronale du NST. De plus, l'utilisation de la réalité virtuelle au moyen d'un casque permet une immersion rapide dans l'expérience et permet de créer des conditions plus écologiques. Par rapport au labyrinthe original qui comprenait une voie finale commune aux deux conditions, nous avons fait varier à la seconde étape la quantité de récompense pouvant être obtenue. Tout comme la tâche réalisée sur ordinateur, le patient aura la possibilité de se diriger dans le bras sans choix mais présentant une grande récompense, ou bien de se diriger vers le bras avec choix qui mène à

son tour vers un embranchement comportant respectivement une petite ou une grande récompense.



**Figure 1.** Capture d'image du labyrinthe au moment de la première étape. Le patient a alors la possibilité de se diriger vers un essais choix ou un essai non choix. Sur cette image, la condition choix est barrée car il s'agit d'un essai d'apprentissage des différents bras du labyrinthe.

## Références

- Académie nationale de médecine. (2003). *Le corps de Luys*. <https://www.academie-medecine.fr/le-corps-de-luys-2/>
- Agnati, L. F., & Fuxe, K. (2000). Volume transmission as a key feature of information handling in the central nervous system possible new interpretative value of the Turing's B-type machine. In *Progress in Brain Research* (Vol. 125, p. 3-19). Elsevier. [https://doi.org/10.1016/S0079-6123\(00\)25003-6](https://doi.org/10.1016/S0079-6123(00)25003-6)
- Alberico, S. L., Cassell, M. D., & Narayanan, N. S. (2015). The Vulnerable Ventral Tegmental Area in Parkinson's Disease. *Basal ganglia*, 5(2-3), 51-55. <https://doi.org/10.1016/j.baga.2015.06.001>
- Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience*, 9, 357-381. <https://doi.org/10.1146/annurev.ne.09.030186.002041>
- Anderson, E., & Nutt, J. (2011). The long-duration response to levodopa : Phenomenology, potential mechanisms and clinical implications. *Parkinsonism & Related Disorders*, 17(8), 587-592. <https://doi.org/10.1016/j.parkreldis.2011.03.014>
- Ardouin, C., Chéreau, I., Llorca, P.-M., Lhommée, E., Durif, F., Pollak, P., & Krack, P. (2009). Évaluation des troubles comportementaux hyper- et hypodopaminergiques dans la maladie de Parkinson. */data/revues/00353787/v165i11/S0035378709003439/*. <https://www.em-consulte.com/en/article/231254>
- Ardouin, C., Pillon, B., Peiffer, E., Bejjani, P., Limousin, P., Damier, P., Arnulf, I., Benabid, A. L., Agid, Y., & Pollak, P. (1999). Bilateral subthalamic or pallidal stimulation for Parkinson's disease affects neither memory nor executive functions : A consecutive series of 62 patients. *Annals of Neurology*, 46(2), 217-223. [https://doi.org/10.1002/1531-8249\(199908\)46:2<217::aid-ana11>3.0.co;2-z](https://doi.org/10.1002/1531-8249(199908)46:2<217::aid-ana11>3.0.co;2-z)
- Atkinson-Clement, C., Cavazzini, É., Zénon, A., Witjas, T., Fluchère, F., Azulay, J.-P., Baunez, C., & Eusebio, A. (2019). Effects of subthalamic nucleus stimulation and levodopa on decision-making in Parkinson's disease. *Movement Disorders: Official Journal of the Movement Disorder Society*, 34(3), 377-385. <https://doi.org/10.1002/mds.27625>
- Augood, S. J., Hollingsworth, Z. R., Standaert, D. G., Emson, P. C., & Penney Jr., J. B. (2000). Localization of dopaminergic markers in the human subthalamic nucleus. *Journal of Comparative Neurology*, 421(2), 247-255. [https://doi.org/10.1002/\(SICI\)1096-9861\(20000529\)421:2<247::AID-CNE9>3.0.CO;2-F](https://doi.org/10.1002/(SICI)1096-9861(20000529)421:2<247::AID-CNE9>3.0.CO;2-F)
- Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric Diffeomorphic Image Registration with Cross-Correlation : Evaluating Automated Labeling of Elderly and Neurodegenerative Brain. *Medical image analysis*, 12(1), 26-41. <https://doi.org/10.1016/j.media.2007.06.004>
- Ballanger, B., van Eimeren, T., Moro, E., Lozano, A. M., Hamani, C., Boulinguez, P., Pellecchia, G., Houle, S., Poon, Y. Y., Lang, A. E., & Strafella, A. P. (2009).

- Stimulation of the subthalamic nucleus and impulsivity : Release your horses. *Annals of Neurology*, *66*(6), 817-824. <https://doi.org/10.1002/ana.21795>
- Bandura, A. (1977). Self-efficacy : Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191-215. <https://doi.org/10.1037/0033-295X.84.2.191>
- Bardinet, E., Bhattacharjee, M., Dormont, D., Pidoux, B., Malandain, G., Schüpbach, M., Ayache, N., Cornu, P., Agid, Y., & Yelnik, J. (2009). A three-dimensional histological atlas of the human basal ganglia. II. Atlas deformation strategy and evaluation in deep brain stimulation for Parkinson disease. *Journal of Neurosurgery*, *110*(2), 208-219. <https://doi.org/10.3171/2008.3.17469>
- Bassett, L., & Buchanan-Smith, H. M. (2007). Effects of predictability on the welfare of captive animals. *Applied Animal Behaviour Science*, *102*(3-4), 223-245. <https://doi.org/10.1016/j.applanim.2006.05.029>
- Baunez, C., Christakou, A., Chudasama, Y., Forni, C., & Robbins, T. W. (2007). Bilateral high-frequency stimulation of the subthalamic nucleus on attentional performance : Transient deleterious effects and enhanced motivation in both intact and parkinsonian rats. *European Journal of Neuroscience*, *25*(4), 1187-1194. <https://doi.org/10.1111/j.1460-9568.2007.05373.x>
- Baunez, C., Dias, C., Cador, M., & Amalric, M. (2005). The subthalamic nucleus exerts opposite control on cocaine and « natural » rewards. *Nature Neuroscience*, *8*(4), 484-489. <https://doi.org/10.1038/nn1429>
- Baunez, C., Nieoullon, A., & Amalric, M. (1995). In a rat model of parkinsonism, lesions of the subthalamic nucleus reverse increases of reaction time but induce a dramatic premature responding deficit. *Journal of Neuroscience*, *15*(10), 6531-6541. <https://doi.org/10.1523/JNEUROSCI.15-10-06531.1995>
- Baunez, C., & Robbins, T. W. (1997). Bilateral lesions of the subthalamic nucleus induce multiple deficits in an attentional task in rats. *The European Journal of Neuroscience*, *9*(10), 2086-2099. <https://doi.org/10.1111/j.1460-9568.1997.tb01376.x>
- Benazzouz, A., Gross, C., Féger, J., Boraud, T., & Bioulac, B. (1993). Reversal of rigidity and improvement in motor performance by subthalamic high-frequency stimulation in MPTP-treated monkeys. *European Journal of Neuroscience*, *5*(4), 382-389.
- Bergman, H., Wichmann, T., & DeLong, M. R. (1990). Reversal of experimental parkinsonism by lesions of the subthalamic nucleus. *Science (New York, N.Y.)*, *249*(4975), 1436-1438. <https://doi.org/10.1126/science.2402638>
- Björklund, A., & Dunnett, S. B. (2007). Dopamine neuron systems in the brain : An update. *Trends in Neurosciences*, *30*(5), 194-202. <https://doi.org/10.1016/j.tins.2007.03.006>
- Bobadilla-Suarez, S., Sunstein, C. R., & Sharot, T. (2017). The intrinsic value of choice : The propensity to under-delegate in the face of potential gains and losses. *Journal of Risk and Uncertainty*, *54*(3), 187-202. <https://doi.org/10.1007/s11166-017-9259-x>
- Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models : A practical guide for ecology and evolution. *Trends in Ecology & Evolution*, *24*(3), 127-135. <https://doi.org/10.1016/j.tree.2008.10.008>
- Boraud, T., Bezard, E., Bioulac, B., & Gross, C. (1996). High frequency stimulation of the internal Globus Pallidus (GPi) simultaneously improves parkinsonian symptoms and

- reduces the firing frequency of GPi neurons in the MPTP-treated monkey. *Neuroscience Letters*, *215*(1), 17-20. [https://doi.org/10.1016/S0304-3940\(96\)12943-8](https://doi.org/10.1016/S0304-3940(96)12943-8)
- Bown, N. J., Read, D., & Summers, B. (2003). The lure of choice. *Journal of Behavioral Decision Making*, *16*(4), 297-308. <https://doi.org/10.1002/bdm.447>
- Braak, H., Del Tredici, K., Rüb, U., de Vos, R. A. I., Jansen Steur, E. N. H., & Braak, E. (2003). Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiology of Aging*, *24*(2), 197-211.
- Brehm, J. W. (1956). Postdecision changes in the desirability of alternatives. *The Journal of Abnormal and Social Psychology*, *52*(3), 384-389. <https://doi.org/10.1037/h0041006>
- Breyse, E., Pelloux, Y., & Baunez, C. (2015). The Good and Bad Differentially Encoded within the Subthalamic Nucleus in Rats. *eNeuro*, *2*(5), ENEURO.0014-15.2015. <https://doi.org/10.1523/ENEURO.0014-15.2015>
- Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: A comparison of neuronal and psychophysical performance. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *12*(12), 4745-4765.
- Bromberg-Martin, E. S., & Hikosaka, O. (2009). Midbrain Dopamine Neurons Signal Preference for Advance Information about Upcoming Rewards. *Neuron*, *63*(1), 119-126. <https://doi.org/10.1016/j.neuron.2009.06.009>
- Bush, G., Vogt, B. A., Holmes, J., Dale, A. M., Greve, D., Jenike, M. A., & Rosen, B. R. (2002). Dorsal anterior cingulate cortex: A role in reward-based decision making. *Proceedings of the National Academy of Sciences*, *99*(1), 523-528. <https://doi.org/10.1073/pnas.012470999>
- Catania, A. C., & Sagvolden, T. (1980). Preference for Free Choice Over Forced Choice in Pigeons. *Journal of the Experimental Analysis of Behavior*, *34*(1), 77-86. <https://doi.org/10.1901/jeab.1980.34-77>
- Cavanagh, J. F., Wiecki, T. V., Cohen, M. X., Figueroa, C. M., Samanta, J., Sherman, S. J., & Frank, M. J. (2011). Subthalamic nucleus stimulation reverses mediofrontal influence over decision threshold. *Nature Neuroscience*, *14*(11), 1462-1467. <https://doi.org/10.1038/nn.2925>
- Chambon, V., Théro, H., Vidal, M., Vandendriessche, H., Haggard, P., & Palminteri, S. (2020). Information about action outcomes differentially affects learning from self-determined versus imposed choices. *Nature Human Behaviour*. <https://doi.org/10.1038/s41562-020-0919-5>
- Chinaglia, G., Alvarez, F. J., Probst, A., & Palacios, J. M. (1992). Mesostriatal and mesolimbic dopamine uptake binding sites are reduced in Parkinson's disease and progressive supranuclear palsy: A quantitative autoradiographic study using [3H]mazindol. *Neuroscience*, *49*(2), 317-327. [https://doi.org/10.1016/0306-4522\(92\)90099-n](https://doi.org/10.1016/0306-4522(92)90099-n)
- Cilia, R., Cereda, E., Akpalu, A., Sarfo, F. S., Cham, M., Laryea, R., Obese, V., Oppon, K., Del Sorbo, F., Bonvegna, S., Zecchinelli, A. L., & Pezzoli, G. (2020). Natural history of motor symptoms in Parkinson's disease and the long-duration response to levodopa. *Brain: A Journal of Neurology*, *143*(8), 2490-2501. <https://doi.org/10.1093/brain/awaa181>

- Cockburn, J., Collins, A. G. E., & Frank, M. J. (2014). A Reinforcement Learning Mechanism Responsible for the Valuation of Free Choice. *Neuron*, *83*(3), 551-557. <https://doi.org/10.1016/j.neuron.2014.06.035>
- Collins, A. G. E., & Cockburn, J. (2020). Beyond dichotomies in reinforcement learning. *Nature Reviews Neuroscience*, *21*(10), 576-586. <https://doi.org/10.1038/s41583-020-0355-6>
- Cools, R., Barker, R. A., Sahakian, B. J., & Robbins, T. W. (2001). Enhanced or Impaired Cognitive Function in Parkinson's Disease as a Function of Dopaminergic Medication and Task Demands. *Cerebral Cortex*, *11*(12), 1136-1143. <https://doi.org/10.1093/cercor/11.12.1136>
- Corvol, J.-C., Artaud, F., Cormier-Dequaire, F., Rascol, O., Durif, F., Derkinderen, P., Marques, A.-R., Bourdain, F., Brandel, J.-P., Pico, F., Lacomblez, L., Bonnet, C., Brefel-Courbon, C., Ory-Magne, F., Grabli, D., Klebe, S., Mangone, G., You, H., Mesnage, V., ... DIGPD Study Group. (2018). Longitudinal analysis of impulse control disorders in Parkinson disease. *Neurology*, *91*(3), e189-e201. <https://doi.org/10.1212/WNL.0000000000005816>
- Crossman, A. R., Sambrook, M. A., & Jackson, A. (1984). Experimental hemichorea/hemiballismus in the monkey : Studies on the intracerebral site of action in a drug-induced dyskinesia. *Brain*, *107*(2), 579-596. <https://doi.org/10.1093/brain/107.2.579>
- Czernecki, V., Schüpbach, M., Yaici, S., Lévy, R., Bardinnet, E., Yelnik, J., Dubois, B., & Agid, Y. (2008). Apathy following subthalamic stimulation in Parkinson disease: A dopamine responsive symptom. *Movement Disorders: Official Journal of the Movement Disorder Society*, *23*(7), 964-969. <https://doi.org/10.1002/mds.21949>
- Damier, P., Hirsch, E. C., Agid, Y., & Graybiel, A. M. (1999). The substantia nigra of the human brain. II. Patterns of loss of dopamine-containing neurons in Parkinson's disease. *Brain: A Journal of Neurology*, *122* ( Pt 8), 1437-1448. <https://doi.org/10.1093/brain/122.8.1437>
- Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*(6), 1204-1215. <https://doi.org/10.1016/j.neuron.2011.02.027>
- Delgado, M. R., Nystrom, L. E., Fissell, C., Noll, D. C., & Fiez, J. A. (2000). Tracking the Hemodynamic Responses to Reward and Punishment in the Striatum. *Journal of Neurophysiology*, *84*(6), 3072-3077. <https://doi.org/10.1152/jn.2000.84.6.3072>
- DeLong, M. R. (1972). Activity of basal ganglia neurons during movement. *Brain Research*, *40*(1). [https://doi.org/10.1016/0006-8993\(72\)90118-7](https://doi.org/10.1016/0006-8993(72)90118-7)
- Dembek, T. A., Hellerbach, A., Jergas, H., Eichner, M., Wirths, J., Dafsari, H. S., Barbe, M. T., Hunsche, S., Visser-Vandewalle, V., & Treuer, H. (2021). DiODe v2: Unambiguous and Fully-Automated Detection of Directional DBS Lead Orientation. *Brain Sciences*, *11*(11), 1450. <https://doi.org/10.3390/brainsci11111450>
- Desbonnet, L., Temel, Y., Visser-Vandewalle, V., Blokland, A., Hornikx, V., & Steinbusch, H. W. M. (2004). Premature responding following bilateral stimulation of the rat subthalamic nucleus is amplitude and frequency dependent. *Brain Research*, *1008*(2), 198-204. <https://doi.org/10.1016/j.brainres.2004.02.032>

- Deuschl, G., Schade-Brittinger, C., Krack, P., Volkmann, J., Schäfer, H., Bötzel, K., Daniels, C., Deutschländer, A., Dillmann, U., Eisner, W., Gruber, D., Hamel, W., Herzog, J., Hilker, R., Klebe, S., Kloss, M., Koy, J., Krause, M., Kupsch, A., ... German Parkinson Study Group, Neurostimulation Section. (2006). A randomized trial of deep-brain stimulation for Parkinson's disease. *The New England Journal of Medicine*, *355*(9), 896-908. <https://doi.org/10.1056/nejmoa060281>
- Eagle, D. M., & Baunez, C. (2010). Is there an inhibitory-response-control system in the rat? Evidence from anatomical and pharmacological studies of behavioral inhibition. *Neuroscience & Biobehavioral Reviews*, *34*(1), 50-72. <https://doi.org/10.1016/j.neubiorev.2009.07.003>
- Espinosa-Parrilla, J.-F., Baunez, C., & Apicella, P. (2013). Linking reward processing to behavioral output : Motor and motivational integration in the primate subthalamic nucleus. *Frontiers in Computational Neuroscience*, *7*, 175. <https://doi.org/10.3389/fncom.2013.00175>
- Evans, A. H., Katzenschlager, R., Paviour, D., O'Sullivan, J. D., Appel, S., Lawrence, A. D., & Lees, A. J. (2004). Punding in Parkinson's disease : Its relation to the dopamine dysregulation syndrome. *Movement Disorders*, *19*(4), 397-405. <https://doi.org/10.1002/mds.20045>
- Evans, A. H., Pavese, N., Lawrence, A. D., Tai, Y. F., Appel, S., Doder, M., Brooks, D. J., Lees, A. J., & Piccini, P. (2006). Compulsive drug use linked to sensitized ventral striatal dopamine transmission. *Annals of Neurology*, *59*(5), 852-858. <https://doi.org/10.1002/ana.20822>
- Evans, A. H., Strafella, A. P., Weintraub, D., & Stacy, M. (2009). Impulsive and compulsive behaviors in Parkinson's disease. *Movement Disorders: Official Journal of the Movement Disorder Society*, *24*(11), 1561-1570. <https://doi.org/10.1002/mds.22505>
- Festinger, L. (1957). *A theory of cognitive dissonance* (p. xi, 291). Stanford University Press.
- Fiorillo, C. D., Tobler, P. N., & Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science (New York, N.Y.)*, *299*(5614), 1898-1902. <https://doi.org/10.1126/science.1077349>
- Fleming, S. M., Thomas, C. L., & Dolan, R. J. (2010). Overcoming status quo bias in the human brain. *Proceedings of the National Academy of Sciences*, *107*(13), 6005-6009. <https://doi.org/10.1073/pnas.0910380107>
- Fonov, V., Evans, A. C., Botteron, K., Almli, C. R., McKinstry, R. C., & Collins, D. L. (2011). Unbiased Average Age-Appropriate Atlases for Pediatric Studies. *NeuroImage*, *54*(1), 313-327. <https://doi.org/10.1016/j.neuroimage.2010.07.033>
- Fox, S. H., Katzenschlager, R., Lim, S.-Y., Barton, B., de Bie, R. M. A., Seppi, K., Coelho, M., Sampaio, C., & Committee, on behalf of the M. D. S. E.-B. M. (2018). International Parkinson and movement disorder society evidence-based medicine review : Update on treatments for the motor symptoms of Parkinson's disease. *Movement Disorders*, *33*(8), 1248-1266. <https://doi.org/10.1002/mds.27372>
- François, C., Yelnik, J., Percheron, G., & Fénelon, G. (1994). Topographic distribution of the axonal endings from the sensorimotor and associative striatum in the macaque pallidum and substantia nigra. *Experimental Brain Research*, *102*(2), 305-318. <https://doi.org/10.1007/BF00227517>

- François, C., Yelnik, J., Percheron, G., & Tandé, D. (1994). Calbindin D-28k as a marker for the associative cortical territory of the striatum in macaque. *Brain Research*, *633*(1), 331-336. [https://doi.org/10.1016/0006-8993\(94\)91557-1](https://doi.org/10.1016/0006-8993(94)91557-1)
- Frank, M. J. (2005). Dynamic dopamine modulation in the basal ganglia: A neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. *Journal of Cognitive Neuroscience*, *17*(1), 51-72. <https://doi.org/10.1162/0898929052880093>
- Frank, M. J., Samanta, J., Moustafa, A. A., & Sherman, S. J. (2007). Hold Your Horses: Impulsivity, Deep Brain Stimulation, and Medication in Parkinsonism. *Science*, *318*(5854), 1309-1312. <https://doi.org/10.1126/science.1146157>
- Frank, M. J., Seeberger, L. C., & O'Reilly, R. C. (2004). By Carrot or by Stick: Cognitive Reinforcement Learning in Parkinsonism. *Science*, *306*(5703), 1940-1943. JSTOR. <https://doi.org/10.1126/science.1102941>
- Friston, K. J. (Éd.). (2007). *Statistical parametric mapping: The analysis of functional brain images* (1st ed). Elsevier/Academic Press.
- Georgopoulos, A., DeLong, M., & Crutcher, M. (1983). Relations between parameters of step-tracking movements and single cell discharge in the globus pallidus and subthalamic nucleus of the behaving monkey. *The Journal of Neuroscience*, *3*(8), 1586-1598. <https://doi.org/10.1523/JNEUROSCI.03-08-01586.1983>
- Grahn, J. A., Parkinson, J. A., & Owen, A. M. (2008). The cognitive functions of the caudate nucleus. *Progress in Neurobiology*, *86*(3), 141-155. <https://doi.org/10.1016/j.pneurobio.2008.09.004>
- Green, N., Bogacz, R., Huebl, J., Beyer, A.-K., Kühn, A. A., & Heekeren, H. R. (2013). Reduction of Influence of Task Difficulty on Perceptual Decision Making by STN Deep Brain Stimulation. *Current Biology*, *23*(17), 1681-1684. <https://doi.org/10.1016/j.cub.2013.07.001>
- Haber, S. N. (2014). The place of dopamine in the cortico-basal ganglia circuit. *Neuroscience*, *282*, 248-257. <https://doi.org/10.1016/j.neuroscience.2014.10.008>
- Haggard, P. (2008). Human volition: Towards a neuroscience of will. *Nature Reviews Neuroscience*, *9*(12), 934-946. <https://doi.org/10.1038/nrn2497>
- Haggard, P. (2017). Sense of agency in the human brain. *Nature Reviews Neuroscience*, *18*(4), 196-207. <https://doi.org/10.1038/nrn.2017.14>
- Hammes, J., Theis, H., Giehl, K., Hoenig, M. C., Greuel, A., Tittgemeyer, M., Timmermann, L., Fink, G. R., Drzezga, A., Eggers, C., & van Eimeren, T. (2019). Dopamine metabolism of the nucleus accumbens and fronto-striatal connectivity modulate impulse control. *Brain*, *142*(3), 733-743. <https://doi.org/10.1093/brain/awz007>
- Harden, D., & Grace, A. (1995). Activation of dopamine cell firing by repeated L-DOPA administration to dopamine-depleted rats: Its potential role in mediating the therapeutic response to L-DOPA treatment. *The Journal of Neuroscience*, *15*(9), 6157-6166. <https://doi.org/10.1523/JNEUROSCI.15-09-06157.1995>
- Harlow, H. F., Harlow, M. K., & Meyer, D. R. (1950). Learning motivated by a manipulation drive. *Journal of Experimental Psychology*, *40*(2), 228-234. <https://doi.org/10.1037/h0056906>

- Harrison, X. A., Donaldson, L., Correa-Cano, M. E., Evans, J., Fisher, D. N., Goodwin, C. E. D., Robinson, B. S., Hodgson, D. J., & Inger, R. (2018). A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ*, *6*, e4794. <https://doi.org/10.7717/peerj.4794>
- Hashimoto, T., Elder, C. M., Okun, M. S., Patrick, S. K., & Vitek, J. L. (2003). Stimulation of the subthalamic nucleus changes the firing pattern of pallidal neurons. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *23*(5), 1916-1923.
- Haynes, W. I. A., & Haber, S. N. (2013). The organization of prefrontal-subthalamic inputs in primates provides an anatomical substrate for both functional specificity and integration : Implications for Basal Ganglia models and deep brain stimulation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(11), 4804-4814. <https://doi.org/10.1523/jneurosci.4674-12.2013>
- Horn, A., Li, N., Dembek, T. A., Kappel, A., Boulay, C., Ewert, S., Tietze, A., Husch, A., Perera, T., Neumann, W.-J., Reisert, M., Si, H., Oostenveld, R., Rorden, C., Yeh, F.-C., Fang, Q., Herrington, T. M., Vorwerk, J., & Kühn, A. A. (2019). Lead-DBS v2 : Towards a comprehensive pipeline for deep brain stimulation imaging. *NeuroImage*, *184*, 293-316. <https://doi.org/10.1016/j.neuroimage.2018.08.068>
- Horn, A., Reich, M., Vorwerk, J., Li, N., Wenzel, G., Fang, Q., Schmitz-Hübsch, T., Nickl, R., Kupsch, A., & Volkmann, J. (2017). Connectivity predicts deep brain stimulation outcome in P arkinson disease. *Annals of neurology*, *82*(1), 67-78.
- Hughes, A. J., Daniel, S. E., Kilford, L., & Lees, A. J. (1992). Accuracy of clinical diagnosis of idiopathic Parkinson's disease : A clinico-pathological study of 100 cases. *Journal of Neurology, Neurosurgery, and Psychiatry*, *55*(3), 181-184. <https://doi.org/10.1136/jnnp.55.3.181>
- Husch, A., V. Petersen, M., Gemmar, P., Goncalves, J., & Hertel, F. (2018). PaCER - A fully automated method for electrode trajectory and contact reconstruction in deep brain stimulation. *NeuroImage: Clinical*, *17*, 80-89. <https://doi.org/10.1016/j.nicl.2017.10.004>
- Iyengar, S. S., & Lepper, M. R. (2000). When choice is demotivating : Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, *79*(6), 995-1006. <https://doi.org/10.1037/0022-3514.79.6.995>
- Jahanshahi, M., Ardouin, C. M. A., Brown, R. G., Rothwell, J. C., Obeso, J., Albanese, A., Rodriguez-Oroz, M. C., Moro, E., Benabid, A. L., Pollak, P., & Limousin-Dowsey, P. (2000). The impact of deep brain stimulation on executive function in Parkinson's disease. *Brain*, *123*(6), 1142-1154. <https://doi.org/10.1093/brain/123.6.1142>
- Justin Rossi, P., Peden, C., Castellanos, O., Foote, K. D., Gunduz, A., & Okun, M. S. (2017). The human subthalamic nucleus and globus pallidus internus differentially encode reward during action control. *Human Brain Mapping*, *38*(4), 1952-1964. PubMed. <https://doi.org/10.1002/hbm.23496>
- Kaplin, A. I., & Montel Williams, M. W. M. F. (2019). *How common are the « common » neurologic disorders?* <https://n.neurology.org/content/how-common-are-common-neurologic-disorders>

- Karachi, C., François, C., Parain, K., Bardinet, E., Tandé, D., Hirsch, E., & Yelnik, J. (2002). Three-dimensional cartography of functional territories in the human striatopallidal complex by using calbindin immunoreactivity. *Journal of Comparative Neurology*, *450*(2), 122-134. <https://doi.org/10.1002/cne.10312>
- Karachi, C., Grabli, D., Baup, N., Mounayar, S., Tandé, D., François, C., & Hirsch, E. C. (2009). Dysfunction of the subthalamic nucleus induces behavioral and movement disorders in monkeys. *Movement Disorders*, *24*(8), 1183-1192. <https://doi.org/10.1002/mds.22547>
- Karachi, C., Yelnik, J., Tandé, D., Tremblay, L., Hirsch, E. C., & François, C. (2005). The pallidosubthalamic projection : An anatomical substrate for nonmotor functions of the subthalamic nucleus in primates. *Movement Disorders: Official Journal of the Movement Disorder Society*, *20*(2), 172-180. <https://doi.org/10.1002/mds.20302>
- Koessler, L., Maillard, L., Benhadid, A., Vignal, J. P., Felblinger, J., Vespignani, H., & Braun, M. (2009). Automated cortical projection of EEG sensors : Anatomical correlation via the international 10-10 system. *NeuroImage*, *46*(1), 64-72. <https://doi.org/10.1016/j.neuroimage.2009.02.006>
- Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology*, *32*(2), 311-328. <https://doi.org/10.1037/0022-3514.32.2.311>
- Lardeux, S., Paleressompoulle, D., Pernaud, R., Cador, M., & Baunez, C. (2013). Different populations of subthalamic neurons encode cocaine vs. Sucrose reward and predict future error. *Journal of Neurophysiology*, *110*(7), 1497-1510. <https://doi.org/10.1152/jn.00160.2013>
- Lavoie, B., & Parent, A. (1990). Immunohistochemical study of the serotonergic innervation of the basal ganglia in the squirrel monkey. *The Journal of Comparative Neurology*, *299*(1), 1-16. <https://doi.org/10.1002/cne.902990102>
- Le Jeune, F., Drapier, D., Bourguignon, A., Péron, J., Mesbah, H., Drapier, S., Sauleau, P., Haegelen, C., Travers, D., Garin, E., Malbert, C. H., Millet, B., & Vérin, M. (2009). Subthalamic nucleus stimulation in Parkinson disease induces apathy : A PET study. *Neurology*, *73*(21), 1746-1751. <https://doi.org/10.1212/wnl.0b013e3181c34b34>
- Leotti, L. A., & Delgado, M. R. (2014). The Value of Exercising Control Over Monetary Gains and Losses. *Psychological Science*, *25*(2), 596-604. <https://doi.org/10.1177/0956797613514589>
- Leotti, L. A., Iyengar, S. S., & Ochsner, K. N. (2010). Born to choose : The origins and value of the need for control. *Trends in Cognitive Sciences*, *14*(10), 457-463. <https://doi.org/10.1016/j.tics.2010.08.001>
- Levy, D. J., & Glimcher, P. W. (2012). The root of all value : A neural common currency for choice. *Current Opinion in Neurobiology*, *22*(6), 1027-1038. <https://doi.org/10.1016/j.conb.2012.06.001>
- Lhommée, E., Klinger, H., Thobois, S., Schmitt, E., Ardouin, C., Bichon, A., Kistner, A., Fraix, V., Xie, J., Aya Kombo, M., Chabardès, S., Seigneuret, E., Benabid, A.-L., Mertens, P., Polo, G., Carnicella, S., Quesada, J.-L., Bosson, J.-L., Broussolle, E., ... Krack, P. (2012). Subthalamic stimulation in Parkinson's disease : Restoring the balance of motivated behaviours. *Brain*, *135*(5), 1463-1477. <https://doi.org/10.1093/brain/aws078>

- Lhommée, E., Wojtecki, L., Czernecki, V., Witt, K., Maier, F., Tonder, L., Timmermann, L., Hälbig, T. D., Pineau, F., Durif, F., Witjas, T., Pinsker, M., Mehdorn, M., Sixel-Döring, F., Kupsch, A., Krüger, R., Elben, S., Chabardès, S., Thobois, S., ... EARLYSTIM study group. (2018). Behavioural outcomes of subthalamic stimulation and medical therapy versus medical therapy alone for Parkinson's disease with early motor complications (EARLYSTIM trial): Secondary analysis of an open-label randomised trial. *The Lancet. Neurology*, *17*(3), 223-231. [https://doi.org/10.1016/s1474-4422\(18\)30035-8](https://doi.org/10.1016/s1474-4422(18)30035-8)
- Maier-Hein, K. H., Neher, P. F., Houde, J.-C., Côté, M.-A., Garyfallidis, E., Zhong, J., Chamberland, M., Yeh, F.-C., Lin, Y.-C., Ji, Q., Reddick, W. E., Glass, J. O., Chen, D. Q., Feng, Y., Gao, C., Wu, Y., Ma, J., He, R., Li, Q., ... Descoteaux, M. (2017). The challenge of mapping the human connectome based on diffusion tractography. *Nature Communications*, *8*(1), 1349. <https://doi.org/10.1038/s41467-017-01285-x>
- McIntyre, C. C., Grill, W. M., Sherman, D. L., & Thakor, N. V. (2004). Cellular effects of deep brain stimulation: Model-based analysis of activation and inhibition. *Journal of neurophysiology*, *91*(4), 1457-1469.
- Middleton, F. A., & Strick, P. L. (2002). Basal-ganglia 'Projections' to the Prefrontal Cortex of the Primate. *Cerebral Cortex*, *12*(9), 926-935. <https://doi.org/10.1093/cercor/12.9.926>
- Mink, J. W. (1996). THE BASAL GANGLIA: FOCUSED SELECTION AND INHIBITION OF COMPETING MOTOR PROGRAMS. *Progress in Neurobiology*, *50*(4), 381-425. [https://doi.org/10.1016/s0301-0082\(96\)00042-1](https://doi.org/10.1016/s0301-0082(96)00042-1)
- Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience*, *16*(5), 1936-1947. <https://doi.org/10.1523/jneurosci.16-05-01936.1996>
- Moyer, J. T., & Danish, S. F. (2007). Stimulation-Induced Dyskinesias Inform Basal Ganglia Models and the Mechanisms of Deep Brain Stimulation. *Journal of Neuroscience*, *27*(8), 1799-1800. <https://doi.org/10.1523/JNEUROSCI.5359-06.2007>
- Nambu, A., Tokuno, H., Hamada, I., Kita, H., Imanishi, M., Akazawa, T., Ikeuchi, Y., & Hasegawa, N. (2000). Excitatory Cortical Inputs to Pallidal Neurons Via the Subthalamic Nucleus in the Monkey. *Journal of Neurophysiology*, *84*(1), 289-300. <https://doi.org/10.1152/jn.2000.84.1.289>
- Nambu, A., Tokuno, H., & Takada, M. (2002). Functional significance of the cortico-subthalamo-pallidal 'hyperdirect' pathway. *Neuroscience Research*, *43*(2), 111-117. [https://doi.org/10.1016/s0168-0102\(02\)00027-5](https://doi.org/10.1016/s0168-0102(02)00027-5)
- Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, *53*(3), 139-154. <https://doi.org/10.1016/j.jmp.2008.12.005>
- Nougaret, S., Baunez, C., & Ravel, S. (2022). Neurons in the Monkey's Subthalamic Nucleus Differentially Encode Motivation and Effort. *Journal of Neuroscience*, *42*(12), 2539-2551. <https://doi.org/10.1523/JNEUROSCI.0281-21.2021>
- Nutt, J. G., Woodward, W. R., Carter, J. H., & Gancher, S. T. (1992). Effect of long-term therapy on the pharmacodynamics of levodopa. Relation to on-off phenomenon. *Archives of Neurology*, *49*(11), 1123-1130. <https://doi.org/10.1001/archneur.1992.00530350037016>

- O'Doherty, J., Critchley, H., Deichmann, R., & Dolan, R. J. (2003). Dissociating Valence of Outcome from Behavioral Control in Human Orbital and Ventral Prefrontal Cortices. *Journal of Neuroscience*, *23*(21), 7931-7939. <https://doi.org/10.1523/jneurosci.23-21-07931.2003>
- O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable Roles of Ventral and Dorsal Striatum in Instrumental Conditioning. *Science*, *304*(5669), 452-454. <https://doi.org/10.1126/science.1094285>
- Pagonabarraga, J., Kulisevsky, J., Strafella, A. P., & Krack, P. (2015). Apathy in Parkinson's disease: Clinical features, neural substrates, diagnosis, and treatment. *The Lancet Neurology*, *14*(5), 518-531. [https://doi.org/10.1016/s1474-4422\(15\)00019-8](https://doi.org/10.1016/s1474-4422(15)00019-8)
- Palminteri, S., Lebreton, M., Worbe, Y., Grabli, D., Hartmann, A., & Pessiglione, M. (2009). Pharmacological modulation of subliminal learning in Parkinson's and Tourette's syndromes. *Proceedings of the National Academy of Sciences*, *106*(45), 19179-19184. <https://doi.org/10.1073/pnas.0904035106>
- Parent, A., Sato, F., Wu, Y., Gauthier, J., Lévesque, M., & Parent, M. (2000). Organization of the basal ganglia: The importance of axonal collateralization. *Trends in Neurosciences*, *23*(10 Suppl), S20-27. [https://doi.org/10.1016/s1471-1931\(00\)00022-7](https://doi.org/10.1016/s1471-1931(00)00022-7)
- Pavlov, I. P. (1928). *Lectures on conditioned reflexes: Twenty-five years of objective study of the higher nervous activity (behaviour) of animals* (p. 414). Liverwright Publishing Corporation. <https://doi.org/10.1037/11081-000>
- Perdue, B. M., Evans, T. A., Washburn, D. A., Rumbaugh, D. M., & Beran, M. J. (2014). Do monkeys choose to choose? *Learning & Behavior*, *42*(2), 164-175. <https://doi.org/10.3758/s13420-014-0135-0>
- Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., & Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*, *442*(7106), 1042-1045. <https://doi.org/10.1038/nature05051>
- Pont-Sunyer, C., Hotter, A., Gaig, C., Seppi, K., Compta, Y., Katzenschlager, R., Mas, N., Hofeneder, D., Brücke, T., Bayés, A., Wenzel, K., Infante, J., Zach, H., Pirker, W., Posada, I. J., Álvarez, R., Ispierto, L., De Fàbregues, O., Callén, A., ... Tolosa, E. (2015). The onset of nonmotor symptoms in Parkinson's disease (the ONSET PD study). *Movement Disorders: Official Journal of the Movement Disorder Society*, *30*(2), 229-237. <https://doi.org/10.1002/mds.26077>
- Postuma, R. B., & Lang, A. E. (2003). Hemiballism: Revisiting a classic disorder. *The Lancet Neurology*, *2*(11), 661-668. [https://doi.org/10.1016/S1474-4422\(03\)00554-4](https://doi.org/10.1016/S1474-4422(03)00554-4)
- Pote, I., Torkamani, M., Kefalopoulou, Z.-M., Zrinzo, L., Limousin-Dowsey, P., Foltynie, T., Spekenbrink, M., & Jahanshahi, M. (2016). Subthalamic nucleus deep brain stimulation induces impulsive action when patients with Parkinson's disease act under speed pressure. *Experimental Brain Research*, *234*(7), 1837-1848. <https://doi.org/10.1007/s00221-016-4577-9>
- Prange, S., Lin, Z., Nourredine, M., Danaila, T., Laurencin, C., Lagha-Boukbiza, O., Anheim, M., Klinger, H., Longato, N., Phillipps, C., Voirin, J., Polo, G., Simon, E., Mertens, P., Rolland, A.-S., Devos, D., Metereau, E., Tranchant, C., Thobois, S., & Predistim study group. (2022). Limbic Stimulation Drives Mania in STN-DBS in Parkinson

- Disease: A Prospective Study. *Annals of Neurology*, 92(3), 411-417. <https://doi.org/10.1002/ana.26434>
- Reneman, L., Pluijm, M., Schranke, A., & Giessen, E. (2021). Imaging of the dopamine system with focus on pharmacological MRI and neuromelanin imaging. *European Journal of Radiology*, 140, 109752. <https://doi.org/10.1016/j.ejrad.2021.109752>
- Romo, R., & Schultz, W. (1990). Dopamine neurons of the monkey midbrain: Contingencies of responses to active touch during self-initiated arm movements. *Journal of Neurophysiology*, 63(3), 592-606. <https://doi.org/10.1152/jn.1990.63.3.592>
- Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs*, 80(1), 1-28.
- Rouaud, T., Lardeux, S., Panayotis, N., Paleressompoulle, D., Cador, M., & Baunez, C. (2010). Reducing the desire for cocaine with subthalamic nucleus deep brain stimulation. *Proceedings of the National Academy of Sciences of the United States of America*, 107(3), 1196-1200. <https://doi.org/10.1073/pnas.0908189107>
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68-78. <https://doi.org/10.1037/0003-066X.55.1.68>
- Sano, K. (1962). SEDATIVE NEUROSURGERY. *Neurologia medico-chirurgica*, 4(0), 112-142. <https://doi.org/10.2176/nmc.4.112>
- Schönecker, T., Kupsch, A., Kühn, A., Schneider, G.-H., & Hoffmann, K.-T. (2009). Automated Optimization of Subcortical Cerebral MR Imaging-Atlas Coregistration for Improved Postoperative Electrode Localization in Deep Brain Stimulation. *AJNR. American journal of neuroradiology*, 30, 1914-1921. <https://doi.org/10.3174/ajnr.A1741>
- Schultz, W. (1986). Responses of midbrain dopamine neurons to behavioral trigger stimuli in the monkey. *Journal of Neurophysiology*, 56(5), 1439-1461. <https://doi.org/10.1152/jn.1986.56.5.1439>
- Schultz, W. (2000). Multiple reward signals in the brain. *Nature Reviews. Neuroscience*, 1(3), 199-207. <https://doi.org/10.1038/35044563>
- Schultz, W. (2007). Reward. *Scholarpedia*, 2(3), 1652. <https://doi.org/10.4249/scholarpedia.1652>
- Schultz, W., Dayan, P., & Montague, P. R. (1997). A Neural Substrate of Prediction and Reward. *Science*, 275(5306), 1593-1599. <https://doi.org/10.1126/science.275.5306.1593>
- Selemon, L., & Goldman-Rakic, P. (1985). Longitudinal topography and interdigitation of corticostriatal projections in the rhesus monkey. *The Journal of Neuroscience*, 5(3), 776-794. <https://doi.org/10.1523/JNEUROSCI.05-03-00776.1985>
- Seymour, B., Daw, N., Dayan, P., Singer, T., & Dolan, R. (2007). Differential Encoding of Losses and Gains in the Human Striatum. *Journal of Neuroscience*, 27(18), 4826-4831. <https://doi.org/10.1523/JNEUROSCI.0400-07.2007>
- Sharot, T., Shiner, T., & Dolan, R. J. (2010). Experience and Choice Shape Expected Aversive Outcomes. *The Journal of Neuroscience*, 30(27), 9209-9215. <https://doi.org/10.1523/JNEUROSCI.4770-09.2010>

- Shenhav, A., Cohen, J. D., & Botvinick, M. M. (2016). Dorsal anterior cingulate cortex and the value of control. *Nature Neuroscience*, *19*(10), 1286-1291. <https://doi.org/10.1038/nn.4384>
- Singh, S., Lewis, R. L., Barto, A. G., & Sorg, J. (2010). Intrinsically Motivated Reinforcement Learning : An Evolutionary Perspective. *IEEE Transactions on Autonomous Mental Development*, *2*(2), 70-82. <https://doi.org/10.1109/TAMD.2010.2051031>
- Skinner, B. F. (1963). Operant behavior. *American Psychologist*, *18*, 503-515. <https://doi.org/10.1037/h0045185>
- Smith, R. E., Tournier, J.-D., Calamante, F., & Connelly, A. (2013). SIFT : Spherical-deconvolution informed filtering of tractograms. *NeuroImage*, *67*, 298-312. <https://doi.org/10.1016/j.neuroimage.2012.11.049>
- Smith, R. E., Tournier, J.-D., Calamante, F., & Connelly, A. (2015). SIFT2 : Enabling dense quantitative assessment of brain white matter connectivity using streamlines tractography. *NeuroImage*, *119*, 338-351. <https://doi.org/10.1016/j.neuroimage.2015.06.092>
- Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement : Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, *44*(1), 83-98. <https://doi.org/10.1016/j.neuroimage.2008.03.061>
- Spillantini, M. G., Schmidt, M. L., Lee, V. M.-Y., Trojanowski, J. Q., Jakes, R., & Goedert, M. (1997).  $\alpha$ -Synuclein in Lewy bodies. *Nature*, *388*(6645), 839. <https://doi.org/10.1038/42166>
- Sutton, R. S., & Barto, A. (2018). *Reinforcement learning: An introduction* (Second edition). The MIT Press.
- Suzuki, S. (1999). Selection of Forced- and Free-Choice by Monkeys (Macaca Fascicularis). *Perceptual and Motor Skills*, *88*(1), 242-250. <https://doi.org/10.2466/pms.1999.88.1.242>
- Tanaka, S. C., Balleine, B. W., & O'Doherty, J. P. (2008). Calculating Consequences : Brain Systems That Encode the Causal Effects of Actions. *The Journal of Neuroscience*, *28*(26), 6750-6755. <https://doi.org/10.1523/JNEUROSCI.1808-08.2008>
- Temiz, G., Sébille, S. B., Francois, C., Bardinet, E., & Karachi, C. (2019). The anatomofunctional organization of the hyperdirect cortical pathway to the subthalamic area using in vivo structural connectivity imaging in humans. *Brain Structure & Function*. <https://doi.org/10.1007/s00429-019-02012-6>
- Temiz, G., Sébille, S. B., Francois, C., Bardinet, E., & Karachi, C. (2020). The anatomofunctional organization of the hyperdirect cortical pathway to the subthalamic area using in vivo structural connectivity imaging in humans. *Brain Structure and Function*, *225*(2), 551-565. <https://doi.org/10.1007/s00429-019-02012-6>
- Thobois, S., Ardouin, C., Lhommée, E., Klingler, H., Lagrange, C., Xie, J., Fraix, V., Coelho Braga, M. C., Hassani, R., Kistner, A., Juphard, A., Seigneuret, E., Chabardes, S., Mertens, P., Polo, G., Reilhac, A., Costes, N., LeBars, D., Savasta, M., ... Krack, P. (2010). Non-motor dopamine withdrawal syndrome after surgery for Parkinson's disease : Predictors and underlying mesolimbic denervation. *Brain: A Journal of Neurology*, *133*(Pt 4), 1111-1127. <https://doi.org/10.1093/brain/awq032>

- Thobois, S., Hotton, G. R., Pinto, S., Wilkinson, L., Limousin-Dowsey, P., Brooks, D. J., & Jahanshahi, M. (2007). STN Stimulation Alters Pallidal—Frontal Coupling during Response Selection under Competition. *Journal of Cerebral Blood Flow & Metabolism*, *27*(6), 1173-1184. <https://doi.org/10.1038/sj.jcbfm.9600425>
- Tobler, P. N., Fiorillo, C. D., & Schultz, W. (2005). Adaptive coding of reward value by dopamine neurons. *Science (New York, N.Y.)*, *307*(5715), 1642-1645. <https://doi.org/10.1126/science.1105370>
- Tomlinson, C. L., Stowe, R., Patel, S., Rick, C., Gray, R., & Clarke, C. E. (2010). Systematic review of levodopa dose equivalency reporting in Parkinson's disease. *Movement Disorders: Official Journal of the Movement Disorder Society*, *25*(15), 2649-2653. <https://doi.org/10.1002/mds.23429>
- Vinding, M. C., Lindeløv, J. K., Xiao, Y., Chan, R. C. K., & Sørensen, T. A. (2018). *Volition in Prospective Memory: Evidence against differences in recalling free and fixed delayed intentions*. PsyArXiv. <https://doi.org/10.31234/osf.io/hsrbt>
- Voon, V., Fernagut, P.-O., Wickens, J., Baunez, C., Rodriguez, M., Pavon, N., Juncos, J. L., Obeso, J. A., & Bezdard, E. (2009). Chronic dopaminergic stimulation in Parkinson's disease: From dyskinesias to impulse control disorders. *The Lancet Neurology*, *8*(12), 1140-1149. [https://doi.org/10.1016/S1474-4422\(09\)70287-X](https://doi.org/10.1016/S1474-4422(09)70287-X)
- Voon, V., Gao, J., Brezing, C., Symmonds, M., Ekanayake, V., Fernandez, H., Dolan, R. J., & Hallett, M. (2011). Dopamine agonists and risk: Impulse control disorders in Parkinson's disease. *Brain*, *134*(5), 1438-1446. <https://doi.org/10.1093/brain/awr080>
- Voon, V., Krack, P., Lang, A. E., Lozano, A. M., Dujardin, K., Schüpbach, M., D'Ambrosia, J., Thobois, S., Tamma, F., Herzog, J., Speelman, J. D., Samanta, J., Kubu, C., Rossignol, H., Poon, Y.-Y., Saint-Cyr, J. A., Ardouin, C., & Moro, E. (2008). A multicentre study on suicide outcomes following subthalamic stimulation for Parkinson's disease. *Brain*, *131*(10), 2720-2728. <https://doi.org/10.1093/brain/awn214>
- Voon, V., Napier, T. C., Frank, M. J., Sgambato-Faure, V., Grace, A. A., Rodriguez-Oroz, M., Obeso, J., Bezdard, E., & Fernagut, P.-O. (2017). Impulse control disorders and levodopa-induced dyskinesias in Parkinson's disease: An update. *The Lancet Neurology*, *16*(3), 238-250. [https://doi.org/10.1016/s1474-4422\(17\)30004-2](https://doi.org/10.1016/s1474-4422(17)30004-2)
- Voon, V., Pessiglione, M., Brezing, C., Gallea, C., Fernandez, H. H., Dolan, R. J., & Hallett, M. (2010). Mechanisms Underlying Dopamine-Mediated Reward Bias in Compulsive Behaviors. *Neuron*, *65*(1), 135-142. <https://doi.org/10.1016/j.neuron.2009.12.027>
- Voss, S. C., & Homzie, M. J. (1970). Choice as a Value. *Psychological Reports*, *26*(3), 912-914. <https://doi.org/10.2466/pr0.1970.26.3.912>
- Wang, K. S., & Delgado, M. R. (2019). Corticostriatal Circuits Encode the Subjective Value of Perceived Control. *Cerebral Cortex*, *29*(12), 5049-5060. <https://doi.org/10.1093/cercor/bhz045>
- Wang, Q., Akram, H., Muthuraman, M., Gonzalez-Escamilla, G., Sheth, S. A., Oxenford, S., Yeh, F.-C., Groppa, S., Vanegas-Aroyave, N., Zrinzo, L., Li, N., Kühn, A., & Horn, A. (2021). Normative vs. Patient-specific brain connectivity in deep brain

- stimulation. *NeuroImage*, *224*, 117307.  
<https://doi.org/10.1016/j.neuroimage.2020.117307>
- Weintraub, D., Koester, J., Potenza, M. N., Siderowf, A. D., Stacy, M., Voon, V., Whetteckey, J., Wunderlich, G. R., & Lang, A. E. (2010a). Impulse control disorders in Parkinson disease: A cross-sectional study of 3090 patients. *Archives of Neurology*, *67*(5), 589-595. <https://doi.org/10.1001/archneurol.2010.65>
- Weintraub, D., Koester, J., Potenza, M. N., Siderowf, A. D., Stacy, M., Voon, V., Whetteckey, J., Wunderlich, G. R., & Lang, A. E. (2010b). Impulse control disorders in Parkinson disease: A cross-sectional study of 3090 patients. *Archives of Neurology*, *67*(5), 589-595. <https://doi.org/10.1001/archneurol.2010.65>
- Weintraub, D., & Mamikonyan, E. (2019). Impulse Control Disorders in Parkinson's Disease. *The American Journal of Psychiatry*, *176*(1), 5-11. <https://doi.org/10.1176/appi.ajp.2018.18040465>
- Welter, M.-L., Schüpbach, M., Czernecki, V., Karachi, C., Fernandez-Vidal, S., Golmard, J.-L., Serra, G., Navarro, S., Welaratne, A., Hartmann, A., Mesnage, V., Pineau, F., Cornu, P., Pidoux, B., Worbe, Y., Zikos, P., Grabli, D., Galanaud, D., Bonnet, A.-M., ... Agid, Y. (2014). Optimal target localization for subthalamic stimulation in patients with Parkinson disease. *Neurology*, *82*(15), 1352-1361. <https://doi.org/10.1212/wnl.0000000000000315>
- Wenke, D., Fleming, S. M., & Haggard, P. (2010). Subliminal priming of actions influences sense of control over effects of action. *Cognition*, *115*(1), 26-38. <https://doi.org/10.1016/j.cognition.2009.10.016>
- Williams, S. M., & Goldman-Rakic, P. S. (1998). Widespread origin of the primate mesofrontal dopamine system. *Cerebral Cortex (New York, N.Y.: 1991)*, *8*(4), 321-345. <https://doi.org/10.1093/cercor/8.4.321>
- Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., & Nichols, T. E. (2014). Permutation inference for the general linear model. *NeuroImage*, *92*, 381-397. <https://doi.org/10.1016/j.neuroimage.2014.01.060>
- Witt, K., Pulkowski, U., Herzog, J., Lorenz, D., Hamel, W., Deuschl, G., & Krack, P. (2004). Deep Brain Stimulation of the Subthalamic Nucleus Improves Cognitive Flexibility but Impairs Response Inhibition in Parkinson Disease. *Archives of Neurology*, *61*(5), 697. <https://doi.org/10.1001/archneur.61.5.697>
- Yelnik, J. (2002). Functional anatomy of the basal ganglia. *Movement Disorders*, *17*(S3), S15-S21. <https://doi.org/10.1002/mds.10138>
- Yelnik, J., Bardinet, E., Dormont, D., Malandain, G., Ourselin, S., Tandé, D., Karachi, C., Ayache, N., Cornu, P., & Agid, Y. (2007). A three-dimensional, histological and deformable atlas of the human basal ganglia. I. Atlas construction based on immunohistochemical and MRI data. *NeuroImage*, *34*(2), 618-638. <https://doi.org/10.1016/j.neuroimage.2006.09.026>
- Zénon, A., Duclos, Y., Carron, R., Witjas, T., Baunez, C., Régis, J., Azulay, J.-P., Brown, P., & Eusebio, A. (2016). The human subthalamic nucleus encodes the subjective value of reward and the cost of effort during decision-making. *Brain: A Journal of Neurology*, *139*(Pt 6), 1830-1843. <https://doi.org/10.1093/brain/aww075>

Zorowitz, S., Momennejad, I., & Daw, N. D. (2020). Anxiety, Avoidance, and Sequential Evaluation. *Computational Psychiatry*, ~~4~~(0), 1-17.  
[https://doi.org/10.1162/CPSY\\_a\\_00026](https://doi.org/10.1162/CPSY_a_00026)