

# SOMMAIRE

---

<b>REMERCIEMENTS.....</b>	<b>1</b>
<b>RESUME.....</b>	<b>7</b>
<b>ABSTRACT .....</b>	<b>9</b>
<b>SOMMAIRE .....</b>	<b>11</b>
<b>PREFACE .....</b>	<b>15</b>
<b>INTRODUCTION.....</b>	<b>17</b>
<b>ETAT DE L'ART .....</b>	<b>21</b>
CHAPITRE 1. LES COMPORTEMENTS ADAPTATIFS : DIVERSITE, COMPLEXITE ET EVOLUTION D'UN REPERTOIRE COMPORTEMENTAL ESSENTIEL A LA SURVIE DES ORGANISMES VIVANTS .....	23
1.    Un spectre large d'études et de résultats .....	24
1.1.    Un point vocabulaire .....	24
1.1.1.    Des représentations collectives à la science .....	24
1.1.2.    Etude raisonnée de la littérature .....	24
1.1.3.    Enjeux scientifiques : identifier les concepts d'étude .....	25
1.2.    Flexibilité comportementale et comportements adaptatifs : de l'écologie à la psychologie.....	28
1.2.1.    Aux racines de la flexibilité comportementale .....	28
1.2.2.    De la flexibilité comportementale aux comportements adaptatifs .....	30
1.2.3.    Enjeux scientifiques : choisir son champ disciplinaire.....	32
1.3.    Flexibilité cognitive, flexibilité comportementale.....	32
1.3.1.    De la psychologie expérimentale à la psychologie cognitive .....	32
1.3.2.    Dissociabilité des flexibilités cognitive et comportementale .....	34
1.3.3.    Enjeux scientifiques : établir des explications mécanistiques à partir d'observations comportementales.....	35
2.    Etude expérimentale des comportements adaptatifs.....	37
2.1.    Comportements naturels ou artificiels ? .....	37
2.1.1.    Méthodes d'étude du comportement : de la nature au laboratoire.....	37
2.1.2.    Validité et intérêt scientifique des jeux de données collectés.....	38
2.1.3.    Enjeux scientifiques : contrôler écologiquement ses expériences au laboratoire.....	39
2.2.    Protocoles expérimentaux classiques d'étude des comportements adaptatifs en cas de changement abrupt de l'environnement .....	40
2.2.1.    Validité écologique et intérêt scientifique de la question .....	40
2.2.2.    Tâche comportementale de renversement d'apprentissage : protocole historique d'étude du comportement face à des changements abrupts .....	41
2.2.3.    Enjeux scientifiques : développer une nouvelle tâche comportementale translationnelle afin de dépasser les limites de la tâche de renversement d'apprentissage ...	44
CHAPITRE 2. EXPLORER ET/OU EXPLOITER, TELLE EST LA QUESTION : DE L'APPRENTISSAGE AUX INFERENCES COMME SUPPORTS DES COMPORTEMENTS ADAPTATIFS.....	47
1.    Explorer, exploiter : mise en œuvre comportementale d'un processus cognitif d'apprentissage.....	48
1.1.    Le dilemme exploration – exploitation.....	48

1.1.1.	Le dilemme du fourrageur .....	48
1.1.2.	Les déterminants de la balance exploration - exploitation .....	50
1.1.3.	Enjeux scientifiques : la balance exploration – exploitation face à la multiplicité et la complémentarité de différentes formes d’exploration.....	52
1.2.	L’apprentissage par renforcement et la balance exploitation - exploration.....	53
1.2.1.	L’apprentissage par renforcement : une boucle de rétrocontrôle décisionnelle .....	53
1.2.2.	Considération mathématique de l’apprentissage par renforcement : des paramètres computationnels aux comportements .....	54
1.2.3.	Enjeux scientifiques : quid de l’apprentissage par renforcement dans le cadre d’un environnement volatil ?.....	57
2.	Explorer et exploiter efficacement : quand l’adaptation comportementale requiert des processus cognitifs au-delà du simple apprentissage par renforcement .....	59
2.1.	Le méta-apprentissage ou l’ajustement dynamique des paramètres d’apprentissage....	59
2.1.1.	Des paramètres aux méta-paramètres : les déterminants critiques du comportement	59
2.1.2.	Facteurs d’ajustement des méta-paramètres en environnement volatil : entre importance de l’historique des récompenses et cadre bayésien.....	61
2.1.3.	Enjeux scientifiques : quid d’une estimation partielle des paramètres de son environnement ? .....	66
2.2.	« Tu as assez d’intelligence pour deviner le reste, par ce que je te laisse entrevoir »....	66
2.2.1.	L’environnement est-il prévisible ? Estimer le prévisible, considerer l'imprévisible	67
2.2.2.	Inférer le changement et s'y adapter.....	68
2.2.3.	Enjeux scientifiques : comment réconcilier ce foisonnement de théories ? .....	72
<b>CHAPITRE 3. LA NEUROBIOLOGIE DES COMPORTEMENTS ADAPTATIFS : DES NEURONES AU COMPORTEMENT.....</b>		73
1.	Etude de la neurobiologie des comportements adaptatifs : corrélation n’est pas cause .....	74
1.1.	Techniques d’étude descriptives de l’activité cérébrale liée à un comportement.....	74
1.1.1.	Observer l’activité cérébrale.....	74
1.1.2.	Enregistrer simultanément comportement et activité cérébrale.....	75
1.1.3.	Enjeux scientifiques : comment lier activité cérébrale et comportement ? .....	76
1.2.	Techniques d’étude causatives de l’activité cérébrale liée à un comportement .....	77
1.2.1.	Les cas cliniques : ce que nous apprennent les études de patients .....	77
1.2.2.	Manipuler le comportement en contrôlant les neurones : la neuromodulation expérimentale .....	77
1.2.3.	Enjeux scientifiques : choisir la technique de neuromodulation la plus adaptée et en connaître ses limites .....	81
2.	Les régions cérébrales et circuits impliqués dans les comportements adaptatifs .....	83
2.1.	Le cortex cingulaire antérieur dorsal face à la volatilité environnementale .....	83
2.1.1.	Le cortex cingulaire antérieur dorsal : une région cérébrale aux multiples fonctions	83
2.1.2.	Le cortex cingulaire antérieur dorsal : région clé de l’adaptation comportementale.	84
2.1.3.	Enjeux scientifiques : cibler le cortex cingulaire antérieur dorsal comme première approche pour l’étude des bases neurobiologiques des comportements adaptatifs .....	88
2.2.	Validité translationnelle des substrats neurobiologiques.....	89
2.2.1.	Etablir des homologies de structures cérébrales.....	89
2.2.2.	Le cortex cingulaire antérieur dorsal : sous-structure du cortex préfrontal .....	89
2.2.3.	Enjeux scientifiques : choisir au mieux les coordonées stéréotaxiques du cortex cingulaire antérieur dorsal chez la souris .....	92

3. Au-delà des circuits : la flexibilité des circuits neuronaux comme support des comportements adaptatifs .....	94
3.1. La neuromodulation : processus physiologique de plasticité des circuits neuronaux et des comportements .....	94
3.1.1. La neuromodulation comme support des modifications de l'information nerveuse .	94
3.1.2. Les grands circuits neuromodulateurs .....	95
3.1.3. Enjeux scientifiques : faire face à la complexité de la neuromodulation .....	96
3.2. Neuromodulation, modèles computationnels et comportements adaptatifs .....	97
3.2.1. Méta-apprentissage et neuromodulation.....	97
3.2.2. Inférences, transitions abruptes et neuromodulation .....	99
3.2.3. Enjeux scientifiques : investiguer l'implication de la neuromodulation dans un nouveau paradigme experimental .....	100
<b>OBJECTIFS ET HYPOTHESES DE TRAVAIL .....</b>	<b>103</b>
<b>TRAVAIL EXPERIMENTAL .....</b>	<b>107</b>
CHAPITRE 1. DEVELOPPEMENT D'UN SYSTEME OPEN-SOURCE AUTOMATISE ET A HAUT RENDEMENT POUR L'ETUDE DES COMPORTEMENTS ANIMAUX COMPLEXES SUR DE LONGUES PERIODES D'ETUDE : LE PROJET BEATBOX .....	109
1. Résumé introductif .....	109
1.1. Historique et enjeux de l'étude.....	109
1.2. La BEATBox : une nouvelle chambre automatisée open-source à haut rendement d'acquisition de données .....	112
2. Développement matériel, logiciel et validation expérimentale : le projet BEATBox .....	113
Article 1 .....	115
CHAPITRE 2. ADAPTATION DE LA SOURIS EN ENVIRONNEMENT VOLATIL : INVESTIGATION DES STRATEGIES COMPORTEMENTALES MISES EN JEU ET DES MECANISMES COMPUTATIONNELS SOUS-JACENTS .....	155
1. Résumé introductif .....	155
1.1. Enjeux de l'étude.....	155
1.2. La tâche de switch : une nouvelle tâche de psychologie expérimentale pour investiguer les mécanismes computationnels des comportements adaptatifs chez la souris.....	156
2. Développement d'une nouvelle tâche comportementale, validation expérimentale et identification des stratégies comportementales et mécanismes computationnels de l'adaptation de la souris en environnement volatil : le projet SWITCH .....	158
Article 2 .....	159
CHAPITRE 3. INVESTIGATION NEUROMODULATRICE DE L'IMPLICATION DU CINGULAIRE MURIN DANS L'ADAPTATION AUX CHANGEMENTS ABRUPTS DE L'ENVIRONNEMENT .....	195
1. Introduction .....	195
2. Méthodes .....	197
3. Résultats comportementaux préliminaires .....	203
4. Discussion .....	208
<b>CONCLUSIONS, PERSPECTIVES ET DISCUSSION GENERALE .....</b>	<b>213</b>
1. « Des souris et des hommes » ou la question de la recherche translationnelle.....	214
1.1. Le choix de la souris comme modèle d'étude .....	214
1.2. Souris versus humains : des stratégies différentes pour s'adapter en milieu volatil....	215

2.	« Le rat de ville et le rat de laboratoire » ou la question de la disparité des animaux étudiés	
	217	
2.1.	Animal « naturel » ou animal « artificiel » ?.....	217
2.2.	Tirer profit des systèmes experimentaux automatisés afin de réaliser des études comportementales à grande échelle.....	220
3.	« Si le laboratoire ne va pas à la nature, la nature ira au laboratoire » .....	221
3.1.	Développer des systèmes experimentaux plus écologiques .....	221
3.2.	Enjeux et avantages de jeux de données complexes et bruités.....	222
4.	« Le chat qui apprit à une mouette à voler » ou la question du choix des paradigmes expérimentaux .....	223
4.1.	Les paradigmes expérimentaux face à l'écologie de l'espèce étudiée.....	223
4.2.	Les big data au service des paradigmes experimentaux .....	225
<b>BIBLIOGRAPHIE</b>	.....	<b>227</b>
<b>ANNEXES</b>	.....	<b>243</b>
1.	Autres publications.....	245
2.	Congrès.....	245
3.	Médiation scientifique .....	245
4.	Activités d'enseignement .....	245
5.	Formations.....	246
<b>POSTFACE</b>	.....	<b>247</b>

# PREFACE

---

« Je suis plus savant que cet homme-là. En effet, il est à craindre que nous ne sachions ni l'un ni l'autre rien qui vaille la peine, mais, tandis que, lui, il s'imagine qu'il sait quelque chose alors qu'il ne sait rien, moi qui effectivement ne sais rien, je ne vais pas m'imaginer que je sais quelque chose. »

*Apologie de Socrate 21d, Platon, Livre de Poche*

J'ai commencé ce travail thèse il y a plus de quatre ans maintenant. J'ai commencé ce travail en ne connaissant ni le mot Arduino, ni l'équation de l'erreur de prédiction de la récompense, ni les rudiments de la programmation. J'ai commencé ce travail avec mille questions et l'émerveillement candide de la jeune chercheuse rêvant secrètement de pouvoir y répondre.

Quatre ans plus tard, après plusieurs dizaines de courts-circuits, plusieurs milliers de lignes de codes Matlab, plusieurs millions de données acquises et plusieurs centaines de pages écrites, je n'ai répondu qu'à une infime partie de mes interrogations, et des milliers de nouvelles se bousculent désormais dans ma tête.

Ce manuscrit est le reflet de ces questionnements qui m'ont habitée tout au long de ces quatre années. Comme Socrate qui suscite la mise en forme de pensées confuses par le dialogue, je me suis employée à appliquer la sage méthode de la maïeutique à mes données. Travail de longue haleine dont l'aboutissement final n'est certainement pas le sujet de ma thèse mais le début j'espère, d'une aventure scientifique palpitante pour les années à venir.

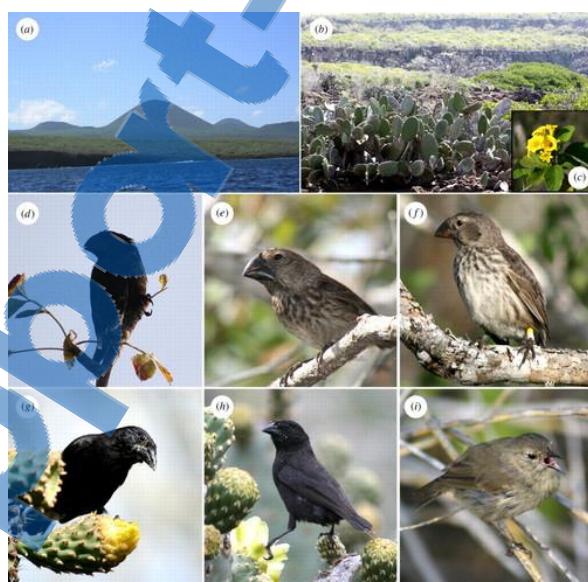
Aujourd'hui, je pense avoir perdu l'exaltation candide de mes débuts. Même si mon cheminement futur s'éloignera probablement de la recherche académique, j'en conserverai un souvenir vif et sage : plus que l'espoir de trouver le résultat révolutionnaire qui défrayera la chronique des grands journaux, c'est la conscience même de l'ignorance qui est peut-être en définitive la clé d'une recherche féconde.



## INTRODUCTION

« *La vie est l'adaptation continue de relations internes à des relations externes* ». Voilà le postulat d'Herbert Spencer dans son ouvrage *Qu'est-ce que la morale ?* (Spencer and Desclos, 1909). Cette assertion, évoque ici le vivant comme un système plastique en immersion dans un environnement face auquel il doit modifier en permanence ses caractéristiques propres afin de survivre. En lisant ce propos, il ne faut pas oublier qu'Herbert Spencer était un sociologue et philosophe anglais, fervent défenseur des théories de l'évolution. En ce sens, discuter des processus adaptatifs revient à discuter des théories de l'évolution telles que formalisées par Charles Darwin (Darwin, 2012).

En biologie évolutive, l'adaptation est en effet définie comme une sélection et un maintien de caractéristiques anatomiques, physiologiques ou comportementales qui procurent à un individu d'une espèce donnée des avantages en termes de survie individuelle et de survie de son espèce dans un environnement donné. Un exemple classique est celui des pinsons de Darwin ou pinsons des Galapagos (Abzhanov, 2010). Les îles Galapagos sont connues pour leur diversité de paysages et donc de ressources alimentaires. Chacun des pinsons observés par Darwin est issu d'une sélection directionnelle et radiative contrainte sur la taille et la forme du bec en fonction des ressources alimentaires à disposition.



**Figure 1 : Adaptation radiative des pinsons de Darwin.** En fonction de leur lieu de vie et des ressources alimentaires à disposition, les pinsons des îles Galapagos présentent des becs de tailles et de formes différentes. (a) L'île Floreana, une des îles volcaniques des Galapagos. (b) Cactus épineux (*Opuntia helleri*) de l'île Genovesa. (c) Fleurs de cordial jaune (*Cordia lutea*). (d) Male *Geospiza magnirostris*. (e) Femelle *G. magnirostris* de l'île Genovesa. (f) Femelle *G. fortis* de l'île Santa Cruz. (g) Male *G. conirostris*. (h) Mâle *G. difficilis*. (i) Mâle *Certhidea fusca*. (Abzhanov, 2010)

Les processus adaptatifs sont dans le cadre de l'évolution des processus assurant un *ajustement fonctionnel passif* des êtres vivants à des milieux relativement stables sur des temps longs. Or ce type d'adaptation serait insuffisant pour expliquer l'aptitude du vivant à se maintenir lorsque des changements surviennent, comme par exemple la disparition d'une source de nourriture suite à sa consommation par des compétiteurs alimentaires. Seule une *réponse active rapide*, à l'échelle d'un individu ou d'un groupe social, permet de répondre à ces modifications garantissant ainsi l'adaptation au changement : c'est ce que l'on appelle la *plasticité phénotypique comportementale* (Wong and Candolin, 2015).

L'environnement dans lequel évolue le vivant est complexe, incertain et une de ces sources d'incertitudes provient des changements permanents qui s'y opèrent. Ces changements peuvent être brutaux ou progressifs, prédictibles ou imprédictibles, facilement détectables ou non, très fréquents ou peu fréquents. L'Histoire regorge d'exemples de changements qui ont modifié l'environnement de manière brutale (déforestation massive pour installer des infrastructures humaines) ou qui le modifient de manière progressive (réchauffement climatique) (Wong and Candolin, 2015). Pourtant le vivant subsiste. Tous les êtres vivants en effet présentent des comportements *adaptés*, au sens où ces comportements répondent aux contraintes de l'environnement, et *adaptatifs/flexibles*, au sens où les stratégies comportementales sont susceptibles de changer lorsque l'environnement se modifie.

L'investigation de ces mécanismes adaptatifs est aujourd'hui une question fondamentale en neurosciences afin de comprendre le vivant. Néanmoins, la complexité des problèmes auxquels celui-ci fait face rend toute tentative de réponse ardue. Il convient donc dans un premier temps d'aborder la question sous un angle *phénoménologique*, c'est-à-dire en décrivant et analysant les manifestations visibles de ce que l'on étudie (Sheets-Johnstone, 2015). L'étude du comportement est donc une approche permettant de collecter un certain nombre de données observables, analysables quantitativement et qualitativement, garantissant ainsi un premier point d'entrée pour l'étude de questions complexes.

Le comportement consiste en l'ensemble des actions et réactions directement observables dans un contexte espace-temps donné. En estimant que les êtres vivants sont des systèmes intentionnels (Dennett and The Journal of Philosophy, Inc., 1971), je considérerai le comportement comme le moyen d'atteindre un/des *objectif(s)* poursuivi(s)<sup>1</sup>.

Classiquement, on distingue les comportements *indépendants de l'apprentissage* ou *canalisés* (Staddon, 2016) (longtemps appelés comportements *innés* ou *instinctifs*), des

---

<sup>1</sup> En référence à la dichotomie *goal-directed* (comportements dirigés vers un but, un objectif) versus *habits* (comportements habituels, routines)

comportements *dépendants de l'apprentissage* (appelés auparavant comportements *apris*). Les comportements *canalisés* sont des comportements présents chez l'ensemble des individus d'une espèce, stéréotypés et peu sensibles à la modification par l'expérience comme le soin parental aux petits (chez les mammifères ou les oiseaux). Les comportements *dépendants de l'apprentissage* quant à eux sont des comportements dont la mise en œuvre dépend de l'expérience passée et ce sont eux qui supportent la plasticité dynamique du vivant : ce sont ces comportements qui sont mis en jeu en cas de changement de l'environnement du fait de la discordance entre passé et présent.

Dans cette perspective, lier comportement et apprentissage n'est pas anodin. Il est en effet essentiel de rappeler que le comportement, chez tous les êtres vivants dotés d'un système nerveux, n'est que la résultante visible de processus neuronaux complexes. Les neurones, ces *unités fonctionnelles* du système nerveux, assurent en continu et quelle que soit leur appartenance à une région cérébrale donnée, l'intégration et le traitement d'informations externes et internes (au sens *input*) générant ou non des sorties (au sens *output*). La manière dont l'activité (ou l'absence d'activité) est intégrée, à l'échelle d'une population de neurones, d'une ou plusieurs aires cérébrales est au cœur d'un vaste champ de recherche. L'ensemble de ces processus intégratifs peut être envisagé sous forme computationnelle et leur compréhension peut apporter des informations précieuses afin de comprendre les processus cognitifs et leurs résultantes comportementales (Bayne et al., 2019; Genon et al., 2018).

Dans la ligne directrice de ces considérations, mon travail de thèse s'est orienté sur l'étude de la capacité et de la vitesse des organismes à détecter les changements abrupts et non prévisibles de l'environnement en évaluant :

- 1) La nature et l'efficacité des stratégies mises en œuvre pour s'adapter au nouvel environnement en considérant cette question d'un point de vue comportemental (quels sont les comportements observés ?) et cognitif (quels sont les algorithmes ou modèles expliquant les résultats observés ?)
- 2) La nature des substrats neuronaux responsables de la computation à l'origine des *comportements adaptatifs*.

Pour introduire mon travail, au cours de l'étude de la littérature à suivre, je reviendrai dans un premier temps plus en détail sur la notion de *comportements adaptatifs* et notamment sur l'historique du concept et de la perception de celui-ci par la communauté scientifique. Je

## *INTRODUCTION*

présenterai ensuite dans quelle mesure les processus d'apprentissage et d'ajustement de l'apprentissage sont essentiels à la mise en place des *comportements adaptatifs* précédemment discutés. Au cours d'une troisième partie, je présenterai les régions cérébrales ainsi et les catégories de neurones supposés être impliqués dans ces processus ainsi que la manière de les étudier. Tout au long de cet état de l'art, je discuterai des enjeux scientifiques que comprend ce sujet d'étude afin de formaliser au fur et à mesure les questions auxquelles mon travail tente de répondre. Je finirai par le développement détaillé des objectifs de mon travail expérimental avant de le présenter puis de le discuter.

## **ETAT DE L'ART**

---



# CHAPITRE 1. Les comportements adaptatifs : diversité, complexité et évolution d'un répertoire comportemental essentiel à la survie des organismes vivants

---

## INTRODUCTION

« *La définition fait connaître ce qu'est la chose* ». La science ne doit pas s'affranchir de ce principe énoncé par Aristote dans son ouvrage *Les seconds analytiques*. D'autant plus si les notions étudiées sont aussi complexes que les *comportements adaptatifs* ou *l'adaptation* souvent mentionnés en neurosciences sous l'expression de *flexibilité comportementale*. Ce concept peut pourtant paraître intuitif de prime abord : si l'on est *flexible* d'un point de vue comportemental, il est très probable que nos comportements se modifient en réponse à un changement dans notre environnement. Néanmoins, il suffit de se documenter extensivement et se plonger dans deux revues approfondies sur le sujet (Audet and Lefebvre, 2017; Lea et al., 2020) pour envisager la complexité du champ d'étude de la *flexibilité comportementale*.

Ethologie, psychobiologie, neurosciences cognitives... Chaque champ disciplinaire qui s'intéresse à ce sujet repose sur des définitions implicites différentes qui ne sont pas toujours mutuellement compatibles. Certes, les conclusions sont toujours plus intéressantes les unes que les autres mais l'hétérogénéité résultante du concept rend son étude complexe. Comme le soulignent les auteurs :

“This rich diversity of behavioral investigations is useful, as it provides a detailed picture of how behaviors are modified under changing conditions. However, there is little chance that all these cases share a similar etiology. Therefore, referring to this huge diversity of traits under the same blanket term is problematic.”(Audet and Lefebvre, 2017)

“Unsurprisingly, given the extent to which the term is used, authors are not consistent in what they mean by it”. (Lea et al., 2020)

Dans le cadre de ma thèse, travailler sur ce sujet nécessite donc une définition claire du cadre conceptuel dans lequel l'étude des *comportements adaptatifs* et de la *flexibilité comportementale* sera réalisée. Ce premier chapitre tâche ainsi de faire un état des lieux de la littérature sur la notion de *flexibilité* en neurosciences et de définir le cadre d'étude des *comportements adaptatifs* au sein duquel j'ai travaillé.

## 1. UN SPECTRE LARGE D'ETUDES ET DE RESULTATS

### 1.1. UN POINT VOCABULAIRE

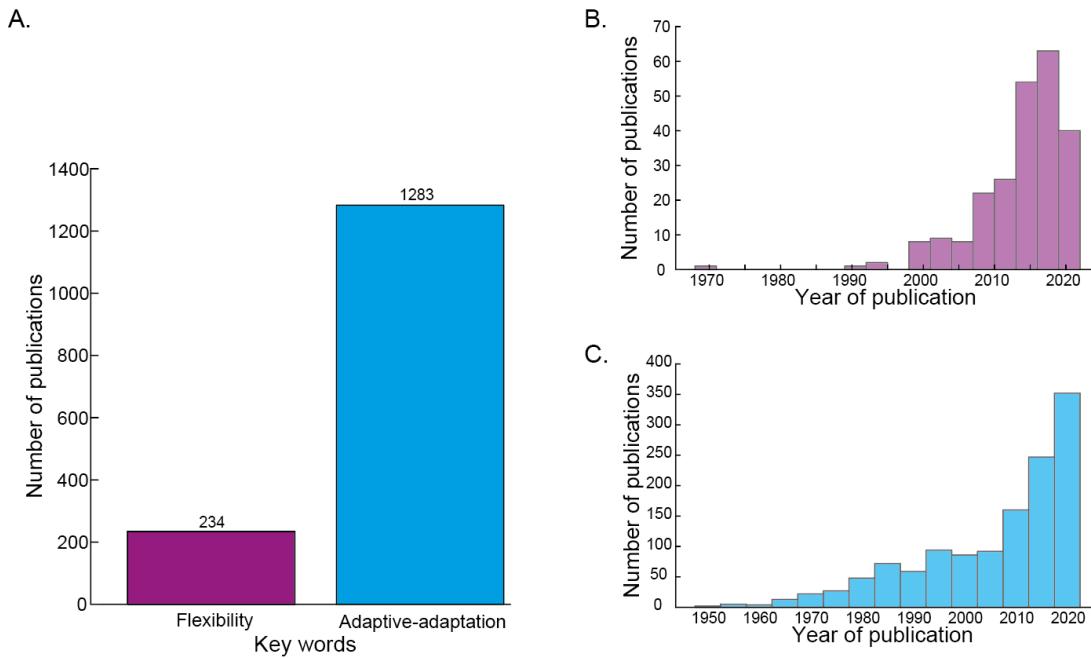
#### 1.1.1. DES REPRESENTATIONS COLLECTIVES A LA SCIENCE

A première vue, discuter de *flexibilité comportementale* ou de *comportements adaptatifs* semble aisément. Les habitants des grandes villes comme Paris, par exemple, savent à quel point les rats, nos petits colocataires citadins, sont capables de s'adapter à tout, et de faire face à toutes les difficultés. En témoignent les campagnes successives d'installation de nouvelles poubelles dans la capitale : années après années les designs ont été modifiés, toujours plus élaborés et complexes afin de lutter contre le chapardage murin. Et pourtant, le dernier modèle en date de 2019, pensé et conçu pour être infaillible, a été vite investi par les rongeurs (20 minutes, M.F, 2019). La vie de tous les jours regorge ainsi de multiples exemples illustrant cette représentation collective que sont les *comportements adaptatifs* ou la *flexibilité comportementale*, mais la question ici est d'en comprendre précisément les bases neuroscientifiques.

#### 1.1.2. ETUDE RAISONNÉE DE LA LITTERATURE

Le point de départ de ce travail bibliographique a été d'évaluer de manière objective dans la littérature l'ensemble des études qui se sont intéressées à la *flexibilité comportementale*, et aux *comportements adaptatifs*.

Dans un premier temps, j'ai défini comme termes de recherche une liste de mots clés centrés autour de la *flexibilité comportementale* (behavioral/behavioural flexibility), et des *comportements adaptatifs* (adaptive behaviour(s)/behavior(s) – behavioral/behavioural adaptation(s)). Puis, j'ai développé via Matlab un algorithme de recherche automatique sur Pubmed de l'ensemble des articles comportant ces termes dans leur titre. Cet algorithme m'a permis de récupérer sous format csv l'ensemble des études et leurs informations relatives (date de publication, auteurs, journal de publication, titre). Le nombre d'études ainsi trouvées tout comme l'évolution du nombre de publications par an se trouvent en **figure 1**.



**Figure 1 : Etat de la littérature sur le sujet de la flexibilité comportementale et des comportements adaptatifs.** A. Nombre de publications identifiées par l'algorithme sur ces thématiques. B. Evolution du nombre de publications par années sur le sujet de la *flexibilité comportementale*. C. Evolution du nombre de publications par années sur le sujet des *comportements adaptatifs*.

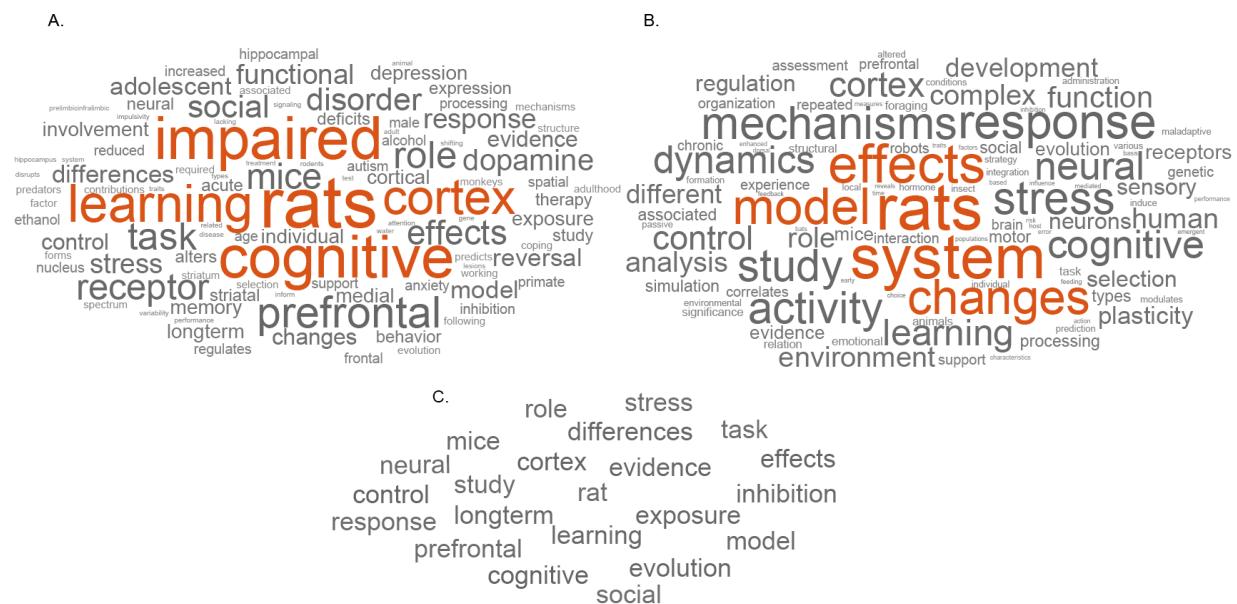
Sur l'échantillon d'articles recensés par cette méthode, il est possible de noter que les premières études s'intéressant aux *comportements adaptatifs* sont un tout petit peu plus anciennes (Pegel'man, 1966) que les premières études discutant de *flexibilité comportementale* (Poirier, 1969). Il est également intéressant de remarquer que le nombre de publications concernant les *comportements adaptatifs* est 5,5 fois plus grand que celui concernant la *flexibilité comportementale* alors que les deux notions semblent pourtant liées. Ceci peut donc laisser penser que l'emploi de ces deux expressions est réalisé par des communautés scientifiques différentes.

Au vu du nombre conséquent de publications identifiées comme potentiellement d'intérêt, il aurait été clairement impossible de tout lire. J'ai donc procédé à une analyse sémantique grâce à un algorithme Matlab que j'ai développé pour déterminer les occurrences de mots-clés afin d'identifier des thématiques récurrentes de recherches et de discuter précisément des notions qui m'intéresseront dans le cadre de ma thèse.

### 1.1.3. ENJEUX SCIENTIFIQUES : IDENTIFIER LES CONCEPTS D'ETUDE

L'analyse sémantique des titres des études portant sur la *flexibilité comportementale* et les *comportements adaptatifs* m'a permis de générer deux nuages de mots-clés associés à

chacune de ces expressions. L'analyse de ceux-ci a révélé une grande disparité dans les termes les plus couramment associés à ces expressions : celles-ci ne sont donc effectivement pas employées par les mêmes communautés scientifiques. L'observation des mots-clés avec la plus grande occurrence a notamment indiqué que les publications concernant les *comportements adaptatifs* sont majoritairement cliniques et dédiées à l'étude des troubles autistiques ou autres troubles (neuro)développementaux ainsi que des déficiences et handicaps causés par l'âge ou certains accidents. Ces études n'étant pas pertinentes dans le cadre de ma thèse, j'ai développé un nouvel algorithme Matlab afin de filtrer ces études, sur la base de mots-clés et de journaux à exclure. Cette étape de filtre m'a permis d'éliminer près de 60% des études. J'ai effectué par la suite une nouvelle analyse sémantique, me permettant également de réaliser une étude croisée des mots-clés retrouvés à l'identique entre les nuages de mots-clés concernant sur *la flexibilité comportementale* et les *comportements adaptatifs* (**Figure 2**).



**Figure 2 : Analyse sémantique des titres de publications sur le sujet de la flexibilité comportementale et des comportements adaptatifs.** A. Nuage de mots obtenus à partir des titres des articles sur la *flexibilité comportementale*. B. Nuage de mots obtenus à partir des titres des articles filtrés sur les *comportements adaptatifs*. C. Mots-clés présents à la fois dans les titres des études sur la *flexibilité comportementale* et les *comportements adaptatifs*.

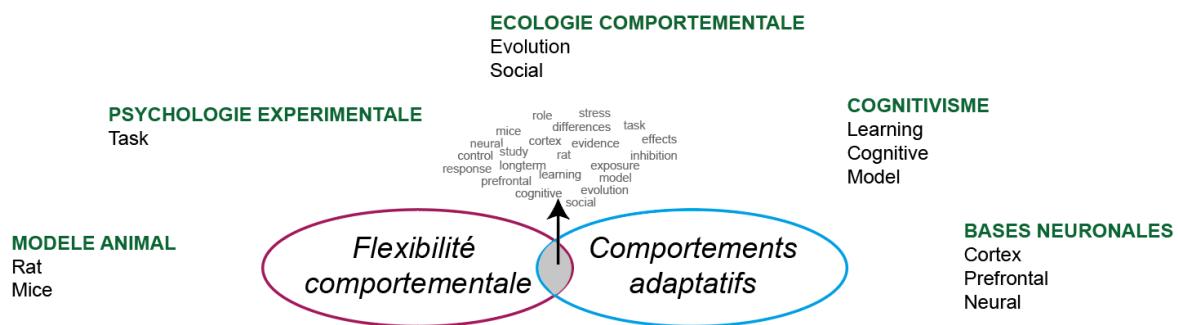
Même si les nouveaux nuages de mots semblent révéler une cohérence plus importante qu'avant filtrage, seule une petite vingtaine de mots-clés apparaît commune entre les titres des deux recherches. Cette étude sémantique révèle que deux expressions qui semblent pourtant proches cachent des définitions différentes en fonction des domaines scientifiques d'étude, voire des équipes de recherche qui s'intéressent à ces sujets. Etudier la *flexibilité*

Stratégies d'adaptation de la souris face à un environnement volatil

*ETAT DE L'ART. Chapitre 1. Flexibilité comportementale et comportements adaptatifs*

*comportementale* et/ou les *comportements adaptatifs* n'est donc pas un sujet aussi trivial qu'il y paraît à première vue et il est donc nécessaire de bien définir son sujet d'étude.

L'analyse croisée des mots-clés retrouvés à l'identique entre les titres des publications sur la *flexibilité comportementale* et les *comportements adaptatifs* est néanmoins très intéressante car elle met en lumière les concepts et les notions importantes qui semblent être associées aux deux expressions. J'ai pu alors établir des sous-catégories de mots (champs lexicaux) permettant d'organiser les concepts et de mettre en lumière les aspects communs auxquels s'intéressent les différents champs disciplinaires. (**Figure 3**). Les mots-clés génériques tels que « effects », « inhibition », « exposure », « response », « study », « control », « evidence », « role » n'ont pas été pris en compte.



**Figure 3 : Champs lexicaux définis à partir de l'analyse sémantique croisée des titres des publications sur la flexibilité cognitive et les comportements adaptatifs.** Plusieurs catégories apparaissent à la suite de cette analyse : les modèles expérimentaux (animaux), les domaines scientifiques (écologie, psychologie...) ainsi que les différents niveaux d'étude (cellulaire, réseau neuronal, aires cérébrales, comportement).

Cette catégorisation fine met en lumière plusieurs aspects. *Flexibilité comportementale* et *comportements adaptatifs* sont étudiés majoritairement dans le cadre de trois grands domaines scientifiques : l'écologie comportementale, la psychologie expérimentale et les neurosciences computationnelles. L'étude de la *flexibilité comportementale* et des *comportements adaptatifs* reflète donc le glissement progressif de l'étude simple du comportement vers son appréhension plus théorique et mécanistique grâce au développement de modèles computationnels et l'investigation des bases neurobiologiques des comportements observés. Ces approches récentes ont été permises grâce à l'accumulation de connaissances dans l'ensemble des champs disciplinaires des neurosciences ainsi qu'au développement de moyens technologiques et techniques permettant d'investiguer ces questions.

Ma thèse ayant oscillé entre la psychologie expérimentale et les neurosciences computationnelles, je m'attacherai dans la suite de ce chapitre à en présenter les principales caractéristiques, leur héritage historique ainsi que leurs enjeux. Mes travaux se sont également

intéressés à la compréhension des phénomènes neurobiologiques mis en jeu dans les comportements étudiés ainsi que les modèles computationnels proposés pour les expliquer.

## **1.2. FLEXIBILITE COMPORTEMENTALE ET COMPORTEMENTS ADAPTATIFS : DE L'ECOLOGIE A LA PSYCHOLOGIE**

### **1.2.1. AUX RACINES DE LA FLEXIBILITE COMPORTEMENTALE**

Comme introduits au paragraphe précédent, deux grands champs d'étude expérimentale existent lorsqu'il s'agit de s'intéresser au comportement animal : la psychologie expérimentale et l'écologie comportementale (Audet and Lefebvre, 2017; Pritchard et al., 2016). L'écologie comportementale est un champ disciplinaire qui cherche à expliquer le comportement et les mécanismes cognitifs comme des marqueurs prédictifs de l'aptitude des animaux à survivre et se reproduire dans leur milieu naturel. La psychologie expérimentale quant à elle est une approche plus anthropocentrale. Ce champ d'étude s'intéresse aux comportements animaux et aux mécanismes cognitifs qui les sous-tendent en comparaison à l'être humain. La psychologie expérimentale tend ainsi à identifier les comportements et les processus cognitifs qui existent ou n'existent pas. Puis, en analysant comparativement les bases neurobiologiques qui divergent entre les espèces, l'objectif est de comprendre les bases évolutives de la cognition humaine (Pritchard et al., 2016).

Bien que différentes dans leurs approches respectives, la psychologie expérimentale et l'écologie comportementale tentent toutes deux de répondre à plusieurs questions : comment les êtres vivants sont-ils en mesure de s'adapter face à un environnement en perpétuel changement ? Comment expliquer que certaines espèces semblent plus facilement s'adapter à des changements variés contrairement à d'autres (nouvel environnement, nouvelle nourriture, nouvel individu...) ?

Jusqu'à la moitié du XXème siècle, les théories éthologistes prônaient un comportement fixé sous forme de schéma inflexible dépendant de l'espèce considérée. Néanmoins, Poirier en 1969, suite à l'observation de singes dont le régime alimentaire était variable, notamment chez les femelles et les petits, a introduit le concept de *flexibilité comportementale* (Poirier, 1969). Depuis, la *flexibilité comportementale* a été le sujet de nombreuses études et les sujets d'intérêt variant en fonction des deux champs disciplinaires,

cela est à l'origine de l'hétérogénéité des théories proposées (Audet and Lefebvre, 2017; Lea et al., 2020) (**Figure 4**).

#### ETUDE DE LA FLEXIBILITE COMPORTEMENTALE EN :



**Figure 4 : Hétérogénéité de la signification de la *flexibilité comportementale* en fonction des champs disciplinaires.**

**En écologie comportementale.** Price et collaborateurs considèrent la différence entre mâles et femelles pour le choix des proies, l'utilisation de l'habitat et le rythme d'activité chez les poisson tête-de-boule comme de la *flexibilité comportementale* (Price et al., 1991). Dans l'étude de Knight, l'allaitement des jeunes par des hyènes femelles autre que la mère porteuse en cas de sécheresse est considéré comme un marqueur d'une telle flexibilité (Knight et al., 1992). L'aptitude à résoudre différents problèmes pour obtenir de la nourriture est une autre manière d'envisager la *flexibilité comportementale* à l'instar de l'étude menée chez différentes populations d'écureuils (Chow et al., 2018)

**En psychologie comportementale.** L'étude translationnelle du comportement de différentes espèces animales dont l'être humain dans le cadre de tâches comportementales au laboratoire tend à mettre en évidence des marqueurs de *flexibilité comportementale*. L'étude du comportement au laboratoire permet de contrôler un maximum de paramètres permettant ainsi de travailler sur une dimension spécifique. L'utilisation de certains dispositifs expérimentaux comme l'usage de chambres opérantes automatisées pour rongeurs permet d'apporter une plus grande validité écologique de la tâche par respect de la physiologie animale.

En écologie comportementale, certains parleront de *flexibilité comportementale* en mentionnant les différences de comportement au sein de populations d'une même espèce vivant dans des endroits différents ou entre individus d'une même espèce (Price et al., 1991). D'autres analyseront les changements d'un point de vue individuel en réponse à des éléments externes (Knight et al., 1992). Enfin, une grande partie de la littérature s'intéresse à la résolution de problèmes ou les processus d'innovation et de créativité (comme l'utilisation d'outils) (Chow et al., 2018).

En ce qui concerne la psychologie expérimentale, celle-ci cherche à expliquer les comportements comme la résultante d'un apprentissage de récompenses et de punitions obtenues par un individu (Audet and Lefebvre, 2017). Dans cette perspective, la *flexibilité comportementale* est donc définie comme une plasticité des comportements appris et est

essentiellement caractérisée par l'étude du comportement dans le cadre de paradigmes expérimentaux implémentés au laboratoire.

En elles-mêmes, toutes ces définitions semblent à première vue assez pertinentes car elles renvoient à la capacité des animaux à s'adapter lorsque leur environnement change, qu'il s'agisse de modifications dans la disponibilité ressources alimentaires, nécessitant d'aller explorer d'autres territoires, de l'accessibilité d'une source de nourriture requérant alors l'utilisation d'un outil ou tout simplement dans le changement d'une réponse au cours d'une tâche artificielle expérimentale au cours de laquelle les paramètres changent.

Une hypothèse de travail assez simple serait donc de considérer qu'un animal flexible, capable de s'adapter dans une situation donnée, est tout autant capable s'adapter face à de multiples autres situations de nature différente.

### **1.2.2. DE LA FLEXIBILITE COMPORTEMENTALE AUX COMPORTEMENTS ADAPTATIFS**

Il est intéressant de constater que l'hypothèse précédemment posée peut être rapidement infirmée par certaines études. Citons notamment celle de Griffin et collaborateurs qui ont travaillé chez le Martin triste. Les auteurs ont proposé aux animaux plusieurs tâches : certaines issues du champ disciplinaire de l'écologie comportementale, une autre issue du champ disciplinaire de la psychologie expérimentale (Griffin et al., 2013).

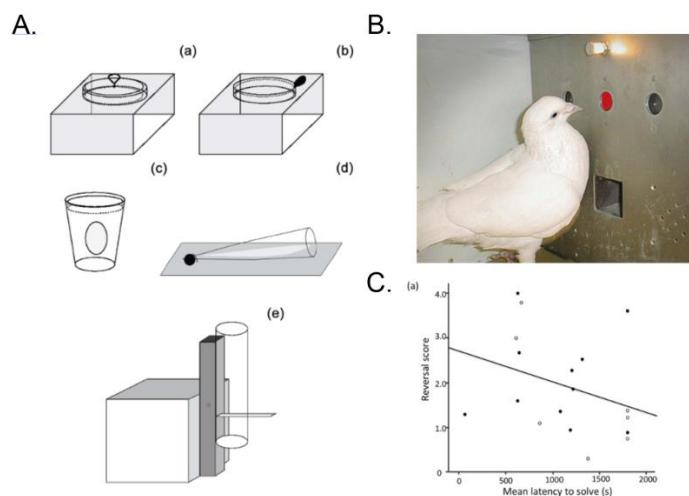
La première partie des tâches avait pour objectif de tester les capacités d'innovation et de résolution de problèmes chez ces oiseaux. Ceux-ci ont été présentés avec plusieurs types de contenants alimentaires requérant une technique d'ouverture particulière pour atteindre la nourriture qu'ils contenaient (**Figure 5A**).

La seconde partie des tâches consistait en une tâche d'apprentissage par renversement, dont nous rediscuterons plus longuement par la suite. En résumé, les oiseaux devaient appuyer sur un disque s'il apparaissait coloré en rouge, leur donnant accès à de la nourriture, mais pas si celui-ci était coloré en bleu (**Figure 5B**). A un moment dans la tâche, les contingences étaient inversées, et les oiseaux devaient alors changer leurs réponses pour appuyer lorsque la couleur bleue était affichée. Plus le changement de réponse est rapide, plus l'animal est considéré comme flexible.

Dans cette étude, les auteurs ont montré une corrélation négative entre la performance dans les tâches d'innovation (résolution de l'ouverture de contenants) et celle d'apprentissage par renversement : la capacité des oiseaux à s'adapter face à deux problèmes de nature

différente (innovation versus changements de contingences) n'est donc pas la même (**Figure 5C**).

En revanche, une récente étude par Audet et collaborateurs a montré que deux oiseaux d'espèces différentes, le pérenoir de Barbade et le cici verdinère, deux petits oiseaux proches des pinsons de Darwin, présentant tous les deux des capacités d'innovation différentes, avaient néanmoins des performances similaires dans des tâches de renversement d'apprentissage : il n'y aurait donc pas de lien entre la *flexibilité comportementale* mesurée par les capacités d'innovation, et la *flexibilité comportementale* mesurée par une tâche de psychologie expérimentale d'apprentissage par renversement (Audet et al., 2018).



**Figure 5 : La flexibilité comportementale étudiée au travers du prisme de deux champs disciplinaires.** A. Résolution de problèmes : accéder à de la nourriture cachée dans des contenants aux modes d'ouverture différents (soulever des couvercles, utiliser des crochets, retirer des morceaux de papier...). B. Tâche de renversement d'apprentissage chez les oiseaux : en fonction de la couleur de la pastille, les oiseaux doivent répondre ou non en appuyant avec leur bec. C. Performance dans la tâche de renversement (mesurée par un score prenant en compte plusieurs paramètres) et performance dans la tâche de résolution de problème (mesurée par le temps mis pour résoudre le problème) sont inversement corrélés. Adapté de (Griffin et al., 2013).

La diversité de conclusions montre à quel point le choix du champ disciplinaire est crucial. Ces études révèlent à quel point discuter de *flexibilité comportementale* comme un concept unique et générique est impossible. Il n'existe pas une forme de *flexibilité comportementale*, mais de multiples formes. La *flexibilité comportementale* consiste donc en un répertoire vaste de comportements qui s'expriment en fonction des contextes et des difficultés rencontrées. Ces *comportements adaptatifs* sont donc divers et spécifiques de ce à quoi il faut s'adapter, qu'il s'agisse d'un changement radical dans son environnement ou à l'introduction d'un nouvel outil.

### 1.2.3. ENJEUX SCIENTIFIQUES : CHOISIR SON CHAMP DISCIPLINAIRE

Comme le soulignent dans leur revue Lea et collaborateurs, « “behavioral flexibility” simply encompasses all learning to adapt to the particular conditions of an individual’s environment – in effect, the whole of animal cognition » (Lea et al., 2020). Quelle que soit l’espèce considérée, voire le règne du vivant concerné (les bactéries par exemple), le vivant est flexible. Mais cette flexibilité ne s’exprime pas de la même manière et les *comportements adaptatifs* qu’elle supporte sont donc différents et extrêmement variés.

D’un point de vue très pragmatique il est donc indispensable de choisir un champ d’étude scientifique et de définir les points qui sont à démontrer. Cherche-t-on à comparer différentes espèces ? Cherche-t-on à étudier une espèce en particulier ? Que cherche-t-on à mettre en évidence d’un point de vue comportemental ? Quelles sont les difficultés que nous souhaitons présenter aux animaux afin de vérifier s’ils sont capables ou non de les résoudre ? En fonction des questions qui se posent, il faut établir un protocole expérimental *ad hoc* car une étude comportementale bien pensée et riche peut apporter de nombreuses explications sur le comportement, et par extension le cerveau (Gao and Ganguli, 2015; Mainen et al., 2016; Niv, 2020). Il est néanmoins nécessaire de garder à l’esprit que les études menées, du fait des conditions expérimentales choisies notamment (Ladouce et al., 2017), ne sont pas nécessairement généralisables et que les résultats obtenus peuvent n’être que des cas particulier de *flexibilité comportementale*.

Pour ma part, mon travail de thèse s’est ancré dans un contexte d’étude de psychologie expérimentale étendu aux neurosciences cognitives et computationnelles. Il est donc essentiel de comprendre désormais le lien entre comportement et cognition.

## 1.3. FLEXIBILITE COGNITIVE, FLEXIBILITE COMPORTEMENTALE

### 1.3.1. DE LA PSYCHOLOGIE EXPERIMENTALE A LA PSYCHOLOGIE COGNITIVE

Les propos de ma thèse sont fondés sur un postulat cognitiviste, à savoir que les observations comportementales sont la résultante de processus mentaux complexes à l’instar de l’apprentissage. Même si ce postulat semble évident, il ne l’était pas jusque dans les années 50. Jusqu’à cette période, les théories behavioristes de Watson (1878–1958), Pavlov (1849–1936), Thorndike (1874–1949) et Skinner (1904–1990) prédominaient et prônaient de limiter

les études à ce qui est directement observable, à savoir le comportement (Nicolas and Ferrand, 2008). Les études behavioristes se limitaient donc à la simple observation de la réponse à la suite d'un stimulus donné. Skinner a étayé le champ d'étude en ajoutant au duo stimulus - réponse un troisième élément : l'ajustement du comportement en fonction des conséquences des réponses données. Néanmoins, ces théories ne pouvaient pas expliquer l'ensemble des aptitudes humaines (comme le langage ou la mémoire) et elles s'en trouvèrent largement critiquées, puisqu'elles laissaient le cerveau au statut de boîte noire (**Figure 6A**).

Les premiers travaux sur l'intelligence artificielle (Turing, 1921-1954), les théories de l'information (Shannon, 1916-2001 et Weaver, 1894-1978), la notion de computation (Neumann, 1903-1957) ont permis par la suite d'envisager de nouvelles manières d'investiguer et de comprendre le comportement du vivant et ce qui le sous-tend. Le cerveau est alors envisagé comme un ordinateur traitant les stimuli comme des entrées, générant des sorties sous forme de réponse suite à des transformations, réductions, élaboration, stockage et récupération d'informations complexes (Neisser, 2014) (**Figure 6B**). En 1949, le physiologiste Hebb dans son ouvrage *The organization of behavior* introduit notamment les notions d'organisation, de plasticité et de cognition, incitant les psychologues à travailler sur les processus cognitifs ainsi que biologiques.



**Figure 6 : Schémas simplifiés des approches behavioriste (A.) et cognitiviste (B.).** Là où les behavioristes se limitaient dans les années 1950-1960 à une approche comportementale stricte, les neuroscientifiques cherchent aujourd'hui à élucider les mécanismes cérébraux (à l'échelle computationnelle et neuronale) induisant les comportements observés.

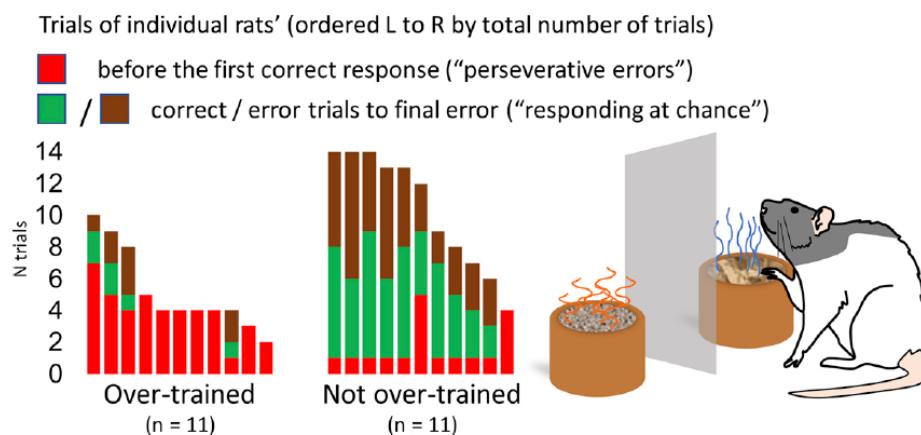
La notion de cognition est donc au cœur de plusieurs champs pluridisciplinaires dont la mise en œuvre croisée est essentielle à la compréhension du « comportement et de l'esprit » (Gomez-Marin and Mainen, 2016). Comprendre ce qui génère le comportement nécessite donc selon ces théories, un formalisme mathématique de modélisation ainsi qu'une étude approfondie des substrats biologiques qui supportent ces computations, sans oublier un regard philosophique critique (Nicolas and Ferrand, 2008). Comme le souligne Lucina Q. Uddin (Uddin, 2021), il n'y a donc pas de comportement sans cognition (réflexes exclus), mais à mon sens, la relation de causalité stricte entre les deux n'est pas aussi triviale qu'il y paraît.

### 1.3.2. DISSOCIABILITE DES FLEXIBILITES COGNITIVE ET COMPORTEMENTALE

The terms “behavioral flexibility” and “cognitive flexibility”, used interchangeably, have been employed for decades in the field of experimental psychology to label a form of cognition that enables animals and humans to adapt their behavior to changing environmental contingencies. (Audet and Lefebvre, 2017)

Ce que soulignent Jean-Nicolas Audet et Louis Lefebvre est révélateur d'un usage mixte de la terminologie scientifique car *flexibilité comportementale* et *flexibilité cognitive* sont deux entités bien distinctes selon certains auteurs. L'usage synonyme des deux est essentiellement lié au fait que les auteurs souhaitent réaliser une distinction claire entre des modifications comportementales de type réflexes ou purement physiologiques (par exemple, grelottement ou piloérection si les températures chutent brutalement), des modifications comportementales résultant d'un traitement d'informations au niveau cérébral (construire un nid douillet s'il fait froid) (Bayne et al., 2019; Brown and Tait, 2010).

Dans la lignée de leur revue de 2010, l'équipe de Brown et Tait a tenté de démontrer formellement que *flexibilité comportementale* et *flexibilité cognitive* sont dissociables (Dhawan et al., 2019). Les auteurs ont entraîné des rats dans une tâche de renversement d'apprentissage. Des rats ont appris à discriminer deux bols aux odeurs différentes, l'un contenant des récompenses, l'autre non. Une fois que les rats choisissaient le bol récompensant à chaque nouvel essai, certains ont été présentés avec la même configuration pendant 30 essais supplémentaires avant que la récompense ne soit alors placée dans l'autre bol : ces rats étaient donc surrentraînés (**Figure 7**).



**Figure 7 : Dissociabilité de la flexibilité comportementale et de la flexibilité cognitive.** Un surrentraînement des rats dans une tâche d'apprentissage par renversement assure un changement de réponse vers l'option récompensante plus rapide malgré des réponses persévérvantes. (Dhawan et al., 2019)

Une fois que les contingences étaient inversées, les auteurs ont remarqué que les rats surentraînés présentaient plus de réponses persévératrices (ils continuaient de vérifier le bol anciennement récompensant) mais étaient tout de même plus rapides que les rats non surentraînés pour finalement opter pour le nouveau bol. De par leur comportement de persévération, les rats surentraînés semblent moins flexibles : leur comportement semble être en effet plus rigide qu'un comportement de réponse au hasard. Pourtant leur capacité à opter rapidement pour la nouvelle option récompensante suggère une plus grande capacité d'adaptation cognitive grâce à une meilleure représentation interne de leur environnement en utilisant notamment les informations contrefactuelles de la tâche (si ce n'est pas ce bol qui contient la récompense, il s'agit nécessairement de l'autre).

*Flexibilité comportementale* et *flexibilité cognitive* sont donc dissociables au sens où mesurer la première ne permet pas nécessairement de donner un proxy fiable quant à la seconde. Tandis que la *flexibilité comportementale* rend compte de processus comportementaux, la *flexibilité cognitive* se réfère aux computations cérébrales sous-jacentes. Dans le cadre de la *flexibilité*, cognition et comportement sont intimement liés, mais il serait peu pertinent de les traiter comme des analogues (Uddin, 2021). Il est donc nécessaire de travailler au sein d'un cadre théorique précis afin d'établir des hypothèses sur les processus cognitifs mis en jeu et en quoi ces processus rendent compte des modifications comportementales observées.

### **1.3.3. ENJEUX SCIENTIFIQUES : ETABLIR DES EXPLICATIONS MECANISTIQUES A PARTIR D'OBSERVATIONS COMPORTEMENTALES**

Rendre compte de la *flexibilité comportementale* requiert donc une compréhension du niveau biologique supérieur qu'est la *flexibilité cognitive*. Dans leur revue « Demystifying cognitive flexibility », Dajani et Uddin tentent de définir ce concept. Ils établissent que celui-ci est une propriété émergente des fonctions exécutives<sup>2</sup> résultant de quatre grandes autres fonctions :

- La détection de saillance et l'attention assurant cette détection en filtrant les d'informations d'intérêt.
- La mémoire de travail permettant de maintenir « actives » les informations recueillies et analysées

---

<sup>2</sup> Ensemble des processus de contrôle mentaux supportant la mise en œuvre de comportements dirigés vers un but. (Dajani and Uddin, 2015)

- L'inhibition responsable de l'arrêt des stratégies précédemment utilisées pour répondre aux contraintes passées
- Le changement de stratégie, appelé *switching*, en anglais afin de répondre aux nouvelles contraintes

Néanmoins, ce type de processus met en jeu des fonctions cognitives ubiquitaires et peu spécifiques qui ne peuvent exactement refléter la capacité d'être flexible. Pour certains chercheurs en effet, c'est la *capacité d'apprendre à apprendre*, et d'adapter ce qui a été précédemment appris à de nouvelles instances d'une situation, qui est le support de la *flexibilité comportementale* (Faraut, 2015). D'autres considèrent le *contrôle adaptatif* comme moyen de guider le comportement à travers des erreurs de prédiction (Narayanan et al., 2013).

Diverses catégories de fonctions exécutives, donc différents modules cognitifs et cérébraux sont impliqués dans ces processus. Etudier les *comportements adaptatifs* et les mécanismes cognitifs mis en jeu requiert donc un cadre théorique précis, mettant en jeu des hypothèses computationnelles clairement définies dont les prédictions sont à comparer avec les résultats comportementaux obtenus.

Dans cette perspective, je m'intéresserai et me limiterai pour ma part aux théories d'apprentissage par renforcement et dérivées. De par mon approche expérimentale et le champ disciplinaire dans lequel ma thèse a été développée, je considère en effet que « l'adaptation comportementale est le processus par lequel les êtres vivants sont capables d'extraire et apprendre les règles de leur environnement » (Logiaco et al., 2015). Je considérerai donc l'adaptation en lien avec les processus d'apprentissage puisque l'apprentissage est cette faculté des êtres vivants à s'adapter en fonction de leur expérience passée.

Tout l'enjeu désormais est de comprendre comment étudier ces processus au laboratoire.

## **2. ETUDE EXPERIMENTALE DES COMPORTEMENTS ADAPTATIFS**

### **2.1. COMPORTEMENTS NATURELS OU ARTIFICIELS ?**

#### **2.1.1. METHODES D'ETUDE DU COMPORTEMENT : DE LA NATURE AU LABORATOIRE**

Suivant les champs disciplinaires de la psychologie expérimentale et écologie comportementale, deux grandes manières expérimentales d'envisager l'étude du comportement existent. D'une part, l'étude des animaux dans la nature, et d'autre part l'étude en laboratoire (Pritchard et al., 2016). Les deux méthodologies peuvent se comparer sur différents aspects.

##### **1) Espèces animales étudiées**

Les expériences dans la nature permettent d'envisager l'étude d'une plus grande diversité d'espèces, non génétiquement modifiées, vivant au sein de leur biotope. Au laboratoire en effet, le nombre d'espèces animales étudiées est plus restreint du fait des règles éthiques qui interdisent l'expérimentation sur certains animaux sauvages ainsi que de la capacité limitée de certains animaux à vivre en captivité (Pritchard et al., 2016). De ce fait, l'ordre animal le plus courant au laboratoire est celui des rongeurs (souris et rats essentiellement) même si certaines équipes tâchent de travailler sur des modèles animaux plus diversifiés (Steurer et al., 2012). Il faut également noter que les études en laboratoire utilisent des animaux issus de lignées transgéniques qui sont donc très probablement différents d'animaux que l'on peut retrouver dans la nature (Saunders, 2020). Néanmoins, ce désavantage est aussi un atout. En effet, grâce aux outils génétiques, il est possible de créer artificiellement des animaux modèles de certaines pathologies, dont l'étude comportementale peut s'avérer utile pour comprendre l'implication de certaines aires cérébrales dans la réalisation de certains comportements.

##### **2) Protocoles expérimentaux et technologies employées**

Historiquement, dans la nature, comme au laboratoire, le comportement animal s'étudie essentiellement par une analyse visuelle qui peut être plus ou moins automatisée. Les comportements sont par exemple annotés à la main par des expérimentateurs dissimulés à proximité. Les animaux peuvent être également filmés par de multiples caméras. L'analyse

vidéo peut se faire à la main, mais il existe de plus en plus d'outils logiciels afin de quantifier automatiquement certains comportements (de Chaumont et al., 2012; Mathis and Mathis, 2020; Nath et al., 2019). Il existe également pléthore de tâches comportementales pouvant être administrées aux animaux à la fois dans la nature (Healy and Hurly, 2013), et au laboratoire afin de tester de multiples aspects comportementaux. Depuis des années, on assiste à une sophistication des tâches employées, afin de tester des comportements et cognitions sous-jacentes de plus en plus complexes. Cette évolution s'accompagne d'une révolution technologique d'automatisation et d'acquisition de données à haut-rendement qu'il est nécessaire de suivre (Schaefer and Claridge-Chang, 2012). Au laboratoire, ces tâches comportementales sont souvent associées à un suivi ou une manipulation de l'activité neuronale afin d'évaluer l'implication de certains substrats neuronaux dans le comportement étudié.

### 3) Présence de l'expérimentateur

Dans la continuité du point 2, il est à noter que la présence des expérimentateurs est variable en fonction des études. Les expériences menées dans la nature s'accompagnent souvent d'une faible présence des expérimentateurs car les espèces animales étudiées sont naturellement peu en contact avec l'espèce humaine. Au laboratoire en revanche, la présence humaine est permanente, qu'il s'agisse des soins aux animaux à la réalisation des expériences.

## **2.1.2. VALIDITE ET INTERET SCIENTIFIQUE DES JEUX DE DONNEES COLLECTES**

Les objectifs poursuivis par l'étude en laboratoire et dans la nature ne sont pas les mêmes. Alors que l'étude dans la nature permet de respecter fidèlement l'environnement écologique de l'animal ainsi que l'animal lui-même (son héritage génétique et épigénétique, son histoire...) tout en limitant la présence humaine, l'étude en laboratoire permet de créer des conditions parfaitement contrôlées (des naissances à l'alimentation en passant par l'entraînement et le paradigme présenté à l'animal) (Pritchard et al., 2016).

La première approche permet donc d'envisager au mieux le comportement et la cognition animale. En effet, travailler sur des lignées génétiquement modifiées n'ayant jamais vécu dans la nature peut biaiser les résultats. Les discordances de résultats comportementaux sur une même tâche comportementale entre différentes lignées de souris en est la preuve (Turner et al., 2017). Le stress induit par la présence d'expérimentateur n'est également pas

sans conséquence sur les animaux (Balcombe et al., 2004; Hurtubise and Howland, 2017), tout comme le caractère « non naturel » du laboratoire car relativement peu « enrichi » et très stéréotypé. Comme le souligne Pritchard : “What animals can do in the laboratory may be quite different from what they are able to do in the wild.” (Pritchard et al., 2016). Néanmoins, l'approche au laboratoire permet de faire varier rigoureusement les paramètres un par un afin d'identifier celui d'intérêt, chose qui est moins aisée, voire impossible dans la nature. Elle permet également d'étudier et de manipuler en parallèle l'activité cérébrale, ce qui est indispensable aujourd'hui en neurosciences.

Par ailleurs, les questions auxquelles il est nécessaire de répondre ne sont pas les mêmes. Alors qu'il est plus pertinent d'étudier certains comportements naturels dans la nature (comme les mouvements migratoires de certains oiseaux en fonction du changement des saisons), les questions relatives à des cognitions plus complexes en comparaison avec celles de l'être humain sont plus facilement étudiables au laboratoire.

Dans le cadre de ma thèse, du fait des objectifs poursuivis (étudier les bases neuronales des *comportements adaptatifs*) c'est une approche au laboratoire qui a été suivie. Il a été néanmoins nécessaire de répondre à plusieurs enjeux afin de pallier le risque d'étudier des comportements totalement artificiels au laboratoire.

### **2.1.3. ENJEUX SCIENTIFIQUES : CONTROLER ECOLOGIQUEMENT SES EXPERIENCES AU LABORATOIRE**

La réalisation de tâches comportementales complexes par des animaux de laboratoire peut vite requérir une présence importante de l'expérimentateur. Ceci peut ainsi stresser les animaux et les perturber dans leur comportement, et il s'agit également d'une contrainte pour l'expérimentateur. De même l'étude de certains comportements complexes requiert une haute fiabilité des données intra-individuelles. Elle peut nécessiter également de longues périodes d'expérimentation (sur plusieurs semaines) ainsi que la collecte d'une grande quantité de données dans le cas où celles-ci seraient le support de modélisation.

Il est donc nécessaire de développer une nouvelle manière d'expérimenter permettant :

- 1) De limiter la présence de l'expérimentateur
- 2) Respecter au mieux la physiologie de l'animal
- 3) Collecter un grand nombre de données sur des périodes plus ou moins longues en fonction des tâches comportementales administrées.

Dans cette perspective, il est également essentiel de repenser les méthodes d'analyses de quantités massives de données pour chaque animal impliqué dans un protocole expérimental. L'analyse de multiples variables, des plus simples aux plus complexes, est alors possible et peut apporter des renseignements précieux (Schaefer and Claridge-Chang, 2012). Par ailleurs, il est nécessaire de prendre en compte l'individualité des animaux (Torquet et al., 2018) et aller au-delà des simples analyses moyennées car souvent trompeuses (Gallistel et al., 2004).

Dans le cadre de ma thèse, il m'a donc été nécessaire de répondre à ces enjeux afin de pouvoir étudier les *comportements adaptatifs* lors d'un changement abrupt de l'environnement.

## **2.2. PROTOCOLES EXPERIMENTAUX CLASSIQUES D'ETUDE DES COMPORTEMENTS ADAPTATIFS EN CAS DE CHANGEMENT ABRUPT DE L'ENVIRONNEMENT**

### **2.2.1. VALIDITE ECOLOGIQUE ET INTERET SCIENTIFIQUE DE LA QUESTION**

Comme discuté précédemment, il existe de multiples possibilités concernant l'étude des *comportements adaptatifs*. Mon travail de thèse a eu pour objectif d'en caractériser un cas particulier, à savoir l'adaptation comportementale à un changement abrupt de l'environnement, dans un cadre théorique de psychologie expérimentale.

L'environnement dans lequel évolue le vivant n'est pas stable : il est donc qualifié de *volatil*. Des changements peuvent y intervenir sans qu'ils ne soient anticipés. Ces situations sont courantes dans la nature, du fait notamment de l'activité humaine. Par exemple, les déforestations massives conduisent à la disparition soudaine d'habitats naturels de certains primates en Amérique du Sud. Ces déforestations mènent d'une part à une extinction majeure de certaines espèces de singes et d'autre part à la modification de certains comportements pourtant déjà bien établis. Du fait de la disparition de certains arbres et de la fragmentation de leur territoire, les singes hurleurs par exemple modifient leur alimentation là où une autre alternative comportementale pourrait consister en une migration du groupe (Cristóbal-Azkarate and Arroyo-Rodríguez, 2007). Toute la problématique est de comprendre ces changements de stratégies comportementales et ce qui les sous-tend d'un point de vue

cognitif. L'être humain quant à lui, est aussi confronté à ce genre de difficultés quotidiennement. Les problèmes rencontrés ne sont pas toujours les mêmes que ceux des animaux, dont une grande partie du temps est en effet dédiée à la recherche alimentaire et la recherche de partenaires, mais ils posent tout de même des questions similaires.

La formalisation au laboratoire de tâches comportementales génériques permettant de mimer ce type de problèmes est donc devenue essentielle.

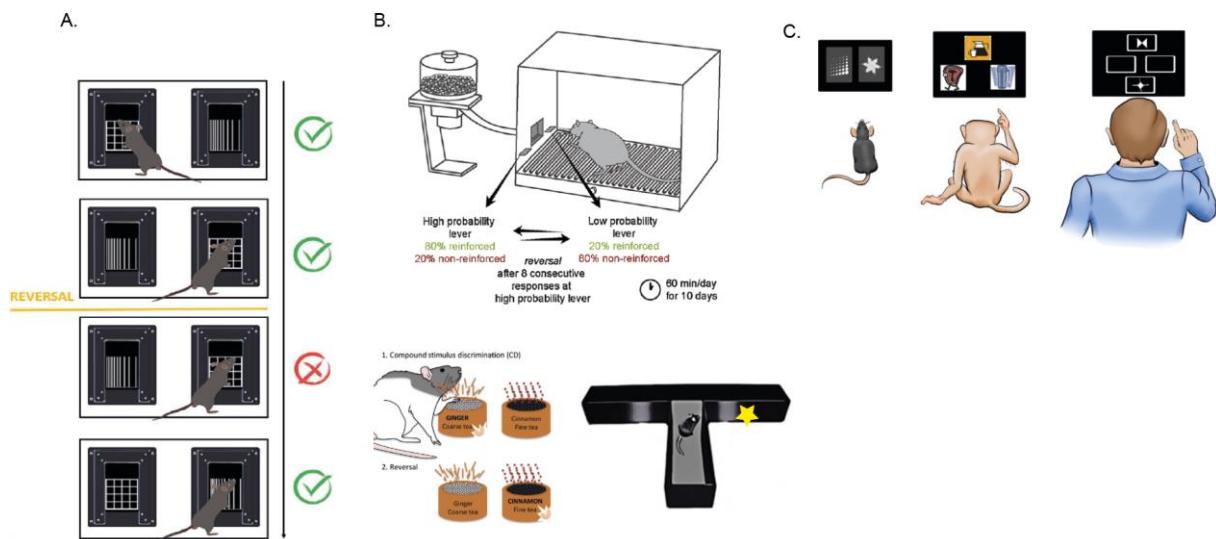
## **2.2.2. TACHE COMPORTEMENTALE DE RENVERSEMENT D'APPRENTISSAGE : PROTOCOLE HISTORIQUE D'ETUDE DU COMPORTEMENT FACE A DES CHANGEMENTS ABRUPTS**

### **1) Paradigme expérimental : mimer un changement abrupt de l'environnement**

La tâche de renversement d'apprentissage (*reversal learning* en anglais) est une des tâches comportementales les plus utilisées que cela soit chez l'être humain, les animaux modèles de laboratoire comme le singe, le rat, la souris (Izquierdo et al., 2017) ou encore les animaux sauvages (Griffin et al., 2013). La possibilité d'implémenter cette tâche chez de multiples espèces lui confère l'avantage majeur de pouvoir en comparer les résultats sous réserve de la similarité des protocoles employés (Benzina et al., 2021). Le *reversal learning* a donc une forte valeur translationnelle (Jarvers et al., 2016) (**Figure 8C**).

Cette tâche a été conçue afin d'évaluer le comportement des animaux lorsqu'ils sont confrontés à une modification de leur environnement (expérimental dans ce cas précis). Le principe de la tâche est tel que décrit ci-après. On présente au sujet deux options, l'une donnant lieu à une récompense (réponse correcte) si elle est choisie, l'autre non (**Figure 8A et B**). Une fois que le sujet atteint un seuil de performance (calculé sur le nombre de réponses correctes sur les X derniers choix) signifiant ainsi qu'il a bien compris la tâche, les contingences sont inversées sans que le sujet ne soit prévenu : l'option précédemment récompensée ne l'est plus, celle qui ne l'était pas, l'est désormais (**Figure 8A**). Ce changement est appelé renversement et mime un changement abrupt de l'environnement dans lequel le sujet évolue. Au cours de la tâche, un renversement, ou de multiples renversements peuvent avoir lieu. En termes de modalités sensorielles, les options présentées sont assez variables en fonction de l'espèce. Pour l'être humain, la tâche de *reversal learning* est administrée par ordinateur, les options sont ainsi des stimuli visuels (**Figure 8C**). Pour les autres animaux, les systèmes expérimentaux et les stimuli peuvent être plus variés. Il peut s'agir d'odeurs, de figures visuelles, de leviers à presser, de bras à choisir dans un labyrinthe

en T ou encore bien d'autres possibilités (**Figure 8B et C**) Les protocoles expérimentaux sont donc extrêmement variés chez les animaux non humains. Un dernier point varie également entre les études. Certaines expériences de *reversal learning* sont déterministes, c'est-à-dire que le choix de la réponse correcte donnera lieu 100 % du temps à une récompense, tandis que d'autres sont probabilistes, c'est-à-dire que le choix de la réponse correcte ne mène pas toujours à une récompense (entre 10% et 40% du temps) (**Figure 8B, haut**). Ceci permet d'augmenter la difficulté de la tâche.



**Figure 8 : Le reversal learning.** **A.** La tâche de *reversal learning* repose sur un changement brutal non indicé des contingences du paradigme expérimental. L'option récompensée devient non récompensée et vice-versa. **B.** Diversité des tâches de *reversal learning*. La tâche peut être probabiliste (Verharen et al., 2019) ou déterministe. Les options peuvent reposer sur de multiples modalités : leviers (Verharen et al., 2019), odeurs (Dhawan et al., 2019), labyrinthes (Brigman et al., 2010)... **C.** Possibilité d'investiguer à travers les espèces les résultats du *reversal learning* (Izquierdo et al., 2017).

Des variantes de la tâche ont été développées comme les tâches de *reversal* avec plus de deux options (Remmelink et al., 2016) ou plusieurs paires de stimuli (Wilson and Gaffan, 2008) permettant d'explorer plus finement certaines stratégies à l'instar des comportements d'exploration et d'exploitation (Addicott et al., 2017). De cette tâche dérivent également d'autres tâches comportementales comme *l'intra/extradiensional shift*, le *task-switching* ou encore les *tâches d'extinction* (Hamilton and Brigman, 2015). Je n'en discuterai pas ici car d'autres facultés cognitives sont mises en jeu comme les facultés attentionnelles sur les différentes dimensions de stimuli (*intra/extradiensional shift*). Par ailleurs, les *tâches d'extinction* ne nécessitent pas de ré-apprentissage à la suite du changement de contingences et dans le cadre du *task-switching*, les changements de règles sont indiqués au sujet. Ce type de tâches limite donc l'accès à certains processus d'intérêt.

2) Intérêts : comprendre les stratégies cognitives sous-jacentes à l'adaptation comportementale à la suite d'un changement abrupt de contingences

Le *reversal learning* permet d'évaluer comment un sujet s'adapte lors d'un changement brutal (Izquierdo et al., 2017). Classiquement, le nombre d'erreurs réalisées ainsi que la performance à la suite du changement abrupt de contingences sont analysés. Lorsque le sujet continue de répondre comme il le faisait avant le changement, il présente un comportement dit de *persévération*. Il finit néanmoins par présenter un *comportement adaptatif*, puisqu'il finit par changer de réponse.

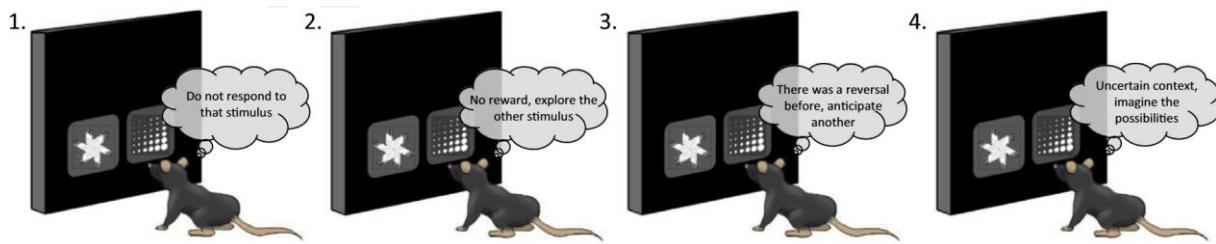
Le comportement de *persévération* observé dans un premier temps peut s'envisager comme un défaut de contrôle inhibiteur et le changement de comportement (réponse sur la bonne option) comme une reprise de contrôle inhibiteur (**Figure 9.1**). Néanmoins, cette hypothèse a été rapidement mise à l'écart : il a notamment été montré que des rats dont le cortex orbitofrontal avait été lésé étaient capables d'inhiber leur réponse face à la présentation de stimuli non récompensés mais présentaient un déficit lors du renversement de contingences de la tâche (Schoenbaum et al., 2002).

De cette première approche très simple découle une analyse plus complexe qui s'intéresse à la manière dont le sujet prend en compte les conséquences de ses actions à la suite de la présentation des options. Comme nous le verrons au Chapitre 2, l'histoire des récompenses et de l'absence de récompense permet au vivant d'apprendre (**Figure 9.2**). Le comportement de *persévération* plus ou moins marqué peut ainsi s'expliquer par une faculté d'apprentissage plus ou moins efficace et ceci peut s'évaluer grâce à des modèles simples comme l'apprentissage par renforcement dont nous discuterons de manière approfondie ultérieurement.

Par ailleurs, des études au cours desquelles plusieurs renversements sont proposés à des animaux, montrent qu'au fur et à mesure des renversements, les sujets changent plus rapidement de réponse (Jarvers et al., 2016). Une des hypothèses suggère ainsi que les sujets sont potentiellement capables de prédire et d'estimer quand les renversements ont lieu, assurant la mise en place de comportements adaptés de manière rapide (Bartolo and Averbeck, 2020; Costa et al., 2015) (**Figure 9.3**).

Enfin, un niveau d'abstraction encore supérieur suggère qu'il existe une capacité à se représenter la tâche sous formes d'états (Tolman, 1948; Wilson et al., 2014a). Ces états consistent en une représentation abstraite entre les différentes options que présente la tâche et leurs conséquences dans le cas où ces options sont choisies. (**Figure 9.4**)

Je rediscuterai plus en détails de ces deux derniers points au Chapitre 2 également.



### 2.2.3. ENJEUX SCIENTIFIQUES : DEVELOPPER UNE NOUVELLE TACHE COMPORTEMENTALE TRANSLATIONNELLE AFIN DE DEPASSER LES LIMITES DE LA TACHE DE RENVERSEMENT D'APPRENTISSAGE

Dans sa forme la plus simple, le *reversal learning* est une tâche certes très intéressante mais elle limite grandement les conclusions que l'on peut en tirer (Manning et al., 2021; Piantadosi et al., 2019).

Dans un premier temps, le fait de ne proposer que deux options permet une résolution de la tâche par la simple mise en jeu d'une stratégie de *win-stay* (je gagne sur cette option – je reste sur cette option), *lose-shift* (je perds sur cette option – je change d'option). Ceci peut donc limiter la mise en œuvre de stratégies plus complexes et ainsi masquer l'existence de certains processus comportementaux et cognitifs comme la recherche de nouvelles informations.

Dans un deuxième temps, la compréhension des processus cognitifs qui sous-tendent le *comportement adaptatif* (arrêter de persévéérer et changer d'option) à la suite des changements de contingences expérimentales requiert la mise en œuvre de stratégies plus ou moins complexes. Les conclusions ou hypothèses pouvant être proposées par la simple analyse des résultats comportementaux sont en effet extrêmement restreintes, car elles se limitent essentiellement à l'analyse de la performance et des erreurs persévérvatives. Ces erreurs ne permettent pas de distinguer dans ce cas s'il s'agit d'un défaut d'inhibition, d'un défaut de détection des changements dans l'environnement ou d'un défaut d'apprentissage.

Dans un dernier temps, les tâches de renversement d'apprentissage ne mettent pas en jeu un ou plusieurs stimuli supplémentaires dont la présence pourrait influencer la prise de décision des sujets (Chau et al., 2014). Or cette simplicité du paradigme ne permet pas d'investiguer les différentes hypothèses proposées quant à l'adaptation comportementale en cas de changement abrupt des contingences (**Figure 11. 2. 3. 4.**) (Izquierdo et al., 2017). Afin de pallier cette limitation, Piantadosi et collaborateurs (Piantadosi et al., 2019) ont par exemple augmenté le niveau de difficulté de la tâche en ajoutant dans leur étude un stimulus jamais récompensé, leur permettant d'étudier ainsi plus finement les stratégies des animaux.

Trois enjeux se dégagent donc comme support du développement d'une nouvelle tâche comportementale :

- 1) La tâche devra permettre de proposer par simple analyse des résultats comportementaux bruts un ensemble d'hypothèses concernant les mécanismes cognitifs sous-jacents afin de savoir si les processus mis en jeu relèvent d'heuristiques simples, et/ou de mécanisme d'apprentissage adaptatif, et/ou d'une représentation plus abstraite de la tâche (anticipation des changements et/ou représentations des états). La validation de ces hypothèses peut s'envisager dans un second temps par une étape de modélisation mathématique.
- 2) La tâche doit pouvoir être compatible avec la mise en place de protocoles d'étude et de manipulation des circuits neuronaux afin de mettre en évidence les circuits critiques support des computations et comportements observés.
- 3) La tâche doit pouvoir être adaptée à différentes espèces afin de pouvoir comparer les résultats et de proposer des hypothèses évolutives.

## CONCLUSION

Ce premier chapitre a permis de faire un état des lieux des concepts et méthodologies essentiels à la compréhension de mon travail de thèse. Le deuxième chapitre s'intéressera de manière approfondie à la nature des comportements mis en jeu à la suite d'un changement abrupt dans l'environnement et aux hypothèses actuelles concernant l'efficacité des mécanismes cognitifs possiblement liés à ces comportements.



## CHAPITRE 2. Explorer et/ou exploiter, telle est la question : de l'apprentissage aux inférences comme supports des comportements adaptatifs

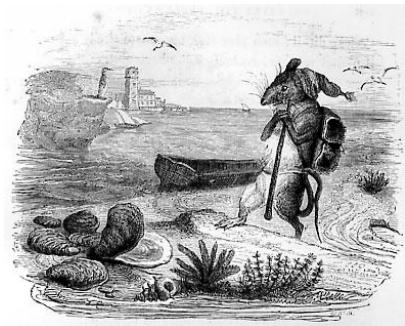
---

### INTRODUCTION

Cette fable contient plus d'un enseignement.  
Nous y voyons premièrement :  
Que ceux qui n'ont du monde aucune expérience  
Sont aux moindres objets frappés d'étonnement ;  
Et puis nous y pouvons apprendre,  
Que tel est pris qui croyait prendre.

« Le Rat et l'Huître », *Fables*, Livre VIII, Jean de La Fontaine

J.J. Grandville



Au-delà de sa plume habile et de son esprit tranchant, Jean de La Fontaine était un fin observateur du comportement animal. La relecture de ses œuvres pourrait sans doute inspirer plus d'un neuroscientifique aujourd'hui ! « Le Rat et L'Huître », par exemple, relate l'histoire d'un rat n'ayant connu que son nid, et qui, un beau jour, se prend l'envie de découvrir le monde. Parcourant montagnes et vallées, campagnes et bords de mer, son aventure se termine lorsqu'il est pris au piège dans la coquille d'une huître. Malgré ses pérégrinations, l'expérience accumulée, il reste ignare du monde, de ses lois et de ses dangers en demeurant un éternel aventurier : maître rat n'a donc pas su s'adapter.

Remise au goût des Neurosciences, cette fable est l'illustration de l'une des nombreux équilibres qui pondèrent le vivant (Del Giudice and Crespi, 2018), à savoir le dilemme entre *exploration* et *exploitation*. Pour assurer leur survie, les organismes vivants doivent en effet assouvir leurs besoins dans un environnement contraignant, parfois changeant. Leur survie à court terme (se nourrir, trouver un abri...) comme à long terme (se reproduire, protéger leur petit des prédateurs...) nécessite donc de maximiser des éléments bénéfiques à leur survie, et de minimiser ce qui peut leur être délétère. Néanmoins, toute la question est de savoir quels sont ces éléments, où et comment ils se trouvent et lesquels sont les plus avantageux en regard du coût que leur recherche et acquisition nécessitent.

Ce compromis entre un mode *exploratoire* et un mode *exploitatoire* est donc au cœur du processus de prise de décision chez tous les êtres vivants (Addicott et al., 2017) et est d'autant plus critique en regard de la *volatilité* inhérente à l'environnement. Lorsque

l'environnement change, ce qui était connu auparavant n'est probablement plus valide. Présenter les mêmes comportements que précédemment peut s'avérer délétère et d'autres comportements doivent ainsi émerger afin de mieux connaître son nouvel environnement et d'en tirer le plus de bénéfices.

Le dilemme *exploration – exploitation* est donc au cœur de la problématique de l'*adaptation comportementale* telle que définie par Logiaco et collaborateurs (Logiaco et al., 2015) (cf Chapitre 1). Lorsque l'environnement change, il semble logique de l'*explorer* pour l'appréhender et déterminer quelles sont les nouvelles meilleures options à *exploiter* par la suite. Ce processus repose nécessairement sur une ou plusieurs formes d'apprentissage. L'apprentissage est en effet la capacité du vivant à modifier son comportement en fonction de l'expérience accumulée dans le passé. Ce deuxième chapitre vise ainsi à mettre en lumière les comportements mis en jeu ainsi que les mécanismes computationnels pouvant les générer au regard de la balance *exploration – exploitation*.

## **1. EXPLORER, EXPLOITER : MISE EN ŒUVRE COMPORTEMENTALE D'UN PROCESSUS COGNITIF D'APPRENTISSAGE**

### **1.1. LE DILEMME EXPLORATION – EXPLOITATION**

#### **1.1.1. LE DILEMME DU FOURRAGEUR**

Les origines du dilemme *exploration – exploitation* remontent aux premiers travaux d'écologie comportementale dans les années 1960 - 1970 sur le *foraging*, terme anglais qui signifie « fourrageage », à savoir la recherche de ressources alimentaires. L'expérimentation a montré chez la grande mésange que celle-ci ne se contente pas d'ingérer les larves de coléoptères les plus avantageuses nutritivement pour elle, mais 10% du temps, elle teste des larves moins optimales (Charron and Collette, 2012). Les oiseaux oscillent donc entre l'*exploitation* de ressources optimales et l'*exploration* d'autres alternatives. Tout l'enjeu des recherches actuelles et passées est de comprendre la nature de ces stratégies et comment passer de l'une à l'autre.

Dans le cadre du fourrageage et de l'acquisition alimentaire au sein d'un site dont la richesse s'appauvrit au fur et à mesure de son exploitation, Charnov a proposé alors le

théorème de la valeur marginale (*marginal value theorem*) : un prédateur doit quitter un site d'acquisition alimentaire lorsque le taux instantané d'acquisition alimentaire atteint la vitesse moyenne maximale d'acquisition possible au sein de ce site (Charnov, 1976). Néanmoins, ce théorème repose sur des hypothèses restrictives comme notamment la décroissance progressive de la consommation des ressources alimentaires par le prédateur, le fait que les sites sont visités successivement et que le prédateur en connaît les propriétés.

Depuis ces premières études, le dilemme *exploration – exploitation* s'est étendu à de nombreux autres sujets que celui de la recherche alimentaire (Cohen et al., 2007; Hills et al., 2015). Les questions initialement posées dans le cadre du champ disciplinaire de l'écologie comportementale se sont progressivement transposées au laboratoire sous le prisme de la psychologie expérimentale et des neurosciences computationnelles. Le théoricien Gittins s'est par exemple penché sur ces questions dans un cadre conceptuel différent, celui des problèmes des machines à sous. L'indice qu'il a proposé pour résoudre ce problème (Gittins, 1979) permet d'envisager comment choisir entre plusieurs options déjà testées par un sujet ou inconnues de lui. Néanmoins même s'il semble que dans la nature les êtres vivants peuvent se comporter approximativement suivant ce modèle (Krebs et al., 1978), l'implémentation cérébrale du calcul de cet indice reste difficilement crédible car trop complexe. Depuis la balance *exploration – exploitation* a été largement étudiée dans le cadre de multiples domaines comme celui de la recherche d'informations visuelles (Chun and Wolfe, 1996), les comportements sociaux (Wilson, 1968; Wisdom et al., 2013), l'intelligence artificielle (Osugi et al., 2005), la neuroéconomie (Hayden, 2017), le management (Laureiro Martinez et al., 2014; Laureiro-Martínez et al., 2010) ou encore l'étude de certaines pathologies (Addicott et al., 2017).

Face à la multiplicité de paradigmes expérimentaux (Addicott et al., 2017), d'hypothèses restrictives posées par les modèles, les contraintes de temporalité, spatialité, variabilité intra-individuelle, phénomènes sociaux (Hills et al., 2015; Mehlhorn et al., 2015), l'ensemble du domaine présente une multiplicité de conclusions :

“The theoretical analysis of exploration-exploitation tradeoffs is complicated in several ways. First, the concepts of exploration, exploitation, and a tradeoff between the two are used in a wide range of literatures, from animal behavior to human behavior and involving different terminologies, methodologies, and perspectives. The disparity and breadth of these concepts across such large and diverse literatures makes it difficult to synthesize existent knowledge into a coherent view. Researchers working in different areas operationalize exploratory and exploitative behavior in different ways, and this diversity motivates disagreements within conclusions about the essence of exploratory and exploitative behaviors.” (Mehlhorn et al., 2015)

Il est donc nécessaire d'investiguer les fondements de ce dilemme afin de comprendre comment ces stratégies d'*exploration* et d'*exploitation* sont mises en jeu par les êtres vivants et comment elles sont essentielles à leur *adaptation*.

### **1.1.2. LES DETERMINANTS DE LA BALANCE EXPLORATION - EXPLOITATION**

Le dilemme *exploration – exploitation* est au cœur des processus de décision. Conséquemment quatre paramètres sont essentiels à la compréhension de ce dilemme, à savoir : la *récompense*, l'*information*, les *valeurs* et l'*incertitude* (Mehlhorn et al., 2015).

#### **1) La récompense comme enjeu essentiel de la survie des êtres vivants**

De manière très simple, il est possible de considérer que le vivant cherche à maximiser les bénéfices qu'il peut tirer de son environnement et d'en limiter les dommages afin de survivre et de pérenniser son espèce. Le vivant est donc à la recherche de *récompenses*. La récompense constitue en quelque sorte l'objectif que le sujet (aussi appelé agent) cherche à atteindre par son choix : il s'agit du gain qu'il peut obtenir et qu'il cherche à maximiser en l'*exploitant* (nourriture par exemple).

#### **2) L'incertitude de l'environnement met en péril l'obtention de récompenses**

Les êtres vivants évoluent dans un environnement bruité, incertain. L'obtention de ces récompenses n'est donc pas aussi facile qu'il y paraît.

L'incertitude de l'environnement est liée à deux aspects. D'une part, l'environnement est *stochastique* ou *non déterministe* c'est-à-dire qu'il n'est pas fiable tout le temps. Choisir la même option plusieurs fois n'amènera pas toujours la récompense espérée : les résultats observés sont donc « bruités ». D'autre part, l'environnement est *volatile*, c'est-à-dire qu'il est changeant, et ce qui était vrai à un instant t peut devenir définitivement caduc à partir de l'instant t+1. La perception de l'agent de ces deux formes d'incertitude environnementale (*stochasticité* et *volatilité*) peut être respectivement regroupée sous deux catégories : l'incertitude attendue (*expected uncertainty*) versus l'incertitude inattendue (*unexpected uncertainty*) (Soltani and Izquierdo, 2019; Yu and Dayan, 2005). En effet, dans un environnement stable mais *stochastique*, le sujet s'attend à ce que les mêmes actions ne produisent pas toujours le même résultat. En revanche, lorsque l'environnement change, le

sujet ne s'attend pas à ce que les mêmes actions ne produisent plus jamais le même résultat. L'agent ne s'adapte donc pas de la même manière aux deux formes d'incertitude.

### 3) Les informations sont nécessaires à une recherche optimale des récompenses

Afin d'obtenir des récompenses, l'agent doit connaître son environnement et ses propriétés : quelles sont les récompenses à obtenir ? où sont-elles ? quels sont les efforts à fournir pour les obtenir ? l'environnement est-il stable ? ou au contraire change-t-il régulièrement ?... Toutes ces informations sont essentielles aux êtres vivants pour évoluer de manière adaptée dans un environnement donné et seule l'exploration de celui-ci permet de les obtenir. L'obtention d'informations permet également de réduire l'incertitude (Gottlieb et al., 2013).

### 3) Des valeurs sont associées aux différentes options pour guider les choix

L'ensemble des informations et des récompenses obtenues permettent au sujet d'attribuer aux différentes options possibles des valeurs. Ces valeurs permettent de quantifier numériquement la probabilité de chaque option à amener une récompense si celle-ci est choisie. Une valeur élevée signifie que l'option considérée a une forte probabilité d'amener une récompense (de meilleure qualité ou plus fréquente) : choisir les options avec les plus hautes valeurs consiste en une stratégie d'*exploitation*.

En fonction de la pondération de ces paramètres, le curseur ne placera donc pas au même endroit sur le continuum *exploration – exploitation* et différents types de stratégies sont donc envisageables. Dans un environnement totalement inconnu par l'agent, celui-ci devra donc apprendre à le connaître et présentera donc une stratégie plus *exploratoire*. L'*exploration* permet d'apporter de l'information. En revanche, si l'agent a évolué depuis longtemps dans un environnement stable, qu'il connaît bien et qu'il a été en mesure d'estimer toutes les valeurs des options possibles, l'agent a tout intérêt à exploiter ses connaissances afin de maximiser ses récompenses. Néanmoins, exploiter sans explorer peut être délétère. L'environnement étant *stochastique*, il peut être intéressant d'explorer de temps à autre afin de vérifier qu'une option n'est pas en réalité plus avantageuse qu'une autre par exemple : l'exploration permet donc également d'identifier les meilleures valeurs. De même, l'environnement étant *volatil*, la nécessité d'obtenir rapidement de nouvelles informations et de mettre à jour ses valeurs requiert probablement une transition très rapide entre *exploration* et *exploitation*.

### **1.1.3. ENJEUX SCIENTIFIQUES : LA BALANCE EXPLORATION – EXPLOITATION FACE A LA MULTIPLICITE ET LA COMPLEMENTARITE DE DIFFERENTES FORMES D'EXPLORATION**

Les travaux sur le fourrageage ont montré que les êtres vivants sont à la fois en quête de récompenses alimentaires mais qu'ils sont également prêts à rechercher des informations sur leur environnement. Cette recherche d'informations peut être envisagée dans le cadre de deux formes d'exploration : une exploration *structurée*, et une exploration *non structurée*.

A l'instar de l'animal qui finit par quitter le site d'acquisition alimentaire qu'il exploitait, afin de voyager et d'en trouver un autre (Charnov, 1976), l'*exploration structurée* prédit que l'agent alterne entre des phases d'*exploration active* et des phases d'*exploitation*, passant ainsi d'un état à un autre (Kolling and O'Reilly, 2018). Il a notamment été montré que chez des sujets humains, à la suite de changements de contingences de la tâche en cours, ceux-ci sont capables de passer quasi instantanément d'un mode *exploitatoire* à un mode *exploratoire* (Domenech et al., 2020). Alterner entre ces deux modes en fonction des changements de l'environnement peut s'avérer être efficace pour faire face à une *incertitude inattendue*.

En revanche, tout comme la grande mésange choisit 10 % du temps des options alimentaires non optimales, l'*exploration non structurée* prédit une co-existence des stratégies d'*exploration* et d'*exploitation*. L'idée est de considérer que l'incertitude peut guider l'*exploration* de deux manières qualitativement différentes : (1) en orientant le choix vers les options pour lesquelles l'agent n'a pas suffisamment d'informations, c'est l'*exploration dirigée*, (2) en ajoutant de l'aléatoire au comportement, c'est l'*exploration aléatoire ou non dirigée*. Alors qu'initialement certaines études ne semblaient pas mettre en évidence ces deux formes d'*exploration* (Daw et al., 2006), il est clair aujourd'hui que ces deux formes d'*exploration, non-dirigée (aléatoire)* et *dirigée*, coexistent chez l'humain (Gershman, 2018; Wilson et al., 2021, 2014b) et peuvent être combinées (Wilson et al., 2021).

Tout l'enjeu est désormais de comprendre comment les êtres vivants sont capables d'intégrer les différentes informations qu'ils reçoivent, d'estimer les valeurs associées à chacune des options auxquelles ils font face tout en adaptant leur choix à l'incertitude de l'environnement, pondérant ainsi ces stratégies d'*exploration* et d'*exploitation* afin de maximiser les récompenses qu'ils peuvent obtenir. L'apprentissage par renforcement est une première solution très simple permettant d'apporter certaines réponses.

## 1.2. L'APPRENTISSAGE PAR RENFORCEMENT ET LA BALANCE EXPLOITATION - EXPLORATION

### 1.2.1. L'APPRENTISSAGE PAR RENFORCEMENT<sup>3</sup> : UNE BOUCLE DE RETROCONTRÔLE DECISIONNELLE

Au sein du vivant, la réalisation de *comportements dirigés vers un but* impose nécessairement de prendre des décisions. La prise de décision par un agent peut s'envisager comme un processus séquentiel.

En premier lieu vient la *volition*, c'est-à-dire la définition de l'objectif à atteindre. Au sein du vivant, cet objectif consiste essentiellement en un évitement du mauvais et une recherche du bénéfice, la *récompense*.

Puis vient l'étape de *représentation*. Au cours de cette étape, il s'agit d'évaluer l'ensemble des états accessibles de l'environnement (états externes) et l'état dans lequel se trouve l'agent (états internes). Les états externes peuvent représenter des endroits physiques, des stimuli de différentes natures (visuels, auditifs...). Les états internes peuvent représenter le niveau de faim ou de peur d'un agent par exemple. L'ensemble de ces états définit *l'espace des états*. Lors de cette étape de *représentation*, il convient également de définir *l'espace des actions*, c'est-à-dire l'ensemble des actions ou choix possiblement réalisables par le sujet (Rmus et al., 2021).

Afin de prendre des décisions appropriées, l'agent doit attribuer une valeur à chaque action possible relativement aux états internes et externes. Ces valeurs doivent être des prédicteurs fiables des avantages qui sont susceptibles de résulter de chaque action. Cette étape est appelée *valuation*.

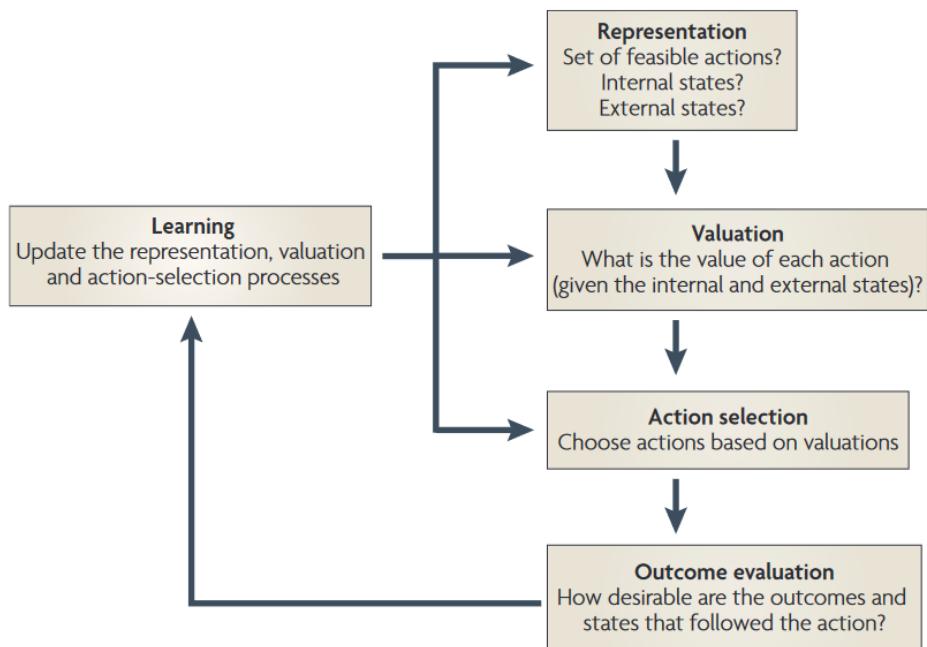
En comparant les valeurs ainsi attribuées pour chaque option, il est alors possible d'en sélectionner une. Cette étape consiste en la *sélection* de l'action et elle précède l'étape d'*exécution* de l'action à l'issue de laquelle les conséquences de l'action et donc de la décision sont révélées (Rangel et al., 2008; Volle and Levy, 2014).

L'*évaluation* finale du résultat obtenu à la suite de son/ses action(s) dans le *passé* permet l'ajustement des prises de décision dans le *futur* grâce à la mise à jour des représentations des différents couples (états – actions). Cet ajustement consiste en une

---

<sup>3</sup>Il existe diverses formes d'apprentissage, les plus simples étant l'habituuation, la sensibilisation ou encore le conditionnement. D'autres formes d'apprentissage existent encore, comme l'apprentissage supervisé et non supervisé. Néanmoins, dans le cadre de ma thèse et des problématiques considérées, je n'en discuterai pas.

véritable boucle de *rétrocontrôle* qui n'est autre que l'*apprentissage* dont la forme la plus simple est l'*apprentissage par renforcement* (**Figure 1**).



**Figure 1 : Cycle de la décision** (adapté de (Rangel et al., 2008))

Tout l'enjeu est de comprendre comment l'apprentissage par renforcement est le modèle le plus simple permettant d'expliquer la capacité des êtres vivants à évoluer de manière adaptée dans leur environnement, et comment ce modèle répond en partie au dilemme *exploration – exploitation*.

### 1.2.2. CONSIDERATION MATHEMATIQUE DE L'APPRENTISSAGE PAR RENFORCEMENT : DES PARAMETRES COMPUTATIONNELS AUX COMPORTEMENTS

#### LA TABLE DES Q-VALEURS

Comme nous l'avons discuté au paragraphe précédent, les êtres vivants se reposent sur des tables qui associent à chaque couple (état – action) une valeur subjective reflétant le gain à long-terme de ce couple. Cette table permet à l'agent de choisir sa *politique*, c'est-à-dire les actions à réaliser afin de maximiser les récompenses. À chaque couple (état ( $s$ ) – action ( $a$ )), à chaque instant  $t$ , est donc associée une valeur, dénommée *Q-valeur* (*Q-value*), que j'annoterai  $Q(s_t, a_t)$ . Le sujet doit les estimer par essai-erreur ces valeurs comme cela a été théorisé par Sutton et Barto (Sutton and Barto, 2018).

## MISE A JOUR DES Q-VALEURS

Ce modèle d'apprentissage a pour objectif l'optimisation quantitative des récompenses au cours du temps, en mesurant la différence entre le résultat attendu à un instant  $t$ , à savoir la valeur  $Q(s_t, a_t)$  et le résultat réellement obtenu,  $r$  après avoir réalisé l'action à ce même instant  $t$ . Cette différence  $\delta_t = (r_t - Q(s_t, a_t))$  est appelée *erreur de prédiction de la récompense* (*reward prediction error*) et permet de mettre à jour la valeur de l'action sélectionnée pour un état donné. A chaque nouveau résultat observé, l'agent est donc capable de mettre à jour ses estimations sur l'environnement et peut ainsi mettre en place des comportements adaptés à son environnement.

Cette différence est en néanmoins modulée par le taux d'apprentissage  $\alpha$  reflétant la vitesse à laquelle un nouveau résultat influence la mise à jour des valeurs. La mise à jours des valeurs des couples (état – action) suit donc la loi de Rescorla-Wagner (Rescorla, 1971) :

$$Q(s_{t+1}, a_{t+1}) = Q(s_t, a_t) + \alpha \cdot \delta_t = Q(s_t, a_t) + \alpha \cdot (r_t - Q(s_t, a_t))$$

Le *taux d'apprentissage* varie entre 0 (pas de mise à jour des valeurs,  $Q(s_{t+1}, a_{t+1}) = Q(s_t, a_t)$ ) et 1 (mise à jour uniquement selon le dernier résultat observé,  $Q(s_{t+1}, a_{t+1}) = r_t$ ). Il est à noter que le résultat  $r$  est souvent associé à une récompense, on parle de *renforcement positif*. Sa valeur est donc positive. Néanmoins, le vivant est tout aussi sensible aux résultats à valence négative. Le résultat  $r$  peut donc en effet être perçu comme une « punition » (pas de récompense voire résultat aversif), on parle alors de *renforcement négatif*. La valeur de  $r$  peut donc être nulle, voire négative.

Il est à noter que certains auteurs ajoutent dans le calcul de l'erreur de prédiction  $\delta_t$ , un paramètre de décompte temporel  $\gamma$  qui quantifie la propension de l'agent à miser sur des récompenses à long terme ( $\gamma$  proche de 1) où sur des récompense immédiates ( $\gamma$  proche de 0) (Khamassi et al., 2013; Schweigofer et al., 2006; Sutton and Barto, 2018)

## **SELECTIONNER UNE ACTION : EXPLOITATION OU EXPLORATIONS**

Apprendre les valeurs nécessite ensuite une sélection d'une action afin d'en observer son résultat. Au début, lorsque les valeurs sont inconnues pour chaque couple (état – action), il n'y a pas de raison de sélectionner une action de manière préférentielle, mais au fur et à mesure que les différentes actions sont testées relativement aux états considérés, et que les valeurs sont mises à jour grâce à l'observation des résultats, certaines valeurs seront alors plus élevées que d'autres. Il existe donc un véritable dilemme entre maximiser à court terme ses profits en exploitant les couples (état – action) avec les plus hautes valeurs et améliorer sa

connaissance de l'environnement en explorant des couples (état – action) dont les valeurs associées sont moins élevées.

En effet, plus la *Q-valeur* pour un couple (état – action) est élevée, plus l'action associée à cet état doit être sélectionnée car elle permet de maximiser la récompense. Choisir l'action avec la plus forte valeur garantit donc une probabilité de récompense très élevée : dans le cadre de l'*apprentissage par renforcement*, ce choix s'apparente à un comportement de type *exploitation*. D'un point de vue expérimental, on s'attend donc à observer une rigidité dans la séquence de choix réalisé par le sujet (toujours le même choix). Néanmoins sélectionner d'autres actions, dont les valeurs ne sont pourtant pas les plus élevées à un instant t, pourrait permettre de découvrir qu'une autre action peut avoir une meilleure.

*Explorer* est donc une solution, mais comment explorer ? Comme il l'a été précédemment discuté, la littérature distingue deux formes d'*exploration*. D'une part l'exploration *non-dirigée* ou *aléatoire*, durant laquelle l'incertitude liée aux options génère du « bruit » dans le modèle computationnel, induisant de ce fait une composante aléatoire dans le comportement : l'agent ne recherche donc pas activement de l'information pour améliorer sa connaissance de l'environnement. D'autre part l'*exploration* dite *dirigée*, durant laquelle l'incertitude liée aux options pousse l'agent à choisir les options les plus incertaines afin de mieux appréhender l'environnement dans lequel il évolue. Ces deux formes d'exploration ont pu être intégrées au modèle d'apprentissage par renforcement par différents calculs.

### 1) Exploration non-dirigée ou aléatoire

Une première manière *d'explorer* serait de choisir une action par la méthode dite de *softmax*, fondée sur une distribution de Gibbs – Boltzmann, permettant de faire varier la probabilité des actions en fonction des valeurs estimées pour chaque couple (état –action) possible. La probabilité de sélection de l'action *a* est calculée comme suit (Sutton and Barto, 2018) :

$$p(a/s) = \frac{\exp(Q(a,s) \cdot \beta)}{\sum_{i=1}^n \exp(Q(a_i, s) \cdot \beta)}$$

Dans cette équation,  $\beta$  est appelé *inverse de la température*  $\tau$  ( $\beta = 1/\tau$ ). Ainsi, plus  $\beta$  est petit (température élevée), plus le comportement tend à l'*exploration* des autres actions possibles puisque les probabilités deviennent proches les unes des autres. À l'inverse, plus  $\beta$  est grand (température faible), plus le comportement tend à l'*exploitation*. La sélection de l'action suivant ce modèle a pu être mise en évidence expérimentalement (Daw et al., 2006).

Certains auteurs proposent d'autres manières d'explorer. Cette *exploration* peut se faire de temps en temps, c'est-à-dire avec une faible probabilité  $\epsilon$  de choisir une action de manière indépendante des valeurs associées à chaque (état –action). C'est ce que l'on appelle une sélection  $\epsilon$ -greedy (Sutton and Barto, 2018).

D'autres formes plus sophistiquées d'*exploration aléatoire* ont été également proposées comme l'échantillonnage de Thompson (*Thompson sampling*) (Thompson, 1933). La valeur de chaque option est échantillonnée de manière aléatoire à partir de la distribution postérieure sur le gain moyen de chaque option. La moyenne de cette valeur est toujours égale à la récompense attendue, mais sa variabilité varie en fonction de l'incertitude de chaque option. Ainsi, l'*exploration aléatoire* dans l'échantillonnage de Thompson est maximale lorsque l'algorithme est incertain et égale à zéro lorsque la structure des gains est connue (Wilson et al., 2021).

## 2) Exploration dirigée

Une des approches les plus connues dans le cadre de l'*exploration dirigée* est l'algorithme de *limite de confiance supérieure* (*upper confidence bound*) qui consiste en l'ajout d'un bonus d'incertitude pour le choix d'une option qui a été peu choisie dans le passé et pour laquelle donc, l'agent dispose de peu d'informations quant aux récompenses qu'il peut obtenir en réalisant un tel choix (Auer and Cesa-Bianchi; Sutton and Barto, 2018). La probabilité de choisir une action  $a$  à chaque instant, choisie déjà  $N$  fois à cet instant associée à la valeur  $Q(a)$  se formalise donc comme suit :

$$p(a) = \operatorname{argmax} [Q_t(a) + c \sqrt{\frac{\ln(t)}{N_t(a)}}]$$

Le paramètre  $c$  contrôle le degré d'*exploration*. Si  $N_t(a) = 0$ , c'est-à-dire que l'option n'a jamais été choisie, alors  $p(a)$  sera maximale et ce sera donc cette option qui sera explorée.

### **1.2.3. ENJEUX SCIENTIFIQUES : QUID DE L'APPRENTISSAGE PAR RENFORCEMENT DANS LE CADRE D'UN ENVIRONNEMENT VOLATIL ?**

Les théories de l'*apprentissage par renforcement* permettent d'apporter une première réponse sur l'organisation des stratégies *exploratoire* et *exploitatoire* en les formalisant mathématiquement. Ce premier modèle simple permet de comprendre comment les êtres vivants sont capables d'apprendre et de se maintenir dans un environnement donné. Néanmoins, comme tout modèle, celui-ci a des restrictions (Dayan and Niv, 2008).

Notamment, dans la perspective de la question qui se pose, à savoir comment les êtres vivants s'adaptent à un changement abrupt de l'environnement, l'apprentissage par renforcement n'est pas le modèle permettant l'ajustement le plus rapide.

- 1) Ce modèle en effet est un modèle continu et aucune transition abrupte reflétant une transition environnementale abrupte n'est possible. Ce modèle ne permet notamment pas d'expliquer les processus d'*exploration structurée* de l'environnement. Lorsque l'environnement change brutalement en effet, les tables de valeurs pour chaque couple (état-action) ne sont plus valides mais ne sont pas réinitialisées abruptement. D'après le modèle, l'agent doit en effet dans un premier temps désapprendre progressivement les anciennes valeurs, puis ré-apprendre progressivement les nouvelles.
- 2) En fonction des paramètres du modèle, comme le *taux d'apprentissage*, apprendre par renforcement peut s'avérer être un processus long : s'adapter à un changement abrupt de son environnement peut donc prendre du temps, ce qui n'est pas optimal. Or dans le cas d'un modèle simple d'*apprentissage par renforcement*, les paramètres  $\alpha$ ,  $\beta$ ,  $\gamma$  et  $\varepsilon$  sont constants. Envisager une variation dynamique des paramètres du modèle semble donc plus appropriée dans le contexte de variations de l'environnement au cours du temps (Khamassi et al., 2013).
- 3) Enfin, l'*apprentissage par renforcement* se limite à découvrir et mettre à jour la *valeur subjective* des actions liés à des états. L'*apprentissage par renforcement* dans sa forme la plus simple se réduit donc à une comparaison de valeurs subjectives en y ajoutant de l'aléatoire ou des bonus d'incertitude. Le modèle d'*apprentissage par renforcement* est donc un modèle sans mémoire des stratégies utilisées par le passé et ne permet pas d'inférer la structure causale de son environnement (Gershman et al., 2015). L'*apprentissage par renforcement* peut donc être suppléé par des modèles qui sont capables d'estimer la structure cachée de l'environnement (relation entre certains états, existence de changements des règles au cours du temps...) afin d'assurer de manière optimale et rapide l'ajustement comportemental sur le continuum *exploration - exploitation*. C'est le cas des modèles bayésiens fondés sur le théorème de Bayes qui supposent que l'agent présente plusieurs hypothèses candidates sur son environnement qu'il met régulièrement à jour en fonction de ses *a priori* et des résultats observés.

Tout l'enjeu de la seconde partie de ce chapitre est donc de discuter de ces différents aspects et de proposer des extensions du modèle d'*apprentissage par renforcement* simple afin de comprendre comment les êtres vivants peuvent s'adapter le plus efficacement et le plus rapidement possible dans un environnement *volatil*.

## **2. EXPLORER ET EXPLOITER EFFICACEMENT : QUAND L'ADAPTATION COMPORTEMENTALE REQUIERT DES PROCESSUS COGNITIFS AU-DELA DU SIMPLE APPRENTISSAGE PAR RENFORCEMENT**

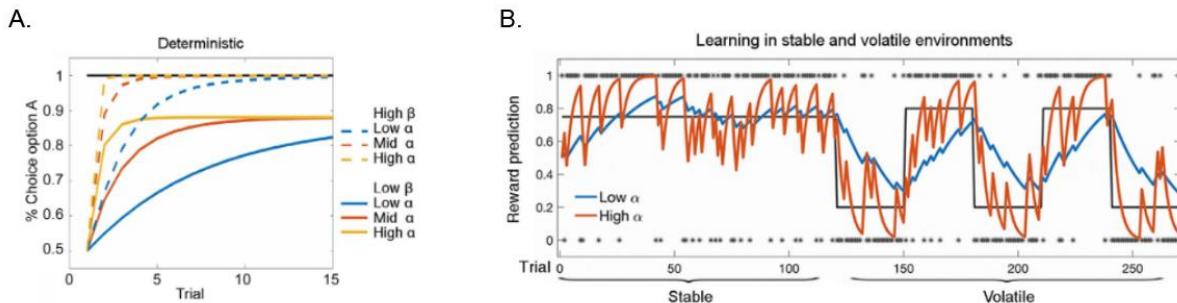
### **2.1. LE META-APPRENTISSAGE OU L'AJUSTEMENT DYNAMIQUE DES PARAMETRES D'APPRENTISSAGE**

#### **2.1.1. DES PARAMETRES AUX META-PARAMETRES : LES DETERMINANTS CRITIQUES DU COMPORTEMENT**

Le concept de méta-apprentissage par renforcement trouve sa définition dans les travaux de Doya en 2002 (Doya, 2002). Les paramètres d'apprentissage dont nous avons précédemment discuté sont des composantes critiques du modèle d'*apprentissage par renforcement* : dépendamment de leur valeur, l'apprentissage ne sera pas le même à la fois dans un environnement stable et dans un environnement volatil. Le comportement en résultant sera également donc très différent.

Par exemple, si l'inverse de la température est élevé, l'agent explore peu et dans le cadre d'une tâche déterministe, celui-ci finira par toujours choisir l'option qu'il juge optimale (**Figure 2A**, courbes pointillées). En revanche, si l'inverse de la température est faible, l'agent explore beaucoup et ne présentera donc jamais 100 % du temps le même choix (**Figure 2A**, courbes continues). De même, plus la valeur du taux d'apprentissage est élevée, plus l'agent atteint son plateau comportemental rapidement (**Figure 2A**, couleurs différentes des courbes). Par ailleurs, en fonction de la stabilité de l'environnement, un taux d'apprentissage élevé n'est pas approprié dans un environnement stable alors qu'il le sera dans un environnement très *volatil*, changeant régulièrement. En effet, si le taux d'apprentissage est élevé, les résultats observés à la suite des actions de l'agent auront une grande influence dans la mise à jour de la

table des valeurs et donc les valeurs de celles-ci varieront de manière importante à chaque nouveau résultat observé (**Figure 2B**).



**Figure 2 : Considération théorique de l'effet de deux méta-paramètres (taux d'apprentissage  $\alpha$  et inverse de la température  $\beta$ ) sur le choix des agents.** Adapté de (Scholl and Klein-Flügge, 2018). **A.** Cinétique de choix d'un agent en fonction de différentes valeurs des paramètres  $\alpha$  et  $\beta$  dans une tâche déterministe. Lorsque  $\beta$  a une faible valeur, l'agent présentera un comportement plus stochastique et ne choisira pas 100 % du temps la même réponse. Lorsque  $\alpha$  augmente, l'agent atteint plus rapidement le plateau de choix et donc sa séquence comportementale définitive. **B.** Estimation de l'obtention d'une récompense au cours du temps en fonction de différentes valeurs de  $\alpha$ . Dans un environnement stable, un petit  $\alpha$  permet d'ajuster au mieux ses choix pour maximiser la récompense. En revanche, lorsque l'environnement change brutalement, un taux d'apprentissage plus élevé permettra au sujet de s'adapter plus rapidement aux nouvelles contingences.

L'ajustement de ces paramètres étant donc critique, Doya considère qu'il s'agit de méta-paramètres (*metaparameters*) appelés aussi hyper-paramètres (*hyperparameters*) et qu'il existe au niveau cérébral des mécanismes assurant leur ajustement dynamique, probablement supportés par les neuromodulateurs cérébraux (Doya, 2002).

Luksys et collaborateurs ont notamment montré que le comportement de souris effectuant une tâche d'apprentissage simple (apprendre à placer son museau dans un trou après un signal lumineux afin de récupérer une récompense alimentaire) n'était pas le même en fonction de la lignée de souris étudiée ou en fonction du niveau de stress de l'animal. Les courbes de performance étant ainsi différentes, il est donc probable que le processus d'apprentissage sous-jacent diffère entre les conditions (Luksys et al., 2009). Les auteurs démontrent alors que le comportement peut être prédit par un modèle d'apprentissage par renforcement dans le cas où les méta-paramètres ( $\beta$ ,  $\gamma$  notamment) sont ajustés en fonction de la lignée de la souris, du stress des animaux ou encore du taux de noradrénaline. Leur tâche ne présentait pas en revanche de changement brutal des conditions expérimentales et il est donc possible d'envisager d'autres mécanismes permettant d'ajuster ces méta-paramètres.

## **2.1.2. FACTEURS D'AJUSTEMENT DES META-PARAMETRES EN ENVIRONNEMENT VOLATIL : ENTRE IMPORTANCE DE L'HISTORIQUE DES RECOMPENSES ET CADRE BAYESIEN**

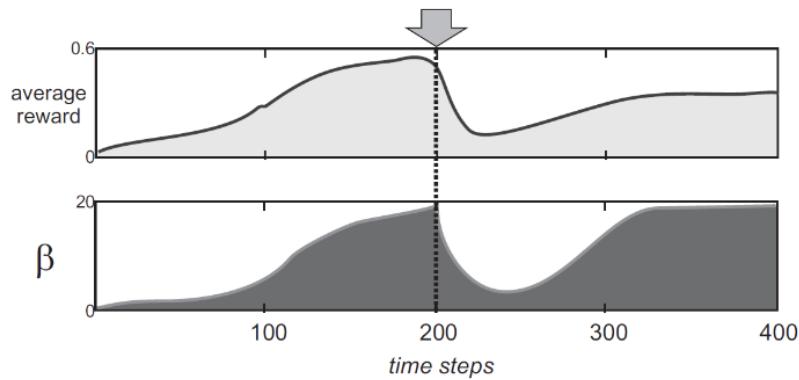
En cas de changement abrupt de l'environnement, plusieurs solutions peuvent être envisagées. L'agent peut notamment mettre à jour plus rapidement sa table des valeurs si le taux d'apprentissage  $\alpha$  augmente. De même, l'*exploration* d'autres options par une augmentation de l'*exploration aléatoire* (augmentation du paramètre  $\epsilon$  ou encore diminution de l'inverse de la température  $\beta$ ) est envisageable. Ces postulats sont illustrés par une récente étude de l'équipe d'Izquierdo et Soltani dans le cadre d'une tâche de renversement d'apprentissage : le taux d'apprentissage et le taux d'exploration sont plus élevés après le renversement que lors de la phase initiale d'acquisition (Harris et al., 2020)

L'ajustement des méta-paramètres sera donc d'autant plus efficace s'il suit une estimation de la *volatilité* de l'environnement. Cette estimation peut-être facilement réalisée grâce à l'historique des récompenses obtenues dans le temps, car si l'environnement change, il est très probable que les options autrefois récompensantes ne le sont plus. Une autre option envisageable est celle proposée par les modèles d'apprentissage bayésien qui estiment des statistiques de l'environnement afin d'optimiser les croyances sur celui-ci. Les paragraphes suivants permettent d'envisager les différents mécanismes mis en jeu.

### 1) Ajustement des méta-paramètres selon le strict suivi de l'historique des récompenses : extension du cadre d'apprentissage par renforcement

Le modèle de Pearce-Hall est un des premiers modèles à donner un mécanisme sur la modulation des méta-paramètres d'apprentissage dans ce cadre théorique. Il est à noter que leur modèle se fonde sur les travaux de conditionnement pavlovien qui évaluent l'occurrence d'une réponse conditionnée à la suite de l'association d'un stimulus conditionné à un autre stimulus non conditionné. Néanmoins, leurs conclusions peuvent s'étendre aux théories de l'*apprentissage par renforcement* puisque dans les deux cas ces modèles sont fondés sur une *erreur de prédiction*. Dans le modèle de Pearce-Hall, l'agent utilise la valeur absolue de l'erreur de prédiction afin de moduler le niveau d'attention accordé à une association. Plus l'erreur est importante, plus l'attention accordée à cette association est importante et donc l'apprentissage de l'association est élevé : (Pearce and Hall, 1980). La valeur de l'*erreur de prédiction* apporte donc une information.

Plus récemment, Schweighofer et Doya ont montré que plus la quantité de récompense moyenne augmente au cours d'une tâche, plus l'*exploration* diminue (et donc le paramètre  $\beta$  augmente). En revanche, dès que les conditions changent, la quantité de récompense moyenne chute et de manière conjointe l'inverse de la température diminue, favorisant ainsi un retour à l'*exploration* et donc la découverte de nouvelles règles afin de cumuler de nouveau des récompenses (Schweighofer and Doya, 2003) (**Figure 3**).



**Figure 3 : Simulation d'un algorithme de métaprendre.** Dans une condition donnée, l'inverse de la température augmente (le taux d'exploration diminue) au fur et à mesure que le sujet apprend et accumule de la récompense. En revanche, dès que les conditions changent (à 200 pas de temps, voir la flèche et la ligne en pointillée), l'inverse de la température diminue (le taux d'exploration augmente) conséquemment à la diminution des récompenses reçues. (Khamassi et al., 2013) adapté de (Schweighofer and Doya, 2003).

Khamassi et collaborateurs proposent quant à eux un modèle mixte d'ajustement de l'inverse de la température  $\beta$  dans le cadre du modèle classique d'apprentissage par renforcement (Khamassi et al., 2011, 2013). Le taux d'exploration est modulé selon deux aspects.

Dans un premier temps,  $\beta$  est modulé par un paramètre  $\beta^*$  reflétant l'historique des résultats des choix. Ce paramètre  $\beta^*$  est calculé à chaque instant en fonction des *erreurs de prédiction*, en pondérant différemment les *erreurs de prédiction* issues des erreurs et des succès. Lorsque l'agent commet des erreurs,  $\beta^*$  diminue et conséquemment,  $\beta$  diminue. L'agent tend donc à plus explorer en choisissant d'autres actions selon l'équation de Gibbs – Boltzmann (voir *Sélection de l'action*, Chapitre 2, 1.2.2).

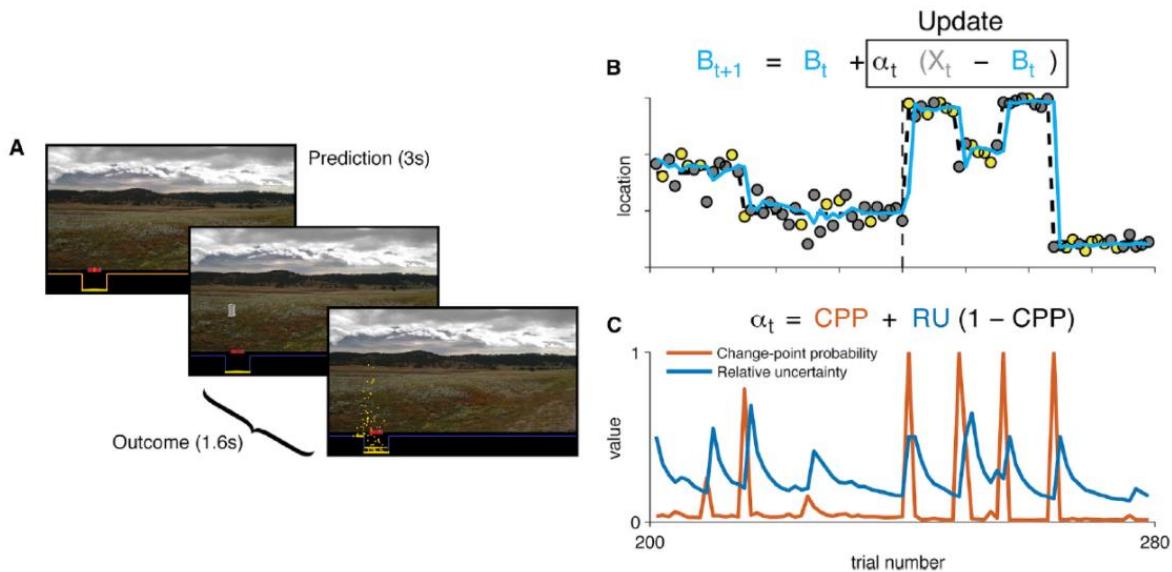
En parallèle, leur modèle calcule une métavaleur qui diminue à chaque fois que le taux moyen de récompense estimé diminue pour un état donné : lorsque cette valeur passe en-dessous d'un seuil, l'inverse de la température  $\beta$  est directement réinitialisé à sa valeur d'initialisation, à savoir une valeur faible permettant ainsi un retour immédiat à une exploration importante.

Ces deux modalités permettent donc d'ajuster dynamiquement le taux d'exploration, de manière fine en ajustant progressivement  $\beta$  grâce à  $\beta^*$  et de manière plus brutale en réinitialisant  $\beta$ . Ce deuxième mécanisme rend ainsi compte d'un ajustement comportemental efficace en regard de la *volatilité* de l'environnement puisqu'il assure une transition rapide d'un mode *exploitatoire* vers un mode *exploratoire* afin que l'agent puisse s'adapter rapidement à son nouvel environnement. Ce modèle est donc bien plus efficace qu'un simple modèle d'*apprentissage par renforcement* pour rendre compte de l'adaptation comportementale en environnement *volatil*.

## 2) Ajustement des méta-paramètres selon un modèle mixte : suivi de l'historique des récompenses et cadre bayésien

Une autre étude, menée par McGuire et collaborateurs, semble montrer que le taux d'apprentissage  $\alpha$  est modulé par d'autres facteurs. Dans leur étude, les participants doivent prédire la position d'un panier au sol dans lequel un sac, largué depuis un hélicoptère (non visible par le participant), doit tomber (**Figure 4A**). La plupart du temps, à chaque nouvel essai, la position de l'hélicoptère oscille légèrement autour d'une position initiale mais de temps en temps, cette position initiale change brutalement, et la position de l'hélicoptère oscillera de nouveau à partir de cette nouvelle position : ces deux aspects permettent donc de générer d'une part un environnement *stochastique*, et d'autre part un environnement *volatil*.

Tout comme un agent mettrait à jour à chaque essai la valeur des couples (état – action) dans le cadre de l'*apprentissage par renforcement* (voir *Mise à jour des Q-valeurs*, Chapitre 2, 1.2.2), les participants mettent à jour leur estimation de l'endroit où le panier devra être placé à l'essai suivant  $t+1$  compte-tenu de leur estimation de la position à l'essai  $t$  et de la position réelle observée à ce même essai (McGuire et al., 2014). L'*erreur de prédiction* entre la croyance sur la position à l'instant  $t$  ( $B_t$ ), et la position réelle observée ( $X_t$ ) est multipliée, comme dans le cas de l'*apprentissage par renforcement* par un taux d'apprentissage  $\alpha$  (**Figure 4B**). Ce taux d'apprentissage en revanche n'est pas fixe mais varie à chaque instant  $t$  en fonction de deux paramètres : d'une part la probabilité d'occurrence d'un *point de rupture* (CPP) et d'autre part l'incertitude relative (RU) (**Figure 4C**). Ces deux paramètres consistent respectivement en une estimation probabiliste de l'occurrence d'un changement brutal dans la position de l'hélicoptère (CPP) et en une estimation probabiliste de l'incertitude (RU) sur la position de l'hélicoptère dépendamment de la variance naturelle de cette position autour de la position initiale et de la survenue de changements brutaux.



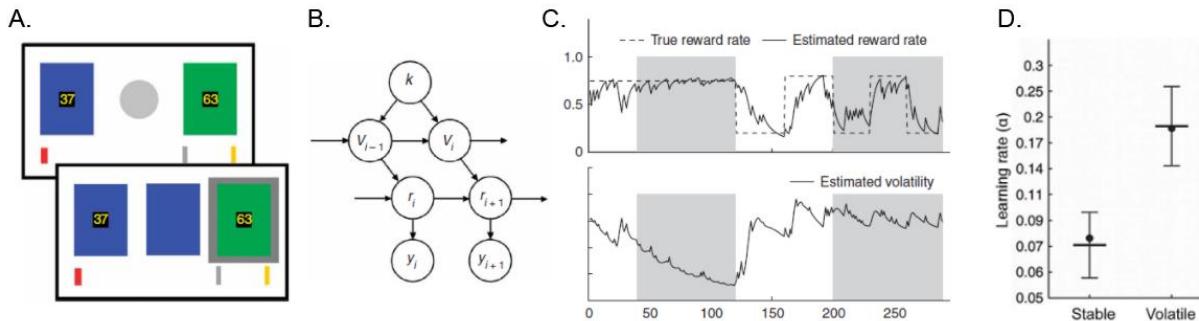
**Figure 4 : Ajustement des métaparamètres d'apprentissage selon un modèle quasi bayésien.**  
**A.** Procédure expérimentale. Les participants doivent positionner un panier (en jaune) dans lequel doit tomber un sac largué depuis le ciel (image du milieu et dernière image). **B.** La position du panier peut varier au cours du temps, légèrement autour d'une position donnée ou brutalement (courbe en pointillés noirs). Les participants doivent donc mettre à jour leur croyance ( $B_{t+1}$ ) après chaque essai en fonction de l'endroit où tombe le sac ( $X_t$ ), de leur croyance actuelle ( $B_t$ ) et du taux d'apprentissage ( $\alpha_t$ ). Les réponses des participants (correctes, point jaunes, incorrectes, point gris) suivent un modèle d'ajustement des croyances quasi bayésien. **C.** Ajustement du taux d'apprentissage en fonction de l'incertitude relative (RU) estimant l'*incertitude attendue* et de la probabilité d'occurrence d'un point de rupture (CPP) estimant l'*incertitude innatendue*. Adapté de (McGuire et al., 2014)

Dans cette étude en particulier, le taux d'apprentissage est donc modulé à chaque instant  $t$  par une estimation probabiliste des *incertitudes attendue* et *inattendue*, dépendamment des résultats observés précédemment. Ces valeurs évoluant constamment et de manière très ample lors d'un changement brutal de position de l'hélicoptère, le taux d'apprentissage change donc très rapidement et amplement, permettant un ajustement quasi immédiat des croyances du sujet sur la nouvelle position de l'hélicoptère.

### 3) Ajustement des métaparamètres selon un modèle bayésien normatif

Ce type d'ajustement a pu être mis en évidence par le travail de Behrens et collaborateurs (Behrens et al., 2007). Les auteurs ont administré une tâche de psychologie expérimentale probabiliste à des sujets humains au cours de laquelle ceux-ci devaient choisir entre un bouton bleu ou un bouton vert, l'un étant plus récompensant que l'autre en moyenne pendant un certain nombre d'essais (Figure 5A). Tout comme dans une tâche de renversement d'apprentissage, les contingences étaient inversées mais la tâche alternait entre deux conditions. Une condition stable durant laquelle la règle était la même pendant 120

essais, puis était inversée au bout de ces 120 essais en transition vers une condition très volatile durant laquelle la règle changeait tous les 30 à 40 essais (**Figure 5C haut**, traits pointillés).



**Figure 5 : Ajustement des métaparamètres d'apprentissage selon un modèle bayésien.**  
**A.** Procédure expérimentale : les sujets doivent choisir entre l'option bleue ou l'option verte, l'une étant plus récompensante que l'autre la plupart du temps (la tâche est en effet probabiliste).  
**B.** Description schématique du modèle hiérarchique bayésien utilisé pour estimer différentes probabilités (récompense  $r$ , volatilité  $v$ , et modification de la volatilité  $k$ ). Le modèle est hiérarchique au sens où un niveau d'estimation de probabilité est influencé par le niveau supérieur. Néanmoins, l'ensemble de l'estimation de ces probabilités repose sur les observations  $y$  réalisées au cours de la tâche (récompenses par exemple). **C.** Evolution de l'estimation du taux de récompense et de la volatilité par le modèle bayésien : les estimations suivent au mieux les paramètres imposés par la tâche (récompense réelle et volatilité). **D.** Les sujets modulent leur taux d'apprentissage en fonction de la volatilité de l'environnement. Plus la volatilité est élevée, plus le taux d'apprentissage est élevé.  
 Adapté de (Behrens et al., 2007; Scholl and Klein-Flügge, 2018).

Les auteurs ont développé un modèle hiérarchique qui estime trois grandeurs : la probabilité de récompense ( $r$ ), la volatilité de l'environnement ( $v$ ) et le taux de modification de la volatilité ( $k$ ), les transitions entre ces valeurs d'un essai  $t$  à un essai  $t+1$  étant affectée par le paramètre au niveau supérieur (**Figure 5B**). A chaque nouvelle observation du résultat d'un choix (récompense ou non), et en appliquant le théorème de Bayes, le modèle calcule les probabilités *a posteriori* de ces grandeurs à partir des probabilités *a priori* et met ainsi à jour ses estimations statistiques de l'environnement (**Figure 5C**).

Un tel modèle a permis aux auteurs de mettre en évidence que les humains dans le cadre de cette tâche sont capables d'ajuster leur taux d'apprentissage comme le ferait un observateur bayésien idéal en fonction de la *volatilité* de l'environnement. Dans la condition durant laquelle l'environnement est très volatil (renversement de contingences tous les 30 – 40 essais), le taux d'apprentissage est beaucoup plus élevé que dans la condition stable (120 essais avec la même règle) (**Figure 5D**). Néanmoins, l'inférence bayésienne est intractable (Cooper, 1990; Kwisthout et al., 2011), puisqu'en l'absence de contrainte sur l'espace des hypothèses et des observations on ne peut pas garantir une solution avec un calcul d'une durée

finie : un tel calcul est donc difficilement envisageable au niveau cérébral (Soltani and Izquierdo, 2019). D'autres études proposent des modèles alternatifs qui illustrent néanmoins elles aussi que le taux d'apprentissage augmente effectivement lorsque la *volatilité* de l'environnement est élevée (Silvetti et al., 2018).

### **2.1.3. ENJEUX SCIENTIFIQUES : QUID D'UNE ESTIMATION PARTIELLE DES PARAMETRES DE SON ENVIRONNEMENT ?**

La modulation des méta-paramètres permettant de mettre à jour plus rapidement ses tables de valeurs, cette approche est donc plus efficace dans un environnement *volatil*. Même si les modèles discutés s'avèrent tout à fait efficaces pour reproduire des données comportementales et estiment de manière plus ou moins directe la *volatilité* de l'environnement, certaines limites sont encore à noter.

Ce type de modèle n'a en effet pas de mémoire des stratégies passées. Or dans un environnement extrêmement volatil, où des régularités de changement s'observent fréquemment (inversion des contingences tous les 30 à 40 essais comme dans l'étude de Behrens et collaborateurs (Behrens et al., 2007)), comprendre cette structure cachée de l'environnement permettrait une adaptation encore plus rapide du comportement que la simple modulation du taux d'apprentissage. De plus, si certaines règles liant des états à des actions reviennent régulièrement, il peut être intéressant de les mémoriser également pour les réutiliser et où les modifier plus rapidement à la suite d'un nouveau changement.

Evoluer dans un environnement complexe nécessite donc probablement d'avoir des *a priori* plus forts sur celui-ci. Les animaux doivent donc être en mesure de comprendre et d'apprendre la structure cachée de l'environnement et de déduire les causes de ce qu'ils observent.

## **2.2. « TU AS ASSEZ D'INTELLIGENCE POUR DEVINER LE RESTE, PAR CE QUE JE TE LAISSE ENTREVOIR »**

Extrait du *Chef-d'œuvre inconnu* d'Honoré de Balzac, cette citation est à mon sens une illustration poétique des enjeux auxquels fait face le vivant. Tout comme l'artiste possède ce don inouï de voir ce que le commun des mortels n'envisage même pas, le vivant dispose de

cette capacité à aller bien au-delà de la simple observation de son environnement afin d'en obtenir une compréhension plus large.

### **2.2.1. L'ENVIRONNEMENT EST-IL PREVISIBLE ? ESTIMER LE PREVISIBLE, CONSIDERER L'IMPREVISIBLE**

Les êtres vivants ont tendance à présenter un comportement et des choix relativement structurés. Les humains ont du mal en effet à produire des séquences aléatoires si on leur demande (Sun et al., 2015). Au laboratoire, lorsque des rongeurs sont laissés libres de leur choix entre plusieurs options également récompensées ou confrontés à des compétiteurs faibles qui n'imposent pas aux rongeurs une grande variabilité dans les réponses données, ceux-ci ont tendance à présenter un comportement extrêmement structuré sous forme de motifs (Belkaid et al., 2020; Tervo et al., 2014).

Dans cette perspective, Sun et collaborateurs ont tenté de reproduire les choix de sujets humains dans une tâche de production de séquences aléatoires (Sun et al., 2015). Le modèle montre que les sujets ont une propension à l'alternance et que ce biais est dépendant de structures statistiques qui émergent naturellement de séquences aléatoires, comme le nombre de fois où un motif peut être observé ou encore le temps au bout duquel ce motif est rencontré. Les auteurs en concluent que le biais de structuration comportementale qui existe chez l'humain est en réalité lié à une estimation de la structure statistique de l'environnement intégré au sein des séquences aléatoires.

Pourtant, des études semblent montrer chez les rongeurs que ceux-ci sont capables de générer des choix stochastiques et extrêmement variables en s'affranchissant notamment de l'historique des résultats obtenus à la suite à leur choix (Belkaid et al., 2020; Tervo et al., 2014). A la différence de l'étude de Tervo et collaborateurs qui mettent en compétition leurs rats avec un compétiteur algorithmique qui les force à choisir entre deux options de manière la plus aléatoire et non structurée possible, l'étude de Belkaid est déterministe et la manière optimale de résoudre la tâche serait d'en apprendre la structure cachée. Or dans cette dernière étude, les souris résolvent la tâche en ayant un comportement pseudo-aléatoire qui leur permet néanmoins d'obtenir 75 % de performance en n'apprenant pas la structure cachée de la tâche et en générant elles-mêmes leur propre variabilité comportementale. Néanmoins, dans ces deux études la complexité inhérente aux tâches est telle qu'une approximation plutôt appropriée serait bien de prévoir que l'environnement est imprévisible et d'adapter en conséquence son comportement à ce haut niveau d'incertitude. Les animaux sont donc

capables d'ajuster de manière adaptive le caractère aléatoire de leur comportement en fonction de la perception qu'ils ont de leur environnement.

A mon sens donc, l'ensemble de ces résultats montre donc que les êtres vivants sont capables d'extraire des règles et informations de l'environnement et de les prendre en considération pour ajuster au mieux leur comportement. Certains auteurs affirment ainsi que l'estimation et la considération de la structure cachée de l'environnement via un apprentissage structuré peut permettre un ajustement plus rapide de son comportement (Braun et al., 2010; Gershman and Niv, 2010). Les travaux de Gallistel montrent notamment une capacité des rats *d'apprendre à apprendre* et de présenter des ajustements comportementaux beaucoup plus rapides une fois que les animaux ont été exposés plusieurs fois à des changements répétés (Gallistel et al., 2001).

## 2.2.2. INFERER LE CHANGEMENT ET S'Y ADAPTER

### 1) L'apport des tâches de renversement d'apprentissage dans le cadre des inférences sur les états cachés de l'environnement

Un grand nombre de données expérimentales allant dans le sens d'une connaissance des états cachés de l'environnement et de la capacité des êtres vivants à inférer les changements qui peuvent survenir provient d'études utilisant la tâche d'apprentissage par renversement, ou d'études dérivées. L'avantage de cette tâche est en effet de proposer un paradigme expérimental simple où des renversements de contingences expérimentales peuvent survenir régulièrement, introduisant donc une structure claire dans la tâche :

- 1) La règle peut changer n'importe quand
- 2) Si une option n'est plus récompensante, c'est l'autre qu'il faut désormais choisir

Wilson et collaborateurs en reprenant les résultats d'une étude d'apprentissage par renforcement chez le singe proposent un modèle à deux états. Au lieu de comparer des valeurs de couples (état-action), le modèle suit la structure de la tâche à savoir « les récompenses sont obtenues en répondant à droite » ou « les récompenses sont obtenues à gauche » (Wilson et al., 2014a). Ainsi, suivant ce modèle, les animaux doivent déduire l'état actuel de la tâche en utilisant les actions et les résultats de plusieurs essais précédents (Gershman et al., 2010). Les états ainsi définis correspondant à des caractéristiques détectables de l'environnement, ils rendent les problèmes plus simples à résoudre et permettent ainsi aux animaux de s'adapter plus rapidement.

Costa et collaborateurs quant à eux ont sur-entraîné des singes dans une tâche de renversement d'apprentissage probabiliste : les singes s'attendent donc à ce que des renversements aient lieu. Les auteurs montrent que contrairement à une transition comportementale graduelle tel que le prédirait un modèle *apprentissage par renforcement* classique, les singes changent très rapidement leur réponse du stimulus anciennement récompensé vers le stimulus nouvellement récompensé. Le comportement des animaux est alors très bien expliqué par un modèle bayésien dont une des hypothèses concerne la connaissance des statistiques de la tâche expérimentale par les animaux (Costa et al., 2015).

Plus récemment, les travaux de Bartolo et collaborateurs ont montré des résultats tout à fait similaires. Les auteurs ont également administré une tâche de renversement d'apprentissage à des singes, en ajoutant un degré de complexité supplémentaire. La tâche alternait entre deux conditions : une condition durant laquelle l'option récompensante était déterminée par la nature du stimulus (le stimulus bleu est récompensant majoritairement pendant un certain nombre d'essai, puis le renversement de contingences implique alors que le stimulus rose est récompensant) et une condition durant laquelle l'option récompensante était déterminée par la position du stimulus (le côté gauche est récompensant majoritairement pendant un certain nombre d'essai, puis le renversement de contingences implique alors que le côté droit est maintenant récompensant). Dans les deux conditions, les conclusions sont les mêmes : les transitions comportementales sont abruptes et ni un modèle simple d'*apprentissage par renforcement*, ni un modèle de Pearce-Hall faisant varier le taux d'apprentissage ne peuvent expliquer les résultats car ces modèles prédisent un renversement plus lent des préférences des animaux lors de la simulation des données. Seul un modèle bayésien testant des hypothèses sur les états cachés de la tâche parvient à les reproduire. Une fois surentraînés, les animaux n'apprennent plus la structure de la tâche mais utilisent plutôt le modèle acquis pour effectuer des inférences et réaliser leurs choix (Bartolo and Averbeck, 2020).

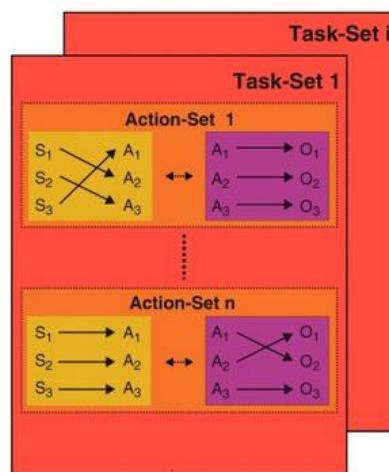
De manière similaire, dans une tâche de fourrageage qui s'apparente grandement à une tâche de renversement (à chaque instant  $t$ , seul un site de distribution alimentaire est actif, l'autre non, puis au moment du renversement l'ancien site devient inactif et l'autre devient actif), les souris présentent un comportement expliqué par la réalisation d'inférences sur les états cachés de la tâche. L'implémentation d'une expérience similaire chez les humains aboutit aux mêmes conclusions (Vertechi et al., 2020).

Néanmoins, une des grandes limitations des tâches par renversement est liée à la présentation de deux options, l'une étant correcte, l'autre non. Les inférences sur les états

cachés sont donc réduites à la simple détection des changements de contingences et à l'inversion stricte du comportement jusqu'alors mis en jeu. Or l'environnement dans lequel les êtres vivants évoluent est plus complexe, nécessitant une part d'*exploration* non négligeable et d'autres mécanismes sont donc probablement mis en jeu.

## 2) Task-set et représentation mentale complexe de son environnement

Une manière encore plus complexe de comprendre et estimer la structure de son environnement reposeraient donc sur l'existence de « modules de tâche », en anglais *task-set*. Un *task-set* consiste en une représentation abstraite d'une tâche qui permet de réaliser celle-ci de manière appropriée (Monsell, 2003; Sakai, 2008). Un *task-set* consiste donc en une représentation mentale très aboutie de l'environnement puisqu'il y regroupe un ensemble de règles comportementales associant différents états / indices / stimuli à différentes actions possibles ainsi que les résultats attendus pour une action donnée (**Figure 6**). Plusieurs *task-sets* peuvent coexister mentalement et en mémoire. Ces *task-sets* sont donc critiques dans le cadre d'une adaptation comportementale efficace puisque d'anciens *task-sets* peuvent être réutilisés dans le cas où une même situation est de nouveau rencontrée. D'autres *task-sets* peuvent émerger en cas de nouvelle situation jamais rencontrée auparavant (Domenech and Koechlin, 2015).



**Figure 6 : Structure d'un task-set.** Les *task-sets* sont des représentations mentales sous forme d'ensemble mnésique regroupant les règles (à savoir quelle action A est associée quel état S pour quel résultat O) en cours et celles utilisées précédemment mais non actuellement mises en jeu. (Domenech and Koechlin, 2015)

Toute la question est de savoir comment ces *sets* sont mis en place et mémorisés et de comprendre comment les agents peuvent passer de l'un à l'autre (en anglais « *switcher* »), et/ou de les modifier en fonction des différentes situations qu'ils rencontrent au gré des

changements de l'environnement. Différents modèles existent comme ceux proposés par Yu et Dayan (Yu and Dayan, 2005), ou encore le modèle dérivé du *Multiple Model-Base Reinforcement Learning (MMBRL)* de Doya mais c'est le modèle PROBE qui apportent les mécanismes les plus complets (Collins and Koechlin, 2012; Donoso et al., 2014).

Dans le cadre de ce modèle PROBE, les agents se reposent sur l'utilisation de *sets* liant des états à des actions. A un instant donné, parmi l'ensemble des *sets* maintenus en mémoire, seul le *set* jugé le plus fiable, c'est-à-dire celui dont la mise en œuvre permet d'obtenir les résultats attendus, est en cours d'utilisation. L'algorithme est considéré comme étant en mode *exploitatoire*. Lorsque l'environnement change et que les règles deviennent obsolètes, les résultats n'étant plus ceux attendus, l'agent est alors confronté à plusieurs options pour s'adapter : maintenir le *set* en cours et l'adapter progressivement par *apprentissage par renforcement*, changer de *set* et utiliser un de ceux stockés en mémoire si la fiabilité estimée de celui-ci est alors élevée dans le nouveau contexte ou *explorer* afin de créer un nouveau *set* en mixant les *sets* stockés en mémoire. L'algorithme finit par retourner en mode *exploitatoire* si la fiabilité du nouveau *set* est satisfaisante.

Ce modèle s'avère particulièrement efficace pour reproduire des données comportementales humaines. Domenech et collaborateurs ont mis en place une tâche de psychologie expérimentale où des sujets humains doivent associer des stimuli visuels à des actions (Domenech et al., 2020). A chacune des trois figures visuelles présentées (ici des chiffres) est associé un unique bouton de réponse parmi quatre, le sujet étant informé après son choix de la rectitude ou non de celui-ci. Au cours de la tâche, les associations changent régulièrement, brutalement et de manière imprévisible de sorte qu'il n'y ait aucun chevauchement d'association (stimulus – action) d'un épisode à l'autre. Trois types de réponses sont observables : des réponses correctes si le sujet choisi l'action effectivement associée à la figure, des réponses persévératives qui associent une figure à l'action qui y était associée dans l'épisode précédent, et des réponses auxiliaires qui ne sont ni correctes ni persévératives mais le reflet de l'*exploration* des différentes associations possibles entre un stimulus donné et les quatre boutons disponibles. L'étude montre que les participants changent brutalement de stratégie comportementale entre deux à huit essais suivant le changement de règles imposés par la tâche, les réponses persévératives chutant brutalement tandis que les réponses auxiliaires augmentent brusquement. Comme prédit par le modèle, les participants changent abruptement d'un mode *exploitatoire* vers un mode *exploratoire* lorsque la fiabilité du *set* en cours est insuffisante, observations non prédictes par un simple *apprentissage par renforcement*.

### **2.2.3. ENJEUX SCIENTIFIQUES : COMMENT RECONCILIER CE FOISONNEMENT DE THEORIES ?**

Les travaux présentés précédemment apportent des preuves indéniables sur la capacité des organismes à s'adapter plus rapidement et efficacement qu'un simple modèle d'*apprentissage par renforcement*.

Néanmoins il est tout de même intéressant de noter que même dans les modèles les plus complexes (comme celui des *task-sets*) on y retrouve la trace de modèles plus simples comme celui de l'*apprentissage par renforcement*. Ceci illustre donc la très forte intrication des différents algorithmes et la possibilité d'envisager encore d'autres modèles mixtes comme un modèle combinant méta-apprentissage et *task-set*. Ce travail théorique est actuellement réalisé par Sami Beaumont, doctorant de l'équipe dans lequel j'ai effectué ma thèse.

Par ailleurs, comme illustré dans les paragraphes précédents, l'étude des *comportements adaptatifs* est menée chez différentes espèces, de la souris à l'homme. Néanmoins, les expériences réalisées sont souvent différentes en fonction de l'espèce et encore peu d'études strictement translationnelles, impliquant la réalisation d'une même tâche de psychologie expérimentale chez différentes espèces, ont été menées. Or l'implémentation de telles études permettrait d'envisager d'un point de vue évolutif les différents mécanismes algorithmiques mis en jeu en fonction des espèces dont les structures cérébrales sont plus ou moins complexes, permettant ainsi d'apporter des réponses essentielles sur la compréhension de ces structures et de leur rôle au cours de l'évolution. Tout l'enjeu serait donc également de proposer un paradigme dont la mise en œuvre chez différentes espèces pourrait apporter de telles réponses.

## **CONCLUSION**

Ce deuxième chapitre a permis d'évaluer les stratégies comportementales mises en jeu en cas de changement abrupt de l'environnement. L'adaptation comportementale repose sur un continuum *exploration – exploitation* dont la mise en œuvre dépend de computations algorithmiques cérébrales dérivées et complexifiées à partir du modèle d'*apprentissage par renforcement*. Le dernier chapitre de cet état de l'art s'intéressera désormais à évaluer la validité biologique de tels modèles en considérant les bases neuronales des comportements adaptatifs.

# CHAPITRE 3. La neurobiologie des comportements adaptatifs : des neurones au comportement

---

## INTRODUCTION

« *J'étais en train de devenir un véritable psychanalyste : j'écoulais les pensées profondes et enfouies de mon écrevisse.* »

Eric Kandel, Prix Nobel de Physiologie et de Médecine, 2000

*A la recherche de la mémoire, Une nouvelle théorie de l'esprit*, édition Odile Jacob, 2007, p.116

Ces propos de Kandel dans son autobiographie font référence à la fascination et l'exaltation qu'il a ressenties lors des premiers enregistrements électrophysiologiques de neurones d'écrevisse. Cette anecdote révèle de manière imagée et humoristique l'ambition de tout neuroscientifique. « Comprendre comment le fonctionnement de l'esprit et du comportement émerge de la biologie et de l'interaction entre les cellules et les molécules » est aujourd'hui la question à laquelle tente de répondre une grande partie des travaux publiés chaque jour (Gomez-Marin and Mainen, 2016).

L'étude des comportements adaptatifs, des comportements d'exploration et d'exploitation, les phénomènes d'apprentissage et d'inférence des états cachés de l'environnement n'échappent pas à cette ambition. Une fois la caractérisation comportementale de ces phénomènes établie, l'investigation de leurs bases neurobiologiques semble être la suite logique. De même la conceptualisation théorique sous formes de modèles computationnels doit être validée d'un point de vue biologique (D'Angelo et al., 2013).

Après une brève présentation des outils techniques et méthodologiques dont disposent les chercheurs aujourd'hui pour étudier l'activité cérébrale et son lien avec le comportement, ce troisième chapitre s'attachera à présenter les bases neurobiologiques identifiées actuellement comme support de l'*adaptation comportementale*. Par bases neurobiologiques j'entends à la fois aires cérébrales, circuits neuronaux, neurones et phénomènes de neuromodulation physiologique. Ces bases neurobiologiques seront mises en lien avec les comportements observés expérimentalement et les modèles algorithmiques présentés au cours du Chapitre 2.

## **1. ETUDE DE LA NEUROBIOLOGIE DES COMPORTEMENTS ADAPTATIFS : CORRELATION N'EST PAS CAUSE**

### **1.1. TECHNIQUES D'ETUDE DESCRIPTIVES DE L'ACTIVITE CEREBRALE LIEE A UN COMPORTEMENT**

#### **1.1.1. OBSERVER L'ACTIVITE CEREBRALE**

De nombreuses techniques existent aujourd’hui au laboratoire pour étudier l’activité cérébrale *in vivo*.

Chez l’Homme, les techniques les plus couramment employées sont l’electroencéphalographie ou encore l’imagerie par résonance magnétique fonctionnelle. D’autres techniques comme la magnétoencéphalographie, la tomographie par émission de positrons ou encore la spectroscopie infrarouge peuvent être utilisées (Xue et al., 2010).

Certaines de ces techniques ne permettent d’avoir qu’une image indirecte de l’activité cérébrale. L’imagerie par résonance magnétique fonctionnelle, par exemple, mesure des variations de flux sanguins (Andreelli and Mosbah, 2014) : cette technique se fonde donc sur la mesure d’un proxy physiologique pour mesurer l’activité cérébrale. D’autres techniques quant à elles, permettent d’obtenir une image mésoscopique de l’activité cérébrale. La magnétoencéphalographie mesure par exemple le champ magnétique induit par l’activité électrique cérébrale (D’Hondt and Sequeira, 2013). Le signal électrique à la base de l’electroencéphalogramme quant à lui résulte de la sommation des potentiels post-synaptiques d’une large population de neurones (Xue et al., 2010) : l’origine spatiale des sources neurales est donc peu précise. Ces technologies même si très informatives sont donc complémentaires d’autres techniques expérimentales.

Dans des cas très particuliers, il peut être en effet possible d’enregistrer l’activité neuronale grâce à des microélectrodes permettant d’obtenir des données plus précises comme des données d’activité de neurones individuels ou des potentiels de champs locaux de régions plus ciblées (Chari et al., 2020). Néanmoins ces études se limitent à des patients nécessitant une implantation médicale d’électrodes comme par exemple les patients atteints d’épilepsie (Chari et al., 2020) ou de troubles obsessionnels compulsifs (Mallet et al., 2008).

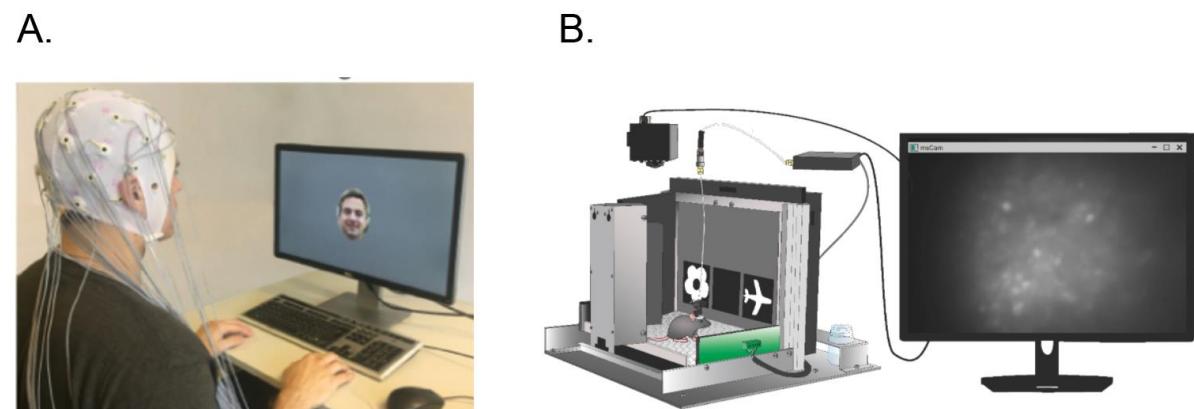
Du fait des contraintes éthiques et expérimentales qu’impose le travail de recherche avec des humains, il n’est donc pas toujours possible d’accéder à l’enregistrement unitaire de l’activité neuronale. Chez les modèles animaux en revanche, il existe une plus grande

diversité de techniques permettant d'enregistrer l'activité neuronale. Dans la limite de l'éthique et du respect du bien-être animal, il est en effet possible d'utiliser des méthodologies plus invasives pour enregistrer l'activité d'une population ciblée de neurones. L'utilisation d'électrodes d'enregistrement plus ou moins complexes (tétrodes, neuropixels, sonde de silicium (Hong and Lieber, 2019)) ou encore d'outils d'imagerie calcique (Grienberger and Konnerth, 2012) permet de suivre en temps réel, à la fois de manière individuelle ou populationnelle, l'activité d'un neurone ou de plusieurs centaines de neurones dans une région d'intérêt donnée.

### 1.1.2. ENREGISTRER SIMULTANEMENT COMPORTEMENT ET ACTIVITE CEREBRALE

Au laboratoire, afin de comprendre comment les comportements sont générés, une première approche consiste en l'enregistrement de l'activité cérébrale au cours de l'administration d'une tâche comportementale.

Chez les humains, les sujets sont le plus souvent face à un écran sur lequel une ou plusieurs figures visuelles s'affichent. Le sujet peut le plus souvent donner ses réponses en appuyant sur des boutons ou un clavier. Tandis qu'il réalise la tâche, l'activité neuronale est enregistrée de manière simultanée par une ou plusieurs des différentes techniques citées précédemment (Warbrick et al., 2014) (**Figure 1A**).



**Figure 1 : Protocoles expérimentaux d'enregistrement de l'activité cérébrale lors de la réalisation d'une tâche comportementale.** **A.** Enregistrement électroencéphalographique chez un sujet humain réalisant une tâche sur un ordinateur (Davis et al., 2020). **B.** Enregistrement de l'activité de neurones par imagerie calcique chez une souris libre de ses mouvements, réalisant une tâche dans une chambre opérante équipée d'écrans tactiles (Mosser et al., 2020).

Le travail avec des animaux comme les rongeurs ou les primates non humains laisse une plus grande liberté dans la réalisation des expériences. Les paradigmes expérimentaux sont plus variés, impliquant des modalités sensorielles différentes (visuelles - (Hvoslef-Eide et al., 2016) - , olfactives ou tactiles (Brown and Tait, 2010)...), et des modalités de réponse différentes (écrans tactiles, levier, *nosepoke*...). Dans tous les cas, l'activité cérébrale peut être enregistrée en même temps que les animaux réalisent la tâche (**Figure 1B**).

Certaines études emploient notamment des systèmes où les animaux sont immobilisés face au dispositif expérimental tandis que l'activité neuronale est enregistrée (Grossman et al., 2020). Ces systèmes permettent de contrôler au mieux les stimuli sensoriels et les réponses motrices des animaux. De plus, ils permettent de limiter dans les enregistrements les artéfacts et le bruit générés par les mouvements des animaux ou d'autres facteurs (Schwarz et al., 2010). D'autres expérimentateurs quant à eux font le choix de laisser les animaux libres de leurs mouvements durant la réalisation de la tâche et de l'enregistrement de l'activité cérébrale grâce à la conception d'implants qui peuvent parfois être sans fil (Fan et al., 2011; Mosser et al., 2020). Contrairement aux études « têtes fixées », ce type de techniques, plus écologiques et respectueuses de la physiologie de l'animal, limite le stress de celui-ci, point essentiel dans toute étude comportementale (Juczewski et al., 2020).

### **1.1.3. ENJEUX SCIENTIFIQUES : COMMENT LIER ACTIVITE CEREBRALE ET COMPORTEMENT ?**

Bien qu'informative, l'enregistrement de l'activité cérébrale présente des limites : limites de résolution spatiale, limites de résolution temporelle, limites quantitative du nombre de neurones enregistrés (Hong and Lieber, 2019). Par ailleurs, observer que l'activité d'une population de neurones ou d'une aire cérébrale augmente lors de la réalisation d'un comportement permet seulement de conclure que ces neurones ou cette aire cérébrale sont impliqués dans la réalisation de comportement : il est en revanche impossible de conclure qu'ils sont causatifs de ce celui-ci (Genon et al., 2018).

D'autres outils méthodologiques sont donc nécessaires pour établir un lien de causalité forte entre activité neuronale et comportements.

## **1.2. TECHNIQUES D'ETUDE CAUSATIVES DE L'ACTIVITE CEREBRALE LIEE A UN COMPORTEMENT**

### **1.2.1. LES CAS CLINIQUES : CE QUE NOUS APPRENNENT LES ETUDES DE PATIENTS**

L'étude des patients atteints de lésions permet d'approcher dans un premier temps le rôle de certaines aires cérébrales (Genon et al., 2018). Les cas historiques de Phineas Gage (O'Driscoll and Leach, 1998) ou encore du patient H.M (Corkin, 1984; Squire, 2009) ont permis respectivement d'entrevoir les fonctions du cortex frontal et de l'hippocampe.

Il est possible alors d'impliquer des patients présentant des lésions d'aires d'intérêt dans des protocoles expérimentaux. En comparaison à des sujets sains non lésés, les patients peuvent réaliser des tâches de psychologie expérimentale et leurs résultats comportementaux sont alors étudiés (di Pellegrino et al., 2007).

Le paradigme expérimental est simple : si les patients présentent un déficit dans la tâche reflétant un déficit moteur et / ou cognitif, il est très probable que l'aire lésée est normalement impliquée dans le processus étudié.

Néanmoins, les cohortes de patients sont souvent compliquées à établir car les lésions ne sont pas nécessairement homogènes et il n'existe pas toujours suffisamment de personnes atteintes des lésions affectant une zone cérébrale d'intérêt pour établir une cohorte. De plus, les phénomènes de plasticité cérébrale qui peuvent compenser certains déficits associés aux lésions sont impossibles à contrôler et peuvent donc biaiser les résultats des études.

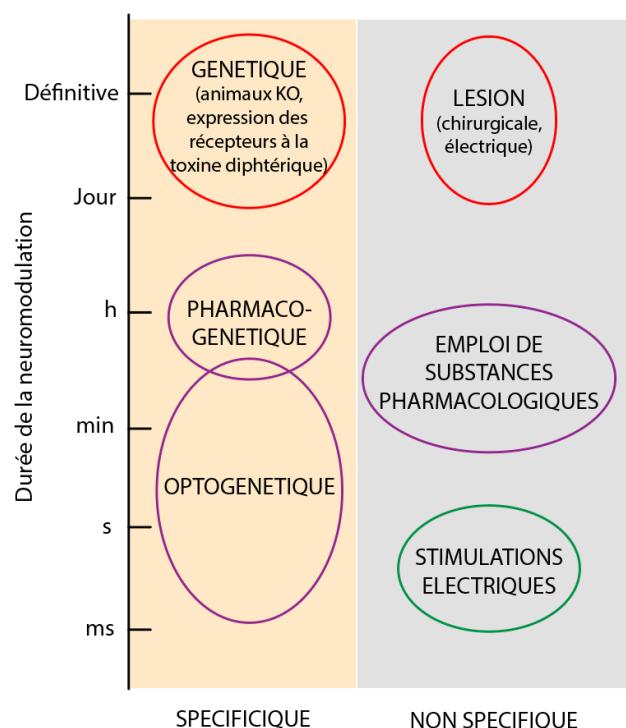
### **1.2.2. MANIPULER LE COMPORTEMENT EN CONTRÔLANT LES NEURONES : LA NEUROMODULATION EXPERIMENTALE**

Afin de pallier les difficultés précédemment évoquées, les scientifiques ont développé de nombreuses techniques visant à manipuler de manière contrôlée l'activité neuronale afin de comprendre son implication fonctionnelle : c'est ce qu'on appelle la neuromodulation expérimentale (Krug et al., 2015; Weber and Dan, 2016).

Ces techniques, visant à disséquer les circuits neuronaux pour en comprendre les fonctions et les conséquences comportementales, sont utilisées chez les animaux modèles et notamment chez la souris (Nakajima and Schmitt, 2020). Néanmoins on peut citer la stimulation magnétique transcrânienne chez l'Homme qui peut altérer de manière totalement

réversible l'activité d'une aire cérébrale par induction magnétique (Valero-Cabré et al., 2011). La stimulation cérébrale profonde à visée thérapeutique chez des patients implantés peut également être utilisée comme outil de neuromodulation expérimentale (Cavanagh et al., 2011). Ayant travaillé avec des souris au cours de ma thèse, je vais de ce fait détailler essentiellement les techniques employées dans le cadre d'études expérimentales impliquant des animaux.

Il existe une très grande diversité méthodologique et les multiples techniques utilisées se caractérisent par des propriétés différentes tant d'un point de vue de leur spécificité que de leur durée d'action ou encore des effets qu'elles produisent sur le système nerveux (gain ou perte d'activité) (**Figure 2**) (Weber and Dan, 2016).



**Figure 2 : Différentes techniques de neuromodulation expérimentale et principales caractéristiques temporelles, d'action et de spécificité.** Cercle rouge : technique inactivant les neurones ou l'aire cérébrale. Cercle vert : technique d'activation. Cercle violet : technique permettant l'activation ou l'inhibition. Adapté de (Weber and Dan, 2016)

- 1) Techniques non spécifiques : lésions chirurgicales, pharmacologie à spectre large, stimulations électriques, cryoinactivation...

Parmi les techniques les moins spécifiques, on peut citer les lésions chirurgicales qui consistent en l'ablation de régions cérébrales complètes soit par incisions chirurgicales « classiques » soit par utilisation d'ultrasons (Bretsztajn and Gedroyc, 2018) : ces lésions irréversibles induisent une perte d'activité à long-terme. Néanmoins, elles sont de moins en

moins utilisées car les problèmes de spécificité empêchent de conclure quant à l'implication de la région dans le comportement. En effet, dans cette technique l'aire cérébrale mais également les fibres de projection autour de cette région sont aussi altérées, ce qui ne permet pas d'éliminer l'hypothèse qu'une autre aire cérébrale peut être impliquée dans les comportements investigués. De plus ces techniques ne permettent pas de cibler spécifiquement une population de cellules en particulier.

L'injection de composés pharmacologiques, comme le muscimole (agoniste des récepteurs GABA-A, inhibant donc l'activité des neurones qui possèdent ce type de récepteurs) ou encore des stimulations électriques peuvent également être employées mais il est impossible de conclure quant à la nature des neurones impliqués dans les processus comportementaux observés à la suite des manipulations d'activité.

D'autres techniques encore comme la cryoinactivation peuvent être employées. Cette technique consiste en un refroidissement du cerveau grâce à l'utilisation de canules. Néanmoins la complexité de cette technique et les dommages cérébraux qu'elle induit dans les tissus cérébraux en font une technique peu courante (Gallo, 2007).

## 2) Techniques hautement spécifiques : l'avènement de la génétique

Grâce aux progrès des outils génétiques, il est aujourd'hui beaucoup plus facile de mettre en œuvre des techniques de neuromodulation plus spécifiques à une population de neurones donnée dans une aire cérébrale donnée.

Les animaux génétiquement modifiés permettent par exemple d'altérer spécifiquement le fonctionnement de certains neurones ou circuits cérébraux. L'observation de déficience comportementale chez ces animaux permet donc de mettre en évidence de manière causative l'implication de ces circuits dans le comportement. Par exemple, les souris SAPAP3-KO (Welch et al., 2007) (souris dont le gène SAPAP3 a été inactivé) et pour lesquelles le fonctionnement des synapses cortico-striatales est spécifiquement altéré présentent un large spectre de comportements répétitifs ce qui met en lumière l'implication de ce circuit dans la régulation de ce type de comportement (Burguière et al., 2013; Lamothe et al., 2020). Dans le cadre d'une tâche de renversement d'apprentissage, ces souris présentent également des déficits comportementaux lors des renversements et notamment une plus grande labilité de réponses (augmentation de la proportion de réponses de type « *win-shift* ») (Benzina et al., 2021) ce qui pourrait suggérer que ce circuit est impliqué dans la gestion de l'incertitude liée aux renversements successifs. Néanmoins, même si ce type de technique permet d'être plus

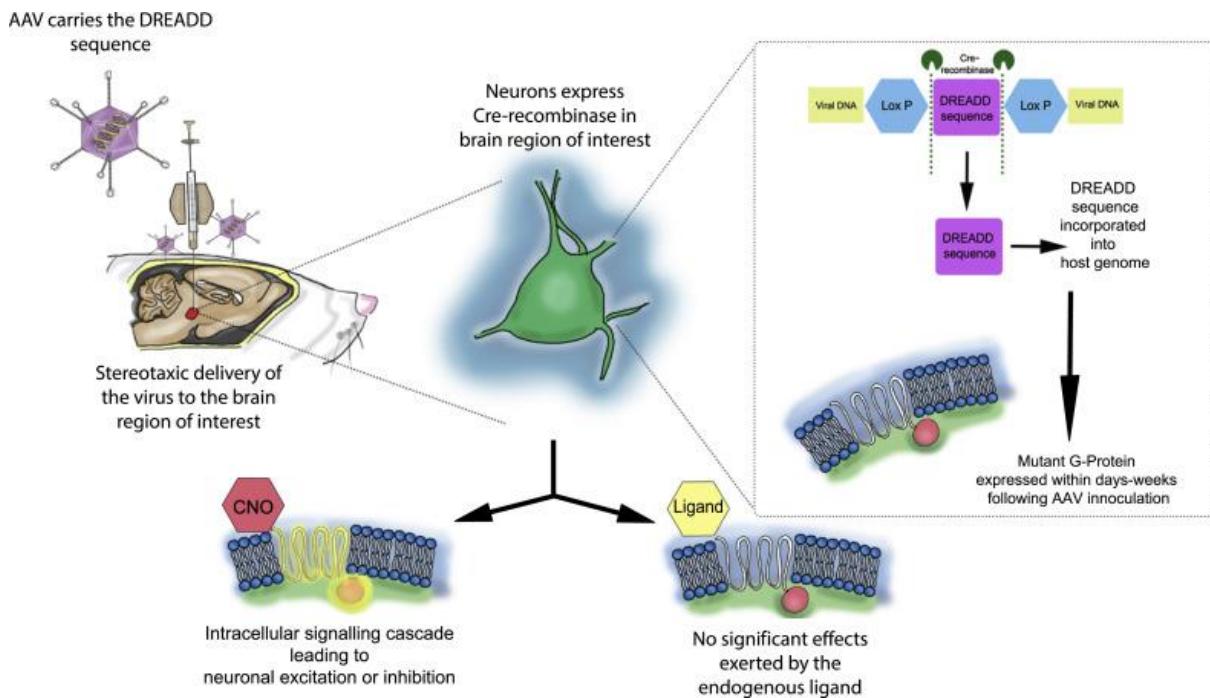
sélectif quant à l'aire cérébrale ciblée, le circuit ou la catégorie de neurones visés, elle ne permet pas en revanche un contrôle temporel de la manipulation.

Toutes les techniques au suffixe « génétique » en revanche permettent de telles manipulations. Optogénétique (Boyden et al., 2005; Deisseroth, 2011; Yizhar et al., 2011), pharmacogénétique (Roth, 2016; Sharma and Pienaar, 2018; Whissell et al., 2016), magnétogénétique (Nimpf and Keays, 2017) et thermogénétique (Bernstein et al., 2012) sont des techniques fondées sur la modification génétique des neurones (ou autre type cellulaire) afin que ces cellules expriment des protéines transmembranaires sensibles respectivement à la lumière, à des composés pharmacologiques de synthèse, à des champs magnétiques ou à la température. Les neurones sont alors activés ou inactivés le temps de la présence du stimulus (lumière, composé pharmacologique...). Ces protéines transmembranaires consistent en des protéines canaux, qui peuvent laisser passer des ions à la suite de leur ouverture (induisant une dépolarisation ou une hyperpolarisation directe des neurones), ou des protéines transmembranaires couplées à une petite protéine G (dont la cascade de signalisation alors enclenchée aboutit à la dépolarisation ou l'hyperpolarisation du neurone).

Le principe de ces techniques est le même et repose sur un mécanisme de transfection (**Figure 3**). Grâce à un virus servant de vecteur de matériel génétique, il est possible de faire exprimer la protéine transmembranaire exogène à une population de neurones donnée. La localisation de l'injection de ce virus grâce à une chirurgie stéréotaxique permet de cibler une aire cérébrale en particulier. En fonction des questions posées, il est également possible de cibler une population spécifique de neurones au sein desquels uniquement la protéine transmembranaire peut être exprimée. L'utilisation d'un promoteur spécifique (comme le promoteur CAMKII qui permet l'expression du récepteur uniquement dans les neurones glutamatergiques pyramidaux) ou encore des technologies Cre-Lox ou FlEx (Atasoy et al., 2008) permet au gène de n'être exprimé que dans une population neuronale d'intérêt.

L'activation des protéines transmembranaires par l'élément spécifique auquel elles sont sensibles (lumière d'une longue d'onde spécifique, composé pharmacologique...) a pour conséquence l'activation ou l'inhibition des neurones au sein desquels elles sont exprimées et donc une perte ou un gain de fonction de la zone ciblée.

Ces techniques réalisées de manière simultanée à l'étude du comportement permettent d'évaluer les conséquences de l'activation ou inactivation de l'activité neuronale sur le comportement. Ces techniques permettent donc d'investiguer la causalité de l'activité cérébrale sur celui-ci.



**Figure 3 : Exemple de la technique de pharmacogénétique.** La transfection virale dans une aire cérébrale restreinte de souris Cre permet l'expression de récepteurs couplés à une protéine G au sein d'une population de neurones spécifiques. Ces récepteurs sont sensibles uniquement à un composé pharmacologique synthétique et à aucune autre molécule endogène : ces récepteurs sont appelés *Designer Receptors Exclusively Activated by Designer Drugs* (DREADD). L'injection sous-cutanée de cette substance ou sa consommation orale par l'animal conduit à l'activation ou l'inhibition des neurones exprimant le récepteur durant la durée de vie du composé (Sharma and Pienaar, 2018).

### 1.2.3. ENJEUX SCIENTIFIQUES : CHOISIR LA TECHNIQUE DE NEUROMODULATION LA PLUS ADAPTEE ET EN CONNAITRE SES LIMITES

Bien que très puissantes, puisqu'elles permettent d'établir un lien de causalité entre l'activation ou l'inactivation de neurones ou de circuits neuronaux et le comportement, il reste indispensable de garder en tête que ces techniques présentent des limites.

Le choix du virus vecteur de la séquence génétique ou de la concentration même de la suspension virale peut avoir des conséquences sur les résultats. Certains virus ont tendance à diffuser de manière très importante et la spécificité de cible peut donc être perdue (Watakabe et al., 2015). L'analyse *a posteriori* de la diffusion de la transfection doit donc être vérifiée par immunohistochimie afin de s'assurer que les effets comportementaux observés sont effectivement bien liés à une activation ou inactivation de la population neuronale d'intérêt.

D'un point de vue purement technique, les réglages des paramètres de neuromodulation peuvent influencer les résultats. En optogénétique, le choix de la fréquence (Deng et al., 2016), de la durée et de la puissance de la stimulation peut faire toute la

différence sur les résultats comportementaux. En pharmacogénétique, le choix de la concentration ou encore du mode d'administration du composé pharmacologique (oral, sous-cutanée, via des canules directement implantées dans le crâne, via des mini-pompes sous-cutanée) peut également avoir des effets sur le comportement en provoquant du stress (Balcombe et al., 2004).

De plus, aucune technique n'est 100 % maîtrisable. L'échauffement des tissus provoqués par les stimulations lumineuses (Yizhar et al., 2011) ou encore les inconnues liées au composé pharmacologique utilisé peuvent avoir des conséquences inattendues. Par exemple, une des molécules couramment utilisées en pharmacogénétique est la clozapine N-oxide (CNO). Il a été découvert quelques temps après le début de son utilisation en neuromodulation expérimentale, qu'il s'agit en réalité un dérivé métabolique de cette molécule, la clozapine, qui agit sur les récepteurs artificiellement exprimés dans les neurones (Gomez et al., 2017). Or il a été montré que la clozapine est susceptible d'interagir avec des récepteurs endogènes, ce qui peut conduire à des effets comportementaux non spécifiques (Bærentzen et al., 2019).

L'ensemble des limites précédemment discutées peut être pallié par la réalisation de contrôles rigoureux : test des différents paramètres (différentes fréquences de stimulation par exemple dans le cadre de l'optogénétique), revue extensive de la littérature pour identifier des biais potentiels, utilisation d'animaux contrôles n'exprimant pas les récepteurs mais soumis aux stimulations lumineuses, magnétiques, thermiques ou administrées avec le composé pharmacologique...

Il existe également certaines inconnues liées à des considérations plus physiologiques. Réaliser les protocoles de neuromodulation en matinée ne produira peut-être pas les mêmes effets que dans la soirée : le rythme circadien a probablement des effets non négligeables sur les résultats (Sidor and McClung, 2014). Par ailleurs, même si de nombreuses données existent aujourd'hui, la connaissance exacte de l'ensemble circuits neuronaux n'est pas absolue. Notamment, les phénomènes de plasticité et de compensation sont difficilement maîtrisables et ceci peut interférer avec le comportement, biaisant donc les interprétations des données comportementales (Otchy et al., 2015). Le choix de la technique est donc à finement considérer en fonction des questions que se pose l'expérimentateur.

Cette partie a permis de présenter la diversité des outils méthodologiques utilisables par les chercheurs pour comprendre la relation entre activité cérébrale et comportements. La deuxième partie permettra d'illustrer comment ces techniques ont permis de mettre en évidence les bases neurobiologiques impliquées dans les *comportements adaptatifs*.

## **2. LES REGIONS CEREBRALES ET CIRCUITS IMPLIQUES DANS LES COMPORTEMENTS ADAPTATIFS**

La force mais aussi la complexité de l'étude des comportements adaptatifs sont essentiellement liés à l'investigation expérimentale chez différentes espèces. Néanmoins, toutes les espèces ne possèdent pas un cerveau identique. Les proportions de certaines régions par rapport à d'autres ne sont pas les mêmes d'une espèce à l'autre. La structure histologique et la connectivité intra-régions ne sont également pas strictement similaires et reflètent des siècles de divergence évolutive (Northcutt, 2002).

Toute la question est néanmoins d'identifier les bases neurobiologiques des comportements et fonctions cognitives. Or il est plus simple de répondre à cette question chez les animaux modèles puisqu'il est moins contraignant et plus facile d'utiliser diverses techniques que chez l'humain. Pourtant, s'il n'existe pas de structure similaire entre les différentes espèces, quel sens y a-t-il à étudier des mécanismes comportementaux et leurs bases neurobiologiques chez la souris et étendre les conclusions à l'humain ? (Carlén, 2017)

Cette partie permet donc de discuter de ces questions en s'intéressant plus particulièrement à la région cérébrale majoritairement impliquée dans les comportements adaptatifs, à savoir le cortex préfrontal. Le cortex préfrontal est en effet une structure cérébrale à la base des capacités cognitives supérieures (Fuster, 2013) telles que les processus d'apprentissage, de planification, de raisonnement, ces fonctions étant essentielles pour s'adapter correctement à un environnement *volatile*. Une sous-région du cortex préfrontal en particulier, le cortex cingulaire antérieur dorsal (dACC), semble être particulièrement impliquée.

Cette partie permettra de mettre en lumière l'implication du dACC dans l'émergence de comportement adaptatifs et de discuter de son existence au sein de différentes espèces.

### **2.1. LE CORTEX CINGULAIRE ANTERIEUR DORSAL FACE A LA VOLATILITE ENVIRONNEMENTALE**

#### **2.1.1. LE CORTEX CINGULAIRE ANTERIEUR DORSAL : UNE REGION CEREBRALE AUX MULTIPLES FONCTIONS**

De multiples études dont les résultats ont été commentés par plusieurs revues ont mis en lumière l'implication du dACC dans de multiples fonctions (Ebitz and Hayden, 2016;

Heilbronner and Hayden, 2016; Kolling et al., 2016a, 2016b). Ces revues permettent d'envisager le dACC comme un véritable centre intégrateur qui s'avère être au cœur de processus indispensables à diverses fonctions cognitives et à la mise en place de multiples comportements. Le dACC est donc placé au cœur de différents modèles computationnels qui soulignent son implication essentielle à l'*adaptation comportementale* (Vassena et al., 2017) et il a été montré que des lésions de cette structure empêchent les animaux de s'adapter correctement à des changements de contingences au cours d'une tâche comportementale (Rushworth et al., 2003).

Il est proposé notamment que le dACC permet de : (1) mettre à jour ses estimations sur l'environnement, (2) ajuster les processus d'apprentissage, (3) modifier les stratégies comportementales en cours (Kolling et al., 2016a). Dans cette perspective, le rôle principal du dACC est donc de mettre à jour dynamiquement les politiques comportementales face à un environnement *volatil* (Ebitz and Hayden, 2016; Kolling et al., 2016b). Les trois aspects précédemment cités seront développés dans les paragraphes suivants.

## **2.1.2. LE CORTEX CINGULAIRE ANTERIEUR DORSAL : REGION CLE DE L'ADAPTATION COMPORTEMENTALE**

### **1) Le dACC est impliqué dans la mise à jour des estimations sur l'environnement**

Lorsque l'environnement change et que les contingences sont modifiées, la première étape permettant à tout être vivant de s'adapter à ces changements consiste en la détection de ceux-ci. Cette détection permet de mettre à jour les croyances du sujet sur son environnement et lui indiquer que celui-ci a probablement changé. Le dACC semble impliqué dans ces processus.

Behrens et collaborateurs ont montré chez l'humain que l'activité du dACC reflète le niveau de *volatilité* estimé de l'environnement : par imagerie par résonance magnétique fonctionnelle, les auteurs ont montré que le dACC présente une activité plus élevée lorsque le sujet évalue la *volatilité* de son environnement qui varie au cours de la tâche de renversement qu'il doit réaliser (Behrens et al., 2007). Ainsi, si l'ACC peut encoder la *volatilité*, il est possible pour le sujet d'établir un modèle interne de son environnement lui permettant de s'adapter au mieux à celui-ci.

Au cours d'une tâche comportementale durant laquelle des singes peuvent obtenir des récompenses plus ou moins grandes après avoir choisi une option parmi deux, Hayden et

collaborateurs ont montré que l'activité du dACC augmente avec la surprise (Hayden et al., 2011a). Ainsi l'activité du dACC reflète à quel point les résultats dévient des attentes initiales. Or à la suite de changements de contingences non indicés, l'agent commet des erreurs auxquelles il ne s'attend pas : l'*incertitude inattendue* est alors probablement encodée par l'augmentation d'activité du dACC signalant ainsi la nécessité d'adaptation.

Des résultats plus détaillés ont été mis en lumière par O'Reilly et collaborateurs. En considérant que le cerveau se fonde sur l'utilisation de modèles prédictifs pour faciliter le traitement des informations et la planification d'actions, les auteurs ont cherché à mettre en évidence comment ces modèles peuvent être mis à jour afin de prendre en compte les modifications de l'environnement en dissociant ces mises à jour de la surprise causée par des événements inattendus. Grâce à leur tâche et à l'utilisation de l'imagerie par résonance magnétique fonctionnelle les auteurs ont pu montrer que le cortex cingulaire antérieur était spécifiquement activé pendant la mise à jour du modèle prédictif (O'Reilly et al., 2013).

Ces résultats semblent cohérents avec les travaux de Tervo sur le rat. Mis en concurrence avec des compétiteurs de différents niveaux de difficulté, les rats sont capables de s'affranchir de l'histoire des récompenses obtenues et de présenter un comportement "stochastique" lorsqu'ils ne peuvent pas vaincre leur compétiteur par contre-prédiction. Les auteurs montrent que dans ce mode stochastique l'engagement du dACC est supprimé (Tervo et al., 2014). L'activité du dACC est donc essentiel à la mise à jour des modèles internes sur son environnement.

Grâce au suivi, à la détection et la prise en considération des changements, il est alors possible de mettre en place des réponses comportementales adaptées.

## 2) Le dACC est impliqué dans les processus d'apprentissage et de modulation de l'apprentissage

Le dACC a été impliqué dans l'encodage des erreurs (Holroyd and Coles, 2002) et des récompenses obtenues par un sujet dans différentes tâches comportementales. Il a notamment été montré que la fréquence de décharge des neurones du dACC encode les récompenses obtenues (Amiez et al., 2006) voire celles qui auraient pu être obtenues (Blanchard et al., 2001; Hayden et al., 2009). Certaines études ont également caractérisé l'activité du dACC comme encodant l'erreur de prédiction de récompense liée à la différence entre la récompense attendue et la récompense obtenue (Kennerley et al., 2011; Matsumoto et al., 2003; Seo and Lee, 2007).

Une étude menée chez des singes dont le dACC a été endommagé montre que ceux-ci présentent un déficit dans le maintien de leur stratégie comportementale en cours (Kennerley et al., 2006). Dans le cadre de cette tâche, les singes doivent choisir entre tourner ou lever une manette : pendant un certain nombre d'essais une action est correcte, puis un changement intervient et l'autre action doit alors être réalisée. Les singes pour lesquels le dACC est endommagé ont du mal à corriger leur réponse à la suite d'erreurs. Les singes semblent donc être incapables de prendre en compte l'historique des résultats pour adapter de manière optimale leur réponse et choisir l'association correctement associée à la règle en cours. Les auteurs concluent donc que le dACC est indispensable pour apprendre les valeurs des actions à réaliser et assure ainsi l'intégration des informations de renforcement au fil du temps pour guider le comportement de choix.

L'ensemble de ces résultats est donc en faveur de l'implication du dACC dans le processus d'apprentissage tel qu'un modèle d'*apprentissage par renforcement* pourrait le prédire. Il est également intéressant de noter que le dACC semble également impliqué dans la modulation des paramètres d'apprentissage.

L'étude de Behrens montre notamment que le changement d'activité dans le dACC reflète la volatilité estimée de l'environnement mais également que les variations d'activité du dACC entre sujets permettent d'expliquer les différences de taux d'apprentissage de chacun des sujets au cours de la tâche : la volatilité mesurée par le dACC permet donc d'ajuster dynamiquement le taux d'apprentissage (Behrens et al., 2007). Ces conclusions sont également supportées par les travaux de McGuire et collaborateurs qui montrent que les paramètres de leur modèle qui influencent le taux d'apprentissage (à savoir l'incertitude attendue, l'incertitude inattendue et la récompense) sont associés une augmentation de l'activité du dACC (McGuire et al., 2014).

Le modèle de Khamassi et collaborateurs propose également un rôle de superviseur du dACC dans l'encodage de l'incertitude des résultats et de l'historique des résultats permettant ainsi d'adapter les paramètres d'apprentissage au gré des variations de l'environnement (Khamassi et al., 2013).

### 3) Le dACC est impliqué dans les changements de stratégies comportementales

Après avoir détecté que l'environnement ou les règles ont changé, il est nécessaire de s'y adapter et donc de mettre en place une nouvelle stratégie comportementale. Le dACC semble également impliqué dans ces processus.

Kolling et collaborateurs ont essayé de comprendre comment des sujets humains alternent entre des stratégies d'*exploration* et d'*exploitation*. Des participants devaient réaliser une tâche durant laquelle ils pouvaient décider de choisir dans un ensemble d'options par défaut ou d'explorer d'autres ensembles d'options. La décision d'explorer dépendait de deux valeurs : (1) d'une valeur de recherche (valeur moyenne des alternatives), (2) d'une valeur d'engagement (valeur des options par défaut). En enregistrant l'activité cérébrale par imagerie par résonance magnétique fonctionnelle, les auteurs ont montré que le dACC permet d'encoder le bénéfice associé à l'exploration d'autres options et que la valeur de recherche était liée au dACC (Kolling et al., 2012). De manière similaire, chez le macaque, il a été montré que le dACC permet d'encoder une variable de décision signalant la valeur relative de quitter une ressource épuisée pour en exploiter une nouvelle (Hayden et al., 2011b).

Dans une autre étude, des singes ont réalisé une tâche déterministe durant laquelle ils devaient identifier une cible récompensante parmi 4. Une fois la cible identifiée, ils devaient continuer de répondre dessus jusqu'à l'observation d'un signal qui indiquait que la cible va changer. À la suite de cette période d'*exploitation*, les singes devaient de nouveau explorer pour trouver la nouvelle cible. La particularité de cette tâche repose sur son aspect déterministe : le premier résultat négatif observé à la fin de la phase d'*exploitation* indique à l'animal qu'il doit explorer, et le premier résultat récompensé lors de la phase d'*exploration* est un indice pour l'animal de passer en mode *exploitatoire*. L'enregistrement simultané des neurones du dACC grâce à des microélectrodes a permis de montrer que le dACC est sensible à différents résultats observés : l'activité du dACC augmentait lors de la première réponse incorrecte à la fin d'une phase d'*exploitation*/début de la nouvelle phase d'*exploration* et à la première réponse correcte du début de la nouvelle phase d'*exploitation*/fin de la phase d'*exploration*. Cette augmentation d'activité s'accompagnait du changement de comportement de l'animal d'un mode *exploratoire* vers un mode *exploitatoire* (et vice-versa). Le dACC permet donc probablement d'encoder un signal de transition comportementale (Quilodran et al., 2008).

Ces données sont cohérentes avec les travaux de Karlsson et collaborateurs (Karlsson et al., 2012). Ces derniers ont montré chez le rat que la détection par l'animal du changement de contingences de la tâche s'accompagne d'une transition brutale de l'activité des neurones du dACC et par l'émergence d'un comportement d'*exploration* de l'animal des différentes options possibles dans la tâche. Ces transitions brutales de l'activité des neurones du dACC ont été également retrouvées précédant des changements de stratégies comportementales à la

suite de changement de règles : les points de rupture comportementaux observés corrélaient parfaitement avec les points de rupture de l'activité neuronale (Durstewitz et al., 2010).

Domenech et collaborateurs ont quant à eux montré par électroencéphalographie intracrânienne que le cortex dorsomédian préfrontal encode la transition de comportements d'exploitation à des comportements d'exploration. (Domenech et al., 2020)

L'ensemble de ces études supporte ainsi l'idée d'un encodage réactif au sein du dACC permettant d'ajuster les comportements en cours en fonction de la perception des états cachés de l'environnement (changement de contingences par exemple). Néanmoins, il est à noter que chez la souris, Vertechi et collaborateurs ont montré que l'inactivation du dACC contrairement à l'inactivation de l'OFC, n'induit pas de changements qualitatifs dans la stratégie des animaux, mais seulement une tendance générale à rester plus longtemps sur l'ancien site de distribution de récompenses plutôt que de se rendre vers le nouveau. Des précautions quant à l'homologie des structures ciblées chez cet animal sont donc à considérer d'autant que les auteurs ont ciblé une aire assez antérieure du cingulaire (Vertechi et al., 2020).

### **2.1.3. ENJEUX SCIENTIFIQUES : CIBLER LE CORTEX CINGULAIRE ANTERIEUR DORSAL COMME PREMIERE APPROCHE POUR L'ETUDE DES BASES NEUROBIOLOGIQUES DES COMPORTEMENTS ADAPTATIFS**

Au vu de la revue réalisée précédemment, il est assez clair que le dACC est une structure critique dans la mise en place des *comportements adaptatifs*, notamment à la suite de changements abrupts de contingences puisque cette région est cruciale aux processus d'adaptation du fait de son rôle dans les processus d'apprentissage, de transition entre *exploration* et *exploitation*, ou encore dans la construction de modèles internes sur l'environnement (Ebitz and Hayden, 2016; Heilbronner and Hayden, 2016; Kolling et al., 2016a, 2016b).

Dans le cadre de la nouvelle tâche développée au cours de ma thèse, cibler de manière spécifique cette structure a donc constitué une première approche pertinente afin de comprendre les bases neurobiologiques des *comportements adaptatifs*. Ayant travaillé chez la souris, alors que la plupart des études précédemment citées ont été menées chez l'humain, le primate voire le rat, il est donc nécessaire de faire un point sur l'existence de structures homologues du dACC chez mon modèle d'étude.

## 2.2. VALIDITE TRANSLATIONNELLE DES SUBSTRATS NEUROBIOLOGIQUES

### 2.2.1. ETABLIR DES HOMOLOGIES DE STRUCTURES CEREBRALES

En biologie, des structures chez des espèces différentes peuvent être qualifiées d'homologues si elles partagent des ressemblances car héritées d'un ancêtre commun (Carlén, 2017).

Pour conclure à l'homologie de deux structures, plusieurs critères sont à observer (Uylings et al., 2003):

- Similarité du développement embryonnaire des structures
- Propriétés fonctionnelles semblables tant d'un point de vue électrophysiologique que comportemental et/ou cognitif
- Connectivité aux autres structures cérébrales proche tant en termes qualitatif que quantitatif
- Expression de récepteurs similaires et présence des mêmes neuromédiateurs
- Cytoarchitecture identique pour des espèces proches

Plus les concordances sont nombreuses, plus il est probable que les structures considérées sont effectivement bien homologues.

### 2.2.2. LE CORTEX CINGULAIRE ANTERIEUR DORSAL : SOUS-STRUCTURE DU CORTEX PREFRONTAL

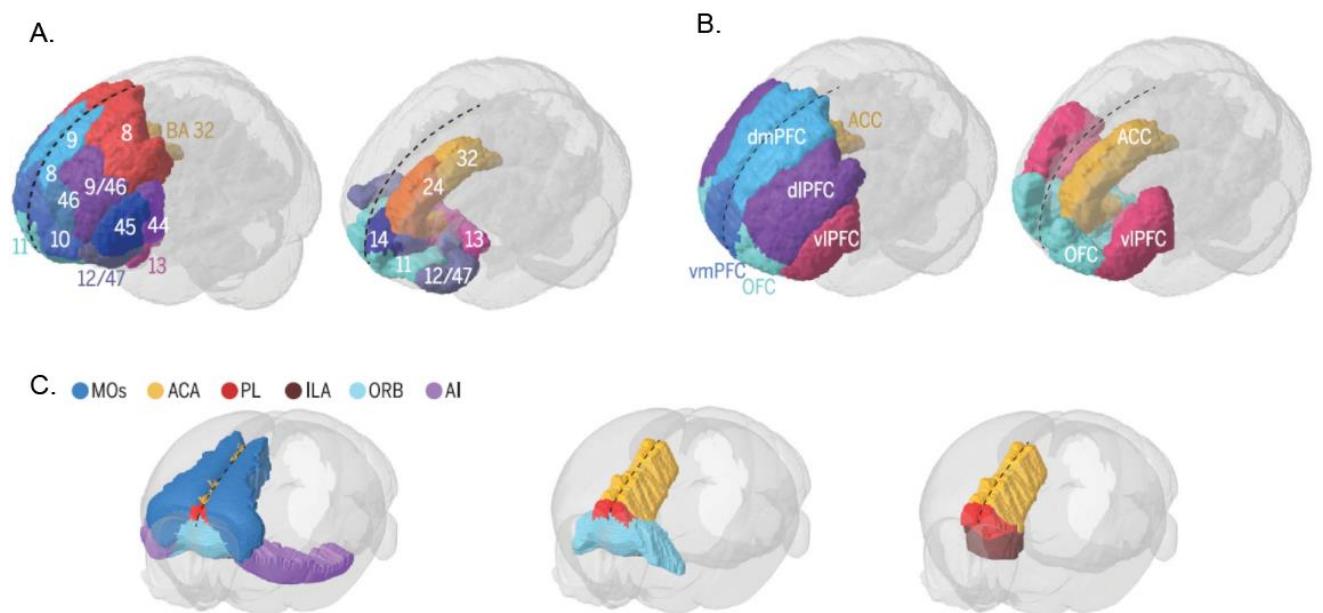
#### 1) Un bref aperçu historique de l'étude du cortex préfrontal

Historiquement, le cortex préfrontal chez l'humain a été décrit de manière topographique par Brodman en étudiant la cytoarchitecture de cette région (Brodmann, 1909). Brodman en a alors établi des aires auxquelles il a attribué un numéro : les aires 8 à 14, et 44 à 47 constituent le cortex préfrontal (**Figure 4A**). Celui-ci semblait notamment se dissocier des autres aires cérébrales par la présence d'une couche granulaire IV. Cette granularité se retrouve chez les primates non humains mais pas chez les rongeurs, ce qui laissait conclure à une absence de cortex préfrontal chez les espèces murines.

Or, par la suite les études de connectivité ont montré que le cortex des rats et des souris reçoit des projections du noyau méiodorsal du thalamus, projections qui étaient alors supposées comme étant une spécificité du cortex préfrontal identifié chez l'humain (Preuss,

1995). Ces résultats ont donc renversé les conclusions précédentes et permis d'envisager l'existence d'un cortex préfrontal chez les rongeurs malgré son agranularité. De plus, chez l'humain, des aires agranulaires (24, 25, et 32, **Figure 4A**) regroupées sous la dénomination de cortex cingulaire, sont venues étendre le cortex préfrontal puisqu'elles recevaient également des projections du noyau médiiodorsal du thalamus (Carlén, 2017).

Différentes études d'histologie (Vogt and Paxinos, 2014), de connectivité (Heilbronner et al., 2016; Sallet et al., 2013; Uylings and van Eden, 1990), d'électrophysiologie, d'évaluation computationnelle et comportementale (Seamans et al., 2008) ont donné de nombreux autres arguments en faveur de l'existence d'un cortex préfrontal chez les rongeurs, homologue à celui des humains et des primates, même si tout de même réduit en volume relatif et en nombre de sous-structures (Uylings et al., 2003; Vogt and Paxinos, 2014).



**Figure 4 : Le cortex préfrontal chez l'homme et la souris.** A. Le cortex préfrontal chez l'homme a été défini initialement par Brodmann grâce à l'étude de l'architecture cytologique des tissus. Brodmann a inclus les aires 8 à 14, et 44 à 47 dans le cortex préfrontal. Les aires 24, 25 et 32, bien qu'agranulaires ont été ajoutées par la suite du fait de leur profil de connectivité. B. Définition du cortex préfrontal humain sous forme d'ensembles ayant des profils fonctionnels et de connectivité semblables. OFC : orbitofrontal cortex (cortex orbitofrontal) | vmPFC : ventromedial prefrontal cortex (cortex préfrontal ventromédian) | vIPFC : ventrolateral prefrontal cortex (cortex préfrontal ventrolatéral) | dmPFC : dorsomedial prefrontal cortex (cortex préfrontal dorsomédian) | dlPFC : dorsolateral prefrontal cortex (cortex préfrontal dorsolatéral) | ACC : anterior cingulate cortex (cortex cingulaire antérieur). C. Délimitations du cortex préfrontal de la souris. MOs : secondary motor areas (aires motrices secondaires) | ORB : orbital areas (aire orbitale) | AI : agranular insular area (aire agranulaire insulaire) | ACA : anterior cingulate area (aire antérieure cingulaire) | PL : prelimbic area (aire prélimbique) | ILA : infralimbic area (aire infralimbique). Adapté de (Carlén, 2017).

## 2) Le cortex préfrontal ou les cortex préfrontaux ?

Le cortex préfrontal en lui-même n'est pas une région cérébrale homogène et différentes aires s'y distinguent du fait de leur spécificité de connexions mais aussi de leurs fonctions.

Chez l'Homme, on distingue classiquement le cortex orbitofrontal, le cortex préfrontal ventromédian, le cortex préfrontal ventrolatéral, le cortex préfrontal dorsomédian (dmPFC pour *dorsomedial prefrontal cortex* en anglais), le cortex préfrontal dorsolatéral, le cortex cingulaire (Carlén, 2017) (**Figure 4B**).

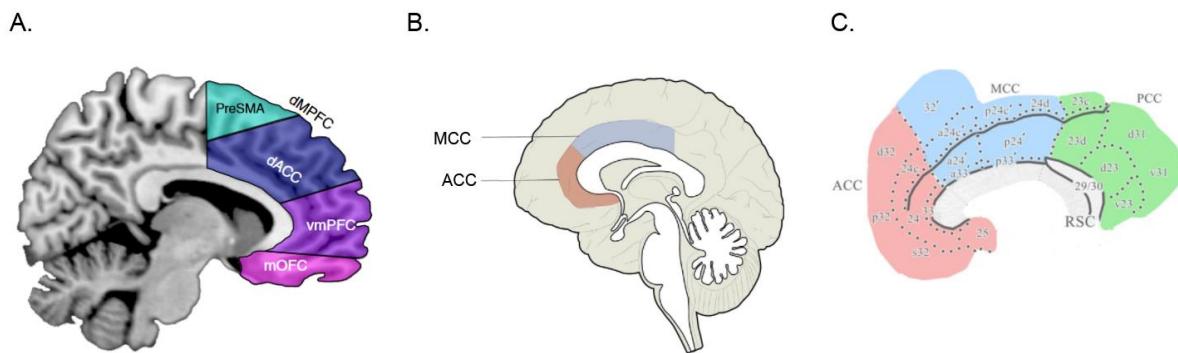
Il est à noter que le dmPFC ne semble pas faire consensus sur sa délimitation. Alors que certains auteurs considèrent qu'il s'agit d'une aire corticale superficielle sous laquelle est enfoui le cortex cingulaire (Carlén, 2017), d'autres auteurs regroupent sous le terme de dmPFC, une partie du cortex cingulaire, les régions corticales plus superficielles à celle-ci ainsi que l'aire prémotrice supplémentaire (Domenech and Koechlin, 2015) (**Figure 5A**).

Chez les rongeurs, le cortex préfrontal regroupe quant à lui les aires orbitofrontales, l'aire infralimbique, l'aire prélimbique, l'aire cingulaire, l'aire insulaire ainsi que les aires prémotrices (Carlén, 2017) (**Figure 4C**). La correspondance par homologie de chacune de ces régions avec les aires précédemment discutées chez l'Homme est au cœur de tous les débats parmi la communauté scientifique. La partie médiane de ce cortex préfrontal (aires infralimbique, prélimbique et cingulaire) est notamment considérée comme proche du cortex médian préfrontal des primates.

## 3) Le dACC au cœur du cortex préfrontal

Le dACC fait partie du cortex préfrontal, mais sa dénomination varie d'une étude à l'autre. Suivant des arguments d'histologie, d'imagerie fonctionnelle et de densité de récepteurs à certains neuromédiateurs, le cortex cingulaire est en effet divisé en deux régions (Vogt and Paxinos, 2014) : une région antérieure et une région médiane. Là où certains auteurs les appelleront respectivement cortex cingulaire antérieur (ACC pour *anterior cingulate cortex* en anglais) et cortex cingulaire médian (MCC pour *midcingulate cortex* en anglais), d'autres auteurs préféreront cortex cingulaire ventral (vACC pour *ventral anterior cingulate cortex*) et cortex cingulaire dorsal (dACC pour *dorsal anterior cingulate cortex*) (Heukelum et al., 2020) (**Figure 5B**). Ces deux régions, MCC/dACC, et ACC/vACC regroupent différentes aires de Brodman (Vogt and Paxinos, 2014). (**Figure 5C**). A noter que

pour certains auteurs, le dACC comprend le MCC mais aussi une petite portion de la partie la plus dorsale de l'ACC (Heilbronner and Hayden, 2016).



**Figure 5 : Le cortex cingulaire antérieur humain : limites et dénominations.** **A.** Chez certains auteurs le cortex cingulaire antérieur dorsal (dACC) fait partie intégrante du cortex préfrontal dorsomédian (dmPFC) au même titre que l'aire supplémentaire prémotrice (preSMA). Adapté de (Domenech and Koechlin, 2015) **B.** Le cortex cingulaire, situé au-dessus du corpus callosum est divisé en deux régions : le cortex cingulaire antérieur (ACC, dénommé aussi cortex cingulaire antérieur ventral, vACC) et le cortex cingulaire médian (MCC, dénommé aussi cortex cingulaire antérieur dorsal, dACC). Adapté de (Heukelum et al., 2020). **C.** Les différentes parties du cortex cingulaire humain et les aires de Brodmann qu'elles englobent. PCC correspond au cortex cingulaire postérieur mais n'est pas considéré comme appartenant au cortex préfrontal. Adapté de (Vogt and Paxinos, 2014)

#### 4) Débats sur les homologies du dACC

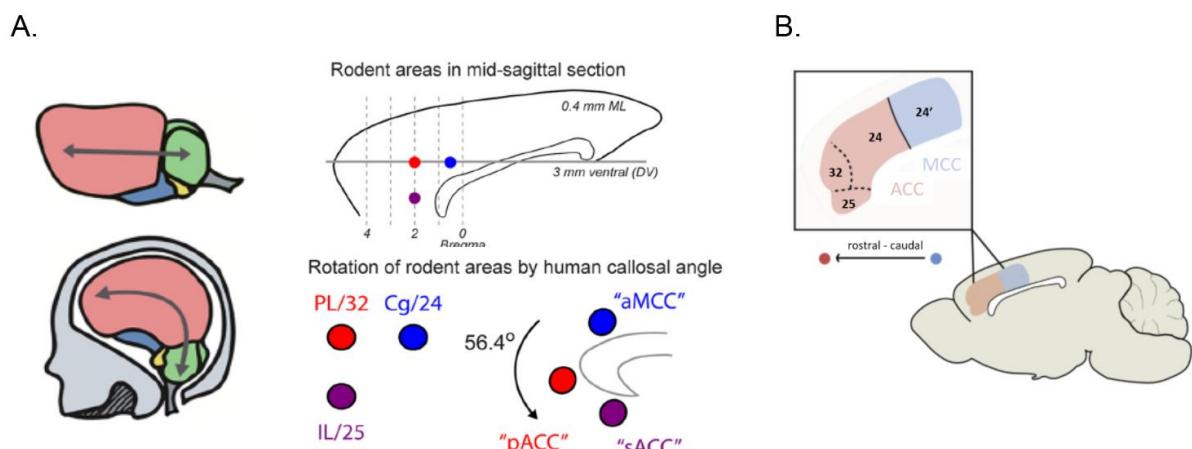
L'extensivité de la littérature sur le sujet du cortex préfrontal ainsi que les hétérogénéités de conclusions quant aux homologies à travers les espèces illustrent à quel point le débat fait encore rage parmi les experts du cortex préfrontal (Carlén, 2017). Certaines revues pointent notamment de manière très détaillée les enjeux et points de discorde (Laubach et al., 2018). Diversité sémantique, mauvais usage terminologique, utilisation d'atlas stéréotaxique différents ou trop anciens sont les causes de divergences dans les conclusions de nombreuses études. Alors que l'homologie du dACC entre primates humains et non-humains est assez bien établie (Kolling et al., 2016b), l'homologie entre primates et rongeurs est plus délicate à évaluer.

### 2.2.3. ENJEUX SCIENTIFIQUES : CHOISIR AU MIEUX LES COORDONNEES STEREOGRAPHIQUES DU CORTEX CINGULAIRE ANTERIEUR DORSAL CHEZ LA SOURIS

Laubach et collaborateurs ainsi qu'Heukelum et collaborateurs proposent de considérer de manière homologue le cortex cingulaire à travers les espèces en dissociant le

cortex cingulaire antérieur (ACC/vACC) et le cortex cingulaire médian (MCC/dACC). En se fondant sur l'étude de la courbure du cortex à travers les différentes espèces, ainsi qu'en établissant des cartes de connectivité et de fonctions, les auteurs proposent alors une nouvelle carte d'homologie entre rongeurs et humains.

Les auteurs proposent que les aires des rongeurs appelées le plus communément infralimbique et prélimbique appartiennent toutes les deux à l'ACC/vACC (**Figure 6**). Entre les deux revues en revanche, la délimitation antéro-postérieure du cingulaire diffère. Là où Laubach et collaborateurs considère l'aire 24 (région antérieure du cingulaire) comme appartenant au MCC/dACC (**Figure 6A**), Heukelum et collaborateurs placent une délimitation entre les aires 24 (région antérieure du cingulaire) et 24' (région postérieure du cingulaire), la première faisant partie de l'ACC/vACC, la seconde faisant partie du MCC/dACC (**Figure 6B**). Cette différence est peut-être liée au fait que les premiers auteurs considèrent l'agrégation de partie la plus postérieure de l'ACC/vACC et du MCC comme un ensemble homogène à l'instar des auteurs qui regroupent ces régions sous la dénomination de dACC (Heilbronner and Hayden, 2016).



**Figure 6 : Homologies des cortex cingulaires de la souris à l'homme.** **A.** La comparaison des degrés de courbure entre les cerveaux d'humains et de rongeurs permet de proposer une hypothèse d'homologie sur les aires infralimbique, prélimbique et cingulaire en comparaison avec les cortex cingulaires antérieur et médian humain. Adapté de (Laubach et al., 2018). **B.** La cartographie de la connectivité et des fonctions des aires infralimbique, prélimbique et cingulaire chez le rongeur en comparaison avec l'humain laisse penser que les aires infralimbique (25) et prélimbique (32) appartiennent bien au cortex cingulaire antérieur ventral (ACC/vACC). En revanche, les auteurs considèrent que les régions 24 et 24' appartiennent respectivement à l'ACC/vACC et au MCC/dACC. Adapté de (Heukelum et al., 2020).

Ces divergences montrent néanmoins que l'étude des homologies du cortex préfrontal à travers les différentes espèces n'a pas encore atteint un consensus généralisé et qu'il reste des zones d'ombre notamment sur les homologies de fonction entre les différentes aires (Heukelum et al., 2020). Laubach et Heukelum insistent donc sur la nécessité d'investiguer

les fonctions des différentes aires en utilisant une méthodologie rigoureuse (cibler une aire en particulier et non plusieurs aires en même temps (Hamilton and Brigman, 2015) en maximisant la spécificité des technologies de neuromodulation par exemple) et en utilisant une terminologie précise ou tout du moins l'explicitation des coordonnées utilisées pour les enregistrements d'activité cérébrale et les techniques de neuromodulation, histologie à l'appui.

### **3. AU-DELA DES CIRCUITS : LA FLEXIBILITE DES CIRCUITS NEURONNAUX COMME SUPPORT DES COMPORTEMENTS ADAPTATIFS**

#### **3.1. LA NEUROMODULATION : PROCESSUS PHYSIOLOGIQUE DE PLASTICITE DES CIRCUITS NEURONNAUX ET DES COMPORTEMENTS**

##### **3.1.1. LA NEUROMODULATION COMME SUPPORT DES MODIFICATIONS DE L'INFORMATION NERVEUSE**

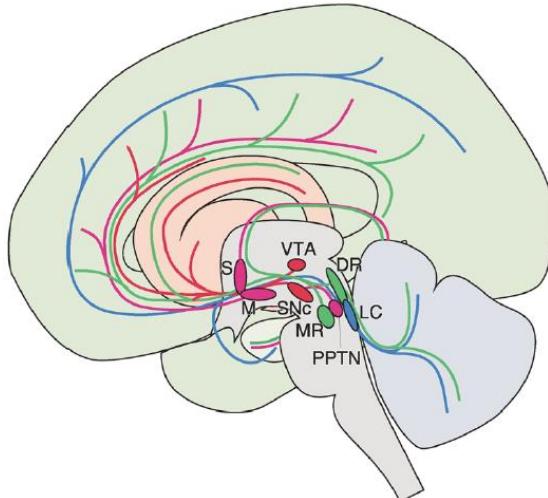
Dans la première partie, j'ai présenté la neuromodulation expérimentale qui consiste en une manipulation contrôlée par un expérimentateur de l'activité des neurones grâce à diverses techniques. La neuromodulation correspond également à un processus physiologique qui existe chez tous les êtres vivants, même ceux disposant d'un système nerveux simple (Brunelli et al., 1976). La neuromodulation est un processus permettant la modification et la modulation des informations transmises au sein du système nerveux (Kaczmarek and Levitan, 1987; Katz and Calin-Jageman, 2008; Marder and Thirumalai, 2002; Nadim and Bucher, 2014). La neuromodulation consiste donc véritablement en un support de la flexibilité et de la plasticité des circuits neuronaux. Par extension, ce processus est essentiel à la mise en place des comportements et à leur modification au cours du temps (Montague et al., 2004). La neuromodulation est donc essentielle à une adaptation comportementale rapide en cas de changement de l'environnement (Bouret and Sara, 2005).

Les systèmes neuromodulateurs étant multiples et complexes, il est essentiel de les présenter rapidement avant de comprendre comment ils supportent le comportement et ses variations.

### 3.1.2. LES GRANDS CIRCUITS NEUROMODULATEURS

Quatre grands groupes de structures neuromodulatrices sont recensés, chacun synthétisant et libérant une classe de neuromodulateur spécifique (Doya, 2002) (**Figure 7**):

- 1) Le locus coeruleus : ce noyau est responsable de la synthèse et libération de noradrénaline (appelée aussi norépinéphrine).
- 2) Les noyaux du raphé : ces noyaux sont responsables de la synthèse et libération de sérotonine.
- 3) L'aire tegmentale ventrale et la substance noire : ces noyaux sont responsables de la synthèse et libération de dopamine.
- 4) Les noyaux cholinergiques : ces noyaux sont responsables de la synthèse libération et d'acétylcholine.



NEUROMODULATEUR SYNTHETISE ET LIBERE	ORIGINE DE LA PROJECTION	CIBLES MAJEURES
Dopamine	Substance noire compacte (SNc) Aire tegmentale ventrale (VTA)	Striatum dorsal, ventral, cortex frontal
Sérotonine	Noyau dorsal du raphé (DR) Noyau médian du raphé (MR)	Cortex, striatum, cervelet, hippocampe
Noradrénaline	Locus coeruleus (LC)	Cortex, hippocampe
Acétylcholine	Noyau septal (S) Noyau de Meynert (M) Noyau pontomésencéphalotegmental (PPTN)	Cortex, amygdale, hippocampe, substance noire, thalamus, colliculus supérieur

**Figure 7 : Caractéristiques, localisation et projections des structures neuromodulatrices.** Adapté de (Doya, 2002)

Les neurones de ces structures sont des neurones de projection divergente, c'est-à-dire que leurs axones sont longs et projettent sur de multiples aires cérébrales, parfois très éloignées des noyaux dans lesquels leur soma est confiné. Le cortex préfrontal et donc le dACC sont notamment largement ciblés par les projections neuronales de ces structures neuromodulatrices.

### **3.1.3. ENJEUX SCIENTIFIQUES : FAIRE FACE A LA COMPLEXITE DE LA NEUROMODULATION**

Dans le cas de la neuromodulation, les récepteurs impliqués sont des récepteurs métabotropiques et plus particulièrement des récepteurs couplés aux protéines G. Pour un neuromodulateur donné, il existe une immense diversité de récepteurs couplés aux protéines G, et donc les conséquences intracellulaires suivant leur activation sont multiples. Les neurones sont d'ailleurs équipés de plusieurs récepteurs métabotropiques pour le même neuromodulateur ainsi que pour plusieurs neuromodulateurs : il existe de ce fait une combinatoire neuromodulatrice importante (Nadim and Bucher, 2014). Par ailleurs, tous ces récepteurs ne présentent pas les mêmes affinités aux molécules et donc en fonction de la concentration des neuromodulateurs, seuls certains récepteurs sont activés.

Par ailleurs, contrairement à la neurotransmission au cours de laquelle les neurotransmetteurs sont rapidement dégradés, réabsorbés ou diffusent rapidement hors de la synapse, garantissant ainsi la pertinence temporelle des informations nerveuses, les neuromodulateurs sont moins rapidement éliminés. De plus, les récepteurs couplés aux protéines G sont plus lentement désactivés : de ce fait, les effets de la neuromodulation peuvent être prolongés dans le temps (Katz and Calin-Jageman, 2008).

L'ensemble de ces remarques constitue donc un point critique méthodologique qu'il est nécessaire de prendre en considération dans toute expérimentation qui s'intéresse aux systèmes neuromodulateurs. La première étape consiste effectivement à caractériser l'effet du système neuromodulateur sur un comportement donné en le manipulant expérimentalement (par utilisation de l'optogénétique par exemple, (Lottem et al., 2018)). Comme je le détaillerai dans les paragraphes suivants, de nombreuses études ont utilisé ce type d'approche, permettant d'immenses progrès sur la compréhension des mécanismes d'apprentissage et d'adaptation. Néanmoins, les neurosciences font aujourd'hui face à des enjeux encore plus complexes puisqu'il s'agit désormais de disséquer de manière plus fine ces processus : déterminer les récepteurs impliqués (Naudé et al., 2016), évaluer les concentrations minimales

de neuromodulateurs nécessaires pour observer des conséquences comportementales ou encore comprendre l'influence des motifs de décharge des neurones neuromodulateurs (Ellwood et al., 2017).

### **3.2. NEUROMODULATION, MODELES COMPUTATIONNELS ET COMPORTEMENTS ADAPTATIFS**

La neuromodulation affecte la neurotransmission et donc la manière dont les informations sont intégrées et générées au sein du système nerveux. Le processus de neuromodulation a donc nécessairement des effets sur les fonctions cognitives et de manière consécutive sur le comportement. Comme souligné auparavant, la neuromodulation assure une véritable plasticité des circuits neuronaux : il semble donc très probable que la neuromodulation soit un support neurobiologique de la *flexibilité comportementale* et de des *comportements adaptatifs* (Amodeo et al., 2020; Ragozzino, 2002; Ragozzino et al., 2009; Seu et al., 2009).

En réponse au chapitre 2 de l'état de l'art, je détaillerai comment les systèmes de neuromodulation sont intégrés dans les différents modèles computationnels précédemment discutés pouvant alors expliquer cette capacité des êtres vivants à modifier leur comportement efficacement en réponse à des changements abrupts de l'environnement.

#### **3.2.1. META-APPRENTISSAGE ET NEUROMODULATION**

Dans sa revue princeps « Metalearning and neuromodulation », Doya effectue une revue extensive de la littérature afin d'illustrer comment les neuromodulateurs peuvent avoir des effets sur les fonctions cognitives et le comportement des agents et comment ces changements peuvent être reflétés par une modification des méta-paramètres du modèle d'*apprentissage par renforcement*. Doya propose alors quatre hypothèses (Doya, 2002) :

- 1) La dopamine est un des premiers neuromodulateurs à avoir été impliqué dans les processus d'*apprentissage par renforcement* puisqu'elle est essentielle dans le traitement de la récompense et l'encodage de l'*erreur de prédiction* (Schultz et al., 1997). La dopamine permet donc d'assurer la représentation globale du signal

d'apprentissage et la prédiction de la récompense : la dopamine permet d'assurer l'encodage l'erreur de prédiction  $\delta$ . A noter que  $\delta$  n'est pas un méta-paramètre.

- 2) La sérotonine contrôle la balance entre la prédiction à court-terme et à long-terme de la récompense : la sérotonine est donc impliquée dans la modification du décompte temporel, reflété par le méta-paramètre  $\gamma$  dans le modèle d'*apprentissage par renforcement*.
- 3) La noradrénaline contrôle la balance entre *exploration* et *exploitation* : ce neuromodulateur a donc des conséquences sur la modulation de l'inverse de la température  $\beta$ .
- 4) L'acétylcholine contrôle le taux de mise à jour de l'information par l'agent, reflété par le méta-paramètre  $\alpha$  dans le modèle d'apprentissage par renforcement.

De plus récentes études semblent néanmoins montrer de nouveaux résultats et l'implication de ces neuromodulateurs dans la modification d'autres paramètres cognitifs et comportementaux.

En ce qui concerne la dopamine, une récente étude chez le rat a montré que l'injection intraveineuse d'antagonistes des récepteurs dopaminergiques D1 et D2, mimant ainsi une interruption de la neuromodulation dopaminergique, induit une plus grande propension des animaux à mettre en œuvre une stratégie de type « je gagne, je change de réponse à l'essai suivant » (*win-shift*). Ce résultat est révélateur dans cette tâche d'un comportement plus *exploratoire*. L'intégration de ces données dans différents modèles d'apprentissage par renforcement permet aux auteurs de conclure que l'inhibition de la neuromodulation dopaminergique se traduit computationnellement par une diminution de l'inverse de la température  $\beta$ , et donc effectivement par une augmentation de l'*exploration aléatoire* (Cinotti et al., 2019).

Ces résultats semblent cohérents avec l'étude Koralek et collaborateurs qui montrent qu'une augmentation pharmacogénétique de l'activité dopaminergique promeut un état *exploitatoire* des souris (Koralek and Costa, 2019). A l'inverse, l'activation des neurones du locus coeruleus, et donc une neuromodulation accrue par la noradrénaline, favorise l'*exploration*, ce qui pourrait peut-être se traduire d'un point de vue computationnel par une diminution de l'inverse de la température par exemple (Koralek and Costa, 2019).

A noter, dans ces études, seule l'*exploration* de type aléatoire semble être affectée par la neuromodulation. Or des souris génétiquement modifiées, n'exprimant pas la sous-unité  $\beta 2$  des récepteurs à l'acétylcholine au niveau de l'aire tegmentale ventrale, semblent montrer un défaut d'*exploration* des options les plus incertaines dans une tâche de bandit. La modélisation par un modèle d'*apprentissage par renforcement* ajoutant un bonus à l'*exploration* des options incertaines semble rendre compte au mieux des résultats (Naudé et al., 2016). Il semblerait donc que l'acétylcholine via son action sur les récepteurs nicotiniques  $\beta 2$  de l'aire tegmentale ventrale promeuve l'*exploration dirigée*. A noter ici qu'il s'agit d'une neuromodulation d'un système neuromodulateur (système dopaminergique) : ceci illustre la possibilité d'une méta-neuromodulation au sein du système nerveux.

Quant à la sérotonine, une étude de Grossman et collaborateurs montre que dans une tâche probabiliste de renversement d'apprentissage durant laquelle les souris réalisent la tâche en système « tête fixée » permettant l'enregistrement de l'activité des neurones sérotoninergiques :

- La fréquence de décharge des neurones sérotoninergiques corrèle avec l'*incertitude attendue*
- La fréquence de décharge des neurones sérotoninergiques corrèle avec l'*incertitude inattendue* lors de l'observation des résultats des actions
- L'inactivation pharmacogénétique des neurones sérotoninergiques altère le comportement des souris qui s'adaptent moins rapidement aux modifications de contingences.

Les auteurs grâce à leur modèle computationnel de méta-apprentissage, pour lequel seul le taux d'apprentissage varie, concluent alors que la sérotonine est responsable de la modulation du taux d'apprentissage grâce à l'estimation des *incertitudes attendue* et *inattendue* (Grossman et al., 2020).

### 3.2.2. INFERENCES, TRANSITIONS ABRUPTES ET NEUROMODULATION

Il est intéressant de noter que les systèmes neuromodulateurs (dopaminergique et noradrénergique notamment) présentent deux modes de décharge différents : un mode tonique et un mode phasique. Cet état binaire de décharge des systèmes neuromodulateurs pourrait donc être à l'origine de transition brutale de l'activité de circuits neuronaux et donc de transitions cognitives et comportementales (Bouret and Sara, 2005; Budygin et al., 2020; Devilbiss and Waterhouse, 2011; Kane et al., 2017).

Des chercheurs ont montré que la neuromodulation noradrénergique du cortex cingulaire antérieur induit une stochasticité comportementale importante chez le rat qui semble alors s'affranchir de l'historique des récompenses et de l'utilisation d'un modèle interne de son environnement. L'inhibition de cette neuromodulation a en revanche l'effet inverse. La neuromodulation noradrénergique semble donc induire l'oscillation du cortex cingulaire antérieur selon deux modes dont les conséquences comportementales se traduisent par plus ou moins de variabilité comportementale (Tervo et al., 2014).

Ces travaux rejoignent notamment les théories d'Aston-Jones qui positionnent le système noradrénargique comme un système neuromodulateur critique de la balance *exploration – exploitation*. L'activité tonique du locus coeruleus induisant une augmentation de la concentration basale de noradrénaline favorise l'*exploration* tandis que l'activité phasique de ce noyau promeut l'*exploitation* (Aston-Jones and Cohen, 2005).

Il semblerait que la sérotonine ait également un rôle important dans la balance *exploration – exploitation*. Pour tester cette hypothèse, l'équipe de Mainen a développé une tâche probabiliste de fourrageage durant laquelle des souris doivent récupérer des récompenses à un des deux ports distribuant de la nourriture selon un calendrier de probabilité aléatoire qui décroît exponentiellement vers zéro. En raison de la nature probabiliste du programme de récompense, à chaque tentative (coup de nez dans un trou), la souris peut obtenir une récompense ou non. Néanmoins, passé un certain temps, plus aucune récompense ne sera distribuée. Le changement n'est donc pas brutal mais toute la question est de savoir quand quitter le port de distribution alimentaire qui s'épuise afin d'aller explorer l'autre. Par manipulation optogénétique du noyau du raphé, augmentant ainsi l'activité sérotoninergique, les auteurs montrent que ce neuromodulateur favorise l'attente en renforçant la persistance et l'*exploitation face à l'incertitude*, retardant ainsi la transition comportementale et donc une adaptation optimale face à un environnement changeant (Lottem et al., 2018).

### **3.2.3. ENJEUX SCIENTIFIQUES : INVESTIGUER L'IMPLICATION DE LA NEUROMODULATION DANS UN NOUVEAU PARADIGME EXPERIMENTAL**

Les derniers paragraphes permettent d'envisager la diversité des processus neuromodulateurs potentiellement mis en jeu dans le cadre des *comportements adaptatifs*. Néanmoins, sans observation préalable du comportement mis en jeu au cours de la tâche,

difficile de proposer un mécanisme neuromodulateur, d'autant plus s'il n'existe pas dans la littérature d'études semblables.

Dans le cadre de ma thèse, l'étude comportementale au sein du nouveau paradigme expérimental a été l'enjeu majeur de celle-ci. La présentation et la discussion des résultats permettront d'envisager les potentiels mécanismes neuromodulateurs impliqués et les perspectives de recherche à plus long-terme.

## **CONCLUSION**

Ce troisième chapitre sur les bases neurobiologiques des comportements adaptatifs a permis d'envisager le rôle du dACC et des systèmes neuromodulateurs dans la mise en place de stratégies comportementales d'adaptation. Ce chapitre permet de clore l'état de l'art nécessaire à la compréhension de mon travail de thèse qui sera présenté après un bref rappel des hypothèses et objectifs de travail.



## OBJECTIFS ET HYPOTHESES DE TRAVAIL

---

Au cours de mes quatre années de thèse, mon travail s'est orienté sur la compréhension des mécanismes de l'adaptation des souris à des changements abrupts de leur environnement en investiguant les stratégies comportementales ainsi que les bases computationnelles et neurobiologiques mises en jeu.

Comme je l'ai précédemment détaillé au cours de l'état de l'art, différentes stratégies et différents mécanismes computationnels ont été proposés pour expliquer la capacité des êtres vivants à s'adapter. Des modèles simples comme l'apprentissage par renforcement, plus complexes comme l'utilisation de task-sets, des processus de méta-apprentissage ou encore d'inférences sur les états cachés de l'environnement ont été investigués. Ces processus sont assez bien caractérisés chez les primates humains et non humains, tout comme leurs bases neurobiologiques qui semblent largement mettre en jeu le cortex cingulaire antérieur dorsal ainsi que les systèmes neuromodulateurs. Néanmoins, chez les rongeurs, du fait de la simplicité et des multiples limitations des paradigmes expérimentaux classiquement utilisés en laboratoire pour étudier cette question, peu de données existent quant à l'existence de ces différents mécanismes chez ces animaux, pourtant couramment employés pour comprendre le fonctionnement du cerveau humain.

L'enjeu principal de ma thèse a donc été de développer une nouvelle tâche comportementale dérivée de la tâche d'apprentissage par renversement classique mais en la complexifiant. Cette tâche est fondée sur l'apprentissage de plusieurs associations entre des images et des actions à réaliser (appuyer sur un bouton de réponse). Au cours de la tâche, deux changements de contingences vont avoir lieu : c'est ce que l'on appelle les *switchs*. Un premier dit *total* où toutes les associations (figure – bouton de réponse) sont inversées (comme un renversement d'apprentissage mais avec plus de deux figures) et un second, dit *partiel*, durant lequel une association reste la même tandis que les autres sont inversées. Cette astuce a été pensée pour désambiguïser les différentes catégories de stratégies adaptatives mises en jeu, ainsi que les classes de modèles sous-jacents, en utilisant uniquement des analyses comportementales descriptives (**Figure 1**).

En considérant l'ensemble de ces perspectives, mon travail de thèse a présenté 3 objectifs.

**OBJECTIF 1 : Mettre en place un système adapté aux besoins expérimentaux de l'étude**

L'implémentation d'une tâche aussi complexe que celle évoquée précédemment requiert indéniablement un système expérimental adapté pour travailler avec des animaux modèles comme la souris afin : (1) de faire en sorte que l'animal soit capable de réaliser la tâche dans les meilleures conditions possibles, (2) de pouvoir collecter un grand nombre de données sur de longues périodes pour en extraire des analyses pertinentes et les utiliser dans le cadre de futurs travaux de modélisation.

Afin de répondre à cet objectif, j'ai utilisé mes connaissances et compétences acquises en électronique et programmation afin de développer un système de chambre opérante automatisée totalement open-source. Ces travaux seront présentés dans le Chapitre 1 de la section « Travail expérimental » sous la forme de mon premier article méthodologique.

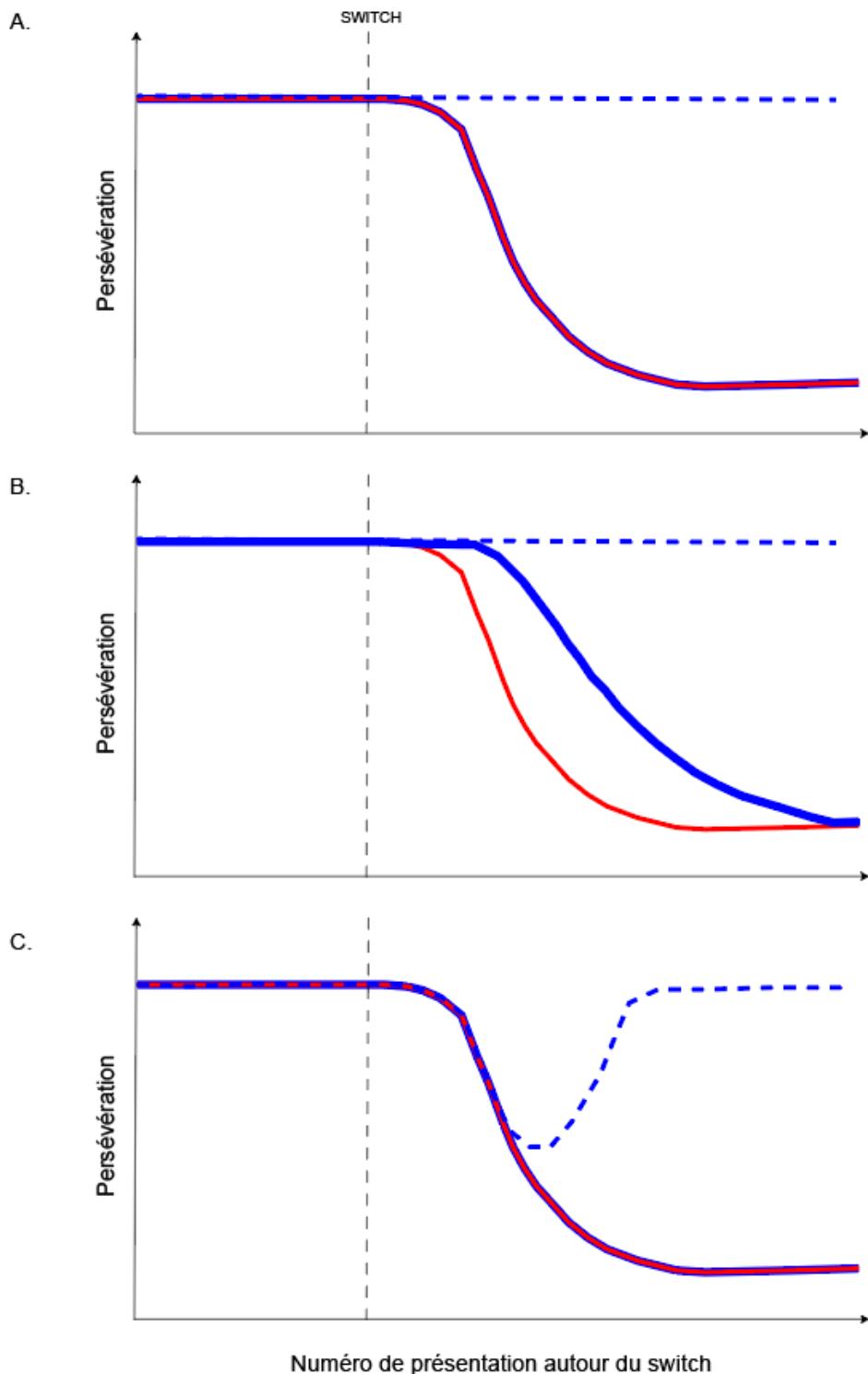
**OBJECTIF 2 : Mettre en place la nouvelle tâche comportementale de *switch* pour comprendre les stratégies d'adaptation de la souris en environnement *volatile* et en analyser les données**

Le deuxième objectif de ma thèse a été de mettre en place chez la souris le paradigme expérimental de *switch* et : (1) de s'assurer que les souris étaient capables de résoudre cette tâche, (2) d'analyser les stratégies comportementales mises en jeu, notamment lors des différents changements de contingence en fonction des hypothèses énoncées (**Figure 1**).

Les souris ont été capables de réaliser la tâche de *switch*. Dix-sept ont pu ainsi effectuer deux changements différents de règles et j'ai pu analyser les milliers de données collectées grâce à des scripts originaux développés par mes soins. Les résultats de cette étude seront présentés et discutés dans le Chapitre 2 de la section « Travail expérimental » dans le cadre d'une première version d'article.

**OBJECTIF 3 : Caractériser les bases neurobiologiques des stratégies comportementales mises en jeu au cours de l'adaptation comportementale de la souris en environnement *volatile***

A la fin de ma thèse, j'ai pu commencer à mettre en place un protocole de neuromodulation pharmacologique afin d'évaluer les conséquences de l'inhibition cortex cingulaire antérieur dorsal sur le comportement des souris. Ces expériences et résultats préliminaires seront présentés au sein du Chapitre 3 de la section « Travail expérimental ».



**Figure 1 : Evolution de la persévération des animaux autour des changements de règles opérés (switchs) au cours de la tâche.** La persévération quantifie la propension de l'animal à répondre selon les règles en vigueur avant le changement. Courbe rouge pleine : évolution de la persévération pour les figures dont le bouton de réponse change lors du *switch total*. Courbe bleue pleine : évolution de la persévération pour les figures dont le bouton de réponse change lors du *switch partiel*. Courbe bleue pointillée : évolution de la persévération pour la figure dont le bouton de réponse ne change PAS lors du *switch partiel*.

**A. Prédictions d'un modèle d'apprentissage par renforcement simple.**

L'agent apprend de manière indépendante chaque association : les valeurs sont mises à jour à chaque résultat observé (récompense ou non) à la suite du choix d'une action pour une figure donnée. Les paramètres du modèle (taux d'apprentissage, taux d'exploration...) sont fixes tout au long de la tâche et identiques pour les différentes associations. De ce fait, toutes les associations sont apprises et désapprisées de la même manière à la suite des changements, sauf l'association qui reste valide après le changement puisque le sujet continue de recevoir des récompenses en appliquant la même politique.

**B. Prédictions d'un modèle de mét-apprentissage par renforcement.**

L'agent également apprend de manière indépendante chaque association comme précédemment. Néanmoins, cette fois, les paramètres du modèle (taux d'apprentissage, taux d'exploration...) ne sont pas fixes et peuvent varier tout au long de la tâche notamment en fonction de la *volatilité estimée* de l'environnement dans lequel évolue l'agent. Or après un changement partiel des règles, l'environnement peut paraître moins volatil, et les paramètres du modèle d'apprentissage peuvent donc être différents de ceux après le switch total, expliquant ainsi le décalage entre les courbes bleu et rouge continues.

**C. Prédiction d'un modèle utilisant des *task-sets*.**

L'agent apprend chacune des associations comme faisant partie d'un ensemble cohérent d'informations. À la suite d'un changement abrupt de l'environnement, lorsqu'un changement est observé sur une des figures il peut être alors possible d'inférer que tout a changé. En suivant cette logique, ce modèle permet de proposer une prédiction essentielle : à la suite d'un changement partiel des contingences, la persévération pour la figure dont le côté de réponse n'a pas changé doit diminuer (courbe bleue en pointillés). En effet, si l'agent résout la tâche en utilisant des sets, l'observation suivant laquelle les autres associations ne sont plus correctes pourrait pousser l'agent à supposer qu'il en est de même toutes les figures : l'agent explore donc temporairement l'autre alternative, ce qui se traduit par une diminution transitoire de la persévération.

## **TRAVAIL EXPERIMENTAL**

---



# **CHAPITRE 1. Développement d'un système open-source automatisé et à haut rendement pour l'étude des comportements animaux complexes sur de longues périodes d'étude : le projet BEATBOX**

---

## **1. RESUME INTRODUCTIF**

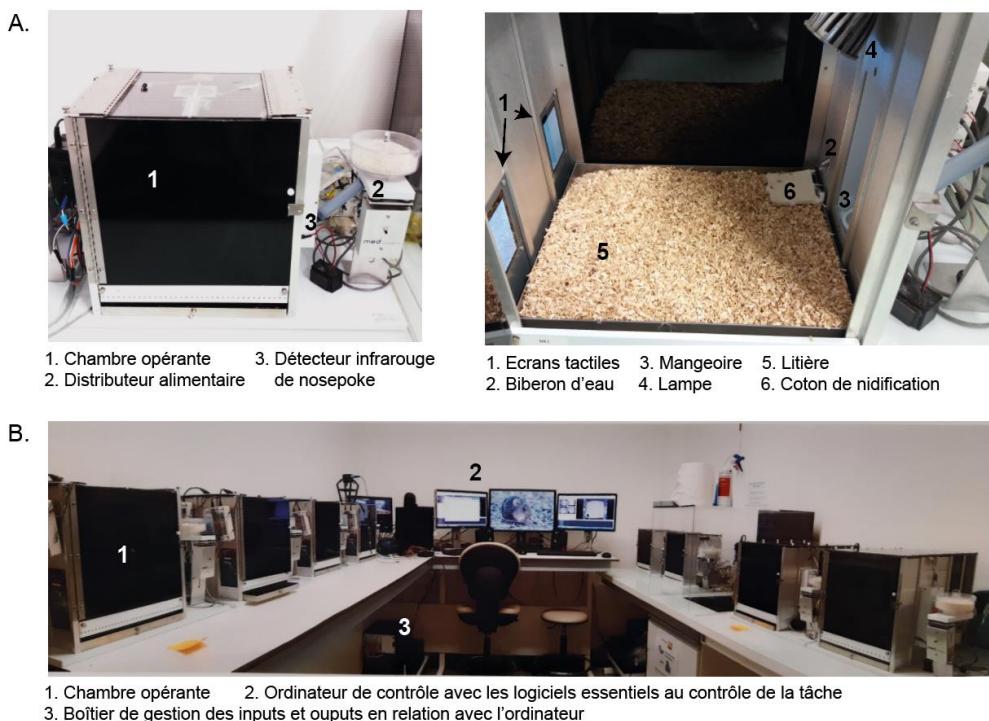
### **1.1. HISTORIQUE ET ENJEUX DE L'ETUDE**

Les systèmes d'expérimentation comportementale les plus couramment utilisés aujourd'hui au laboratoire présentent plusieurs limitations. Les tâches comportementales, par exemple, sont le plus souvent réalisées pendant 30 minutes à 2 heures par jour. Outre l'aspect chronophage que ce défaut d'automatisation représente pour l'expérimentateur, la quantité de données collectées est assez restreinte, et le nombre de répétitions de la mesure d'intérêt est limité. Dans le cadre du *reversal learning* chez la souris par exemple, un à deux renversements sont le plus souvent obtenus en comparaison aux dizaines réalisés par les humains ou les primates (Izquierdo et al., 2017). Ceci pose donc de nombreux problèmes lorsque la réalisation de tâches très complexes est requise à l'instar de la tâche de *switch* nouvellement développée dans le cadre de ma thèse qui met en jeu trois associations (stimulus - réponse) ainsi que deux renversements de contingences expérimentales complexes.

De plus, bien que les protocoles expérimentaux impliquant des animaux respectent la règle des 3R (remplacer, réduire, raffiner) (Richmond, 2000), les procédures menées en laboratoire, les plus légères soient-elles, ne sont pas exemptes de causer du stress aux animaux. Une simple routine de changement de cage a un effet significatif sur le bien-être de l'animal (Balcombe et al., 2004). Les paradigmes comportementaux implantés au laboratoire sont de ce fait encore plus stressants car ils impliquent le plus souvent de la privation alimentaire ou hydrique afin de motiver les animaux, une manipulation quotidienne en dehors des cycles physiologiques d'éveil (les rongeurs sont nocturnes) par des expérimentateurs plus ou moins stressants (Sorge et al., 2014). Le stress ainsi causé peut largement altérer les capacités des animaux dans les tâches comportementales, biaisant ainsi

les résultats (Hurtubise and Howland, 2017) et empêchant leur reproductibilité à travers les laboratoires (Aguillon et al., 2020).

Dans cette perspective, le doctorant me précédant sous la direction du Dr Eric BURGUIERE a développé une nouvelle méthodologie pour implémenter une tâche de *reversal learning* de discrimination visuelle chez la souris (Mar et al., 2013). Il a adapté des chambres opérantes commerciales afin d'héberger les rongeurs 24 h / 24, 7 jours / 7 pendant plusieurs semaines (**Figure 1A**), les souris réalisant la tâche à leur rythme, sans restriction alimentaire (Benzina et al., 2021). L'adaptation tant matérielle que logicielle a permis l'automatisation de sa tâche et la mise en fonctionnement de huit chambres opérantes simultanément, lui permettant de collecter un grand nombre de données quasiment sans aucune intervention humaine (**Figure 1B**) (Benzina, 2019).



**Figure 1 : Système expérimental automatisé dérivé d'un dispositif commercial.** A. Chambre opérante commerciale modifiée pour une tâche automatisée de *reversal learning* de discrimination visuelle et adaptée pour la vie des souris en continu. A gauche : vue extérieure. A droite : vue intérieure, étape d'entraînement. B. Aperçu de la salle expérimentale avec les huit cages comportementales en série.

Le système mis en place par le précédent doctorant répondait parfaitement à mes contraintes expérimentales du fait de son automatisation et de la durée des tâches qu'il permettait d'implémenter. Grâce à ces chambres opérantes, il m'était alors possible de tester chez les souris ma tâche comportementale de *switch*. J'ai néanmoins migré la gestion

logicielle sous Matlab, me permettant de développer un code plus facile d'utilisation et adaptable pour d'autres tâches comportementales de l'équipe.

Malgré son efficacité, le système utilisé dérivait d'un système commercial excessivement cher (près de 30 000 euros pour 8 chambres) et surtout, très peu flexible. J'ai notamment souhaité remplacer au début de ma thèse la réponse sur les écrans capacitifs par un système de détection infrarouge placé devant les écrans. Ce dispositif beaucoup plus sensible diminuait les risques de non-détection de la réponse des souris si celles-ci répondaient avec les griffes. La réalisation de ce nouveau module de détection des réponses a été particulièrement laborieuse puisque cela a nécessité l'inversion de la logique de l'activité des inputs ainsi que la mise en place d'une alimentation électrique alternative puisque le système commercial utilise une tension de 21V, non compatible avec des composés électroniques classiques. Deux circuits imprimés, un Arduino, deux pièces 3D et deux mois plus tard, le dispositif était fonctionnel (**Figure 2A**). Cette faible flexibilité du dispositif se retrouvait également dans la construction même de la cage. Le matériel ne pouvait s'ajouter que le long de rails dont l'emplacement était prédéfini et non modifiable (**Figure 2B**). Ceci entravait donc grandement l'imagination des membres de l'équipe en termes de conception de nouvelles tâches comportementales, sans compter le fort encombrement spatial et la complexité du câblage qui limitaient la mobilité et le partage du matériel entre les différents expérimentateurs (**Figure 2C**). Enfin, ces chambres opérantes n'étant initialement pas conçues pour héberger des animaux 24 h /24, 7 jours / 7, leur entretien était particulièrement complexe (**Figure 2D**).



**Figure 2 : Difficultés rencontrées avec le système commercial adapté.** **A.** Boîtier développé pour passer d'une réponse tactile à une détection par des faisceaux infrarouges. L'adaptation a requis l'usage d'un Arduino, de deux circuits imprimés et deux pièces 3D créées spécifiquement pour cet usage. **B.** La conception de la cage avec l'emploi de rails limite la variété de disposition et de choix du matériel. **C.** Le dispositif est particulièrement encombrant et délicat au niveau des connexions. **D.** Les chambres ne sont pas adaptées à la vie en continu des animaux : ici l'exemple d'une souris urinant en permanence dans la mangeoire.

## 1.2. LA BEATBOX : UNE NOUVELLE CHAMBRE AUTOMATISEE OPEN-SOURCE A HAUT RENDEMENT D'ACQUISITION DE DONNEES

Afin de contourner les difficultés tout en conservant tous les avantages du système développé avant mon arrivée dans l'équipe, il devenait nécessaire de proposer un nouveau système de chambre opérante pour lequel l'expérimentateur aurait une maîtrise complète depuis la conception jusqu'à la réalisation de tâches comportementales. Ces chambres ainsi développées sont (**Figure 3**) :

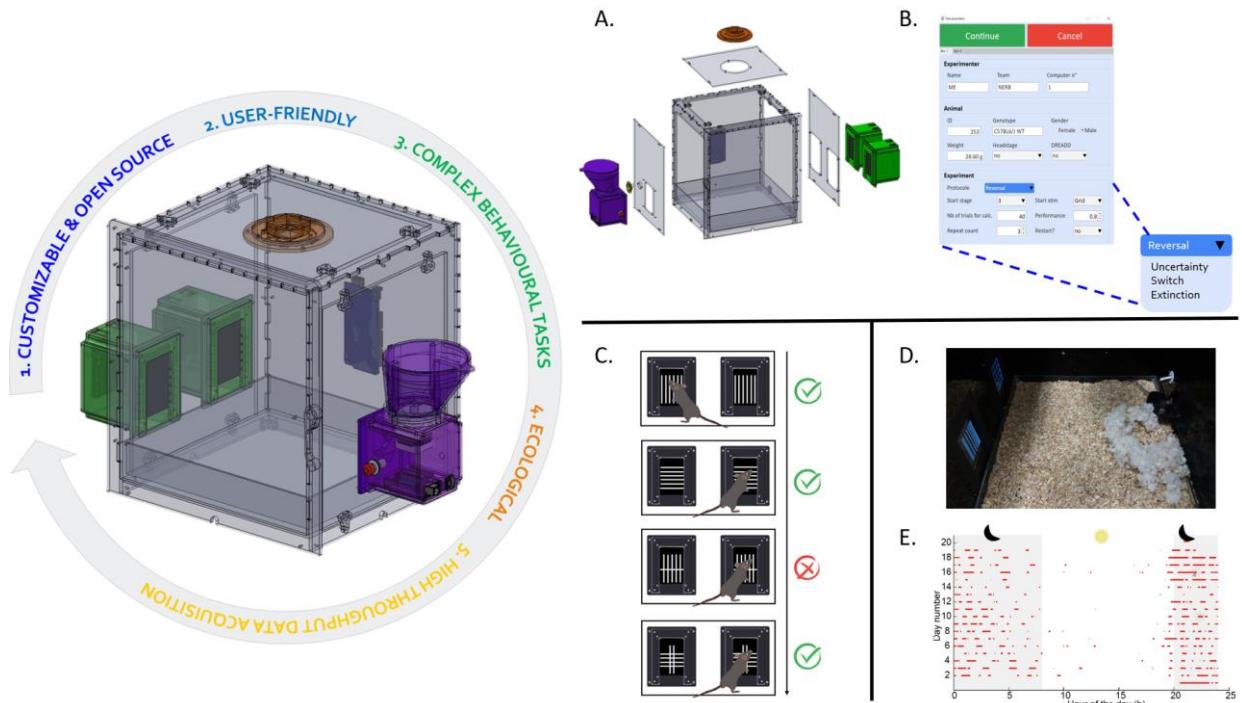
- 1) Hautement modulables tant d'un point de vue matériel que logiciel, permettant le développement de n'importe quelle tâche comportementale. Par ailleurs, le système mis en place est open-source : tous les plans, codes et conseils sont fournis permettant ainsi une utilisation libre au sein de n'importe quel laboratoire. Le matériel est également peu cher, la construction et la prise en main aisées, facilitant ainsi l'accès à cette technologie par le plus grand nombre.
- 2) Automatisées assurant ainsi un rendement maximal de la collecte de données (plusieurs milliers d'essais par animal) sur des semaines sans requérir la présence de l'expérimentateur. En plus de garantir un gain de temps précieux pour ce dernier, l'automatisation permet aux animaux de travailler à leur rythme, suivant leurs besoins et sans stress. Ce système expérimental est donc particulièrement écologique.

Une fois développées, ces chambres comportementales ont été testées avec des souris réalisant une tâche de *renversement d'apprentissage*. Placés de manière continue dans les chambres opérantes, les animaux ont été capables de réaliser la tâche à leur rythme et il a alors été possible de collecter en un temps record plusieurs milliers d'essais par souris, un voire plusieurs renversements par animal, contrairement aux souris contrôle qui ont réalisé la même tâche selon les procédures expérimentales classiques de la littérature (une session d'une heure par jour). L'ensemble de ce travail a abouti à l'article à suivre.

La BEATBOX est aujourd'hui utilisée en routine par plusieurs personnes de l'équipe et a été adaptée à leurs besoins expérimentaux permettant ainsi de valider ce nouveau dispositif expérimental. Ce nouvel outil permet la réalisation de tâches comportementales originales et novatrices garantissant ainsi l'étude de processus comportementaux et mécanismes

Stratégies d'adaptation de la souris face à un environnement volatil  
**TRAVAIL EXPERIMENTAL. Chapitre 1. Projet BEATBOX**

computationnels complexes grâce à la collecte de milliers d'essais et de multiples variables comportementales.



**Figure 3 : Aperçu de la BEATBOX et de ses avantages.** Adapté de (Euvrard et al., 2019). **A.** La BEATBOX est une chambre opérante modulaire fondée sur une conception en panneaux de Plexiglass vissables / dévissables sur lesquels peuvent être ajoutés des modules répondant aux besoins de l'expérimentateur (écrans, distributeur de récompenses...). L'ensemble des modules est piloté par la technologie open-source Arduino. **B.** Le système logiciel est également hautement modulaire, permettant la sélection et la réalisation de multiples protocoles expérimentaux. L'utilisation d'interfaces très simples fait de la BEATBOX un système accessible par tous. **C.** Exemple d'une tâche comportementale complexe durant laquelle la souris doit apprendre qu'elle doit répondre à gauche lorsque le stimulus qu'elle observe présente une majorité de barres verticales, et à droite lorsque le stimulus présente une majorité de barres horizontales. Tâche étudiée par (Schreiweis et al., 2019). **D.** La souris peut vivre en continu pendant plusieurs semaines dans la BEATBOX puisque la chambre est équipée selon ses besoins (litière, coton de nidification, eau...) et lui permet de réaliser la tâche à son rythme. **E.** Exemple d'un actogramme de souris réalisant la tâche de *renversement d'apprentissage*. Chaque trait rouge indique l'enchaînement de plusieurs essais. La souris travaille la nuit, selon son rythme circadien et il est ainsi possible de cumuler des milliers d'essais en 19 jours d'expérimentation.

## 2. DEVELOPPEMENT MATERIEL, LOGICIEL ET VALIDATION EXPERIMENTALE : LE PROJET BEATBOX

Article en cours (soumission prochaine dans un premier temps sur BioRxiv) : page 117 à 151

Supports (extraits d'un tutoriel de construction d'un des modules) : page 153 et 154

Stratégies d'adaptation de la souris face à un environnement volatil  
*TRAVAIL EXPERIMENTAL. Chapitre 1. Projet BEATBOX*

---

## ARTICLE 1

Stratégies d'adaptation de la souris face à un environnement volatil  
*TRAVAIL EXPERIMENTAL. Chapitre 1. Projet BEATBOX*

# The BEhavioural and AuTonomous Box (BEATBox): an open-source automated, customizable, and low-cost system for high-throughput data acquisition

M. Euvrard<sup>1</sup>, N. Benzina<sup>1</sup>, P. Tissier<sup>1</sup>, Y. Travert Jouanneau<sup>1</sup>, M. Bacci<sup>1</sup>, G. Penderia<sup>1</sup>, S.L. Mondragón-González<sup>1</sup>, E. Burguiere<sup>1,\*</sup>

1. Sorbonne Université, Institut du Cerveau - Paris Brain Institute – ICM, Inserm U 1127, CNRS UMR 7225, AP-HP, 47 bd Hôpital de la Pitié Salpêtrière, 75013 Paris, France

\*Corresponding author: [eric.burguiere@icm-institute.org](mailto:eric.burguiere@icm-institute.org)

## SUMMARY

Understanding the complexity of behaviours as well as their underlying neurobiological processes in normal and pathological conditions has become a major goal in Neurosciences. For that purpose, the use of rodent models has become predominant to monitor, map or manipulate their neural activity while they are performing behavioural tasks. Nevertheless, with conventional behavioural procedures, often performed in commercial systems, the amount of data can be limited, and their interpretation can be biased by the way they are collected. In fact, daily manipulation of the animals, food or water deprivation, timing of the experiment can perturbate their circadian rhythm, anxiety level, and motivational state. Thus, improving behavioural setups and protocols has become a major challenge to refine the quality of the collected data by better considering the animal's natural physiology and behaviours. In this respect, we have developed the Behavioural and Autonomous operant Box (BEATBox) where the animal can live while it is performing the task during the entire behavioural procedure. The animal self-paces its engagement in the task according to its needs with barely any interaction with the experimenter. We show that these automated operant chambers allow the administration of complex behavioural procedures and ensure high quality results thanks to the massive amount of collected data per subject. The entire hardware and software are open source, making the BEATBox easily improvable and adaptable by the scientific community at a lower cost than commercial devices.

**Keywords:** Animal behaviour, open-source, high-throughput, operant chamber, ecological, user-friendly, reversal learning, behavioural tasks, neuromodulation, recording

## INTRODUCTION

Behavioural neurosciences is aiming at understanding the complexity of behaviours and their underlying neurobiological processes at stake (Gomez-Marin and Mainen, 2016; Mainen et al., 2016; Niv, 2020), in both normal and pathological conditions. In this purpose, animal models have become essential over the past years as there are multiple powerful tools to monitor (Siegle et al., 2017; Steinmetz et al., 2018) and/or manipulate their neural activity (Park and Carmel, 2016; Whissell et al., 2016) while they are performing behavioural tasks. Nevertheless, several challenges remain to refine and fully exploit animal models behavioural data. First, we need to accurately reflect the behavioural outputs by collecting enough data per subject. To ensure the good quality of these data, the animal should be tested in appropriate conditions where its naturalistic rhythm is preserved. Second, current neuroscientific questions call for the design of novel complex tasks that can reflect, as much as possible, some sensorimotor or cognitive process observed in humans (Nithianantharajah et al., 2015). Thus, we need to translate challenging behavioural tasks for mice that require highly flexible design of the experimental apparatus and sometime long-lasting training of the animals.

For this purpose, the development of operant chambers to implement complex behavioural tasks like the human paradigms have been extensively used. Recently, the diversity and complexity of learning procedures have increased thanks to the use of visual touchscreens (Bussey et al., 2008; Hvoslef-Eide et al., 2016; Mar et al., 2013). Yet, these kinds of commercial devices remain overly expensive with little flexibility. As a result, some teams have developed open-source and low-cost systems with the same purpose (Devarakonda et al., 2016; O’Leary et al., 2018; Pineño, 2014; Steurer et al., 2012).

Although these systems have drastically improved the quality of the implemented behavioural tasks, some limitations remain. Despite a large progress in automation (Schaefer and Claridge-Chang, 2012), the experimenter is still involved in numerous steps along the experimental process since most of the studies rely on a discontinuous protocol during which the animals perform the task every day for 30 minutes to several hours (O’Leary et al., 2018; Pineño, 2014). Not only this limits the amount of acquired data but it also introduces a great amount of variability as multiple factors can alter the results of animal studies (Balcombe et al., 2004; Vissers et al., 2003). For example, an everyday handling of the animals can be particularly stressful for the animals, especially if the handling is not properly done (Gouveia and Hurst, 2017; Hurst and West, 2010). It has also been shown that the presence during the procedure and/or the sex of the experimenter can affect apparent baseline response of the animal (Neely et al., 2018; Sorge et al., 2014). Another common deleterious aspect of behavioural protocol that rely on the motivational state of the animal is that experimenters often practice water or food restriction before and during the experimental protocol. However, it has been shown that food restriction may increase stress hormones (Guarnieri et al., 2012; Levay et al., 2010) and may lead to behavioural performance alteration (Tucci et al., 2006). Last, the time of the day at which the experiment is run is also a crucial point to take into

account (Francis et al., 2019). Indeed, rodents are nocturnal animals and most of the current experiments are performed during the day, which may bias the data and lead to inconsistent results between studies. To palliate these limitations, we have developed the Behavioural and Autonomous operant Box (BEATBox). This is an operant chamber highly customizable by design and entirely produced with 3D technology and laser cut Plexiglas panels. It is operated by the flexible Arduino technology (D'Ausilio, 2012) and is entirely piloted by an in-house software run on a computer. The BEATBox system allows the administration of a various range of complex and long-lasting behavioural procedures to multiple animals at the same time. The main interest of this autonomous system is that mice can live and freely move in the operant chamber. Thus, the animals self-pace their activity to engage in the tasks with very few interventions of the experimenter. The automation of the system also guarantees a high throughput data acquisition. This ensures the development of standardized and novel protocols, the use of complex model-free and model-based analysis and a higher reproducibility of the results as it gets rid of the variability relative to the experimenter.

To demonstrate the utility of the BEATBox, we here compare a discontinuous procedure of a stimulus-based reversal learning task (1h in the task everyday) to our continuous procedure (24/7 behavioural training) in the BEATBox. We show that in the continuous condition, our system is highly efficient to quickly collect a large amount of high-quality data per animal even in complex tasks while respecting its overall physiology and drastically reducing experimenter's time.

Thanks to the BEATBox, we can refine the behavioural results obtain per animal by repeating the procedure over thousands of trials in a more naturalistic environment. Consequently, we can reduce the numbers of animals needed, which definitively abide by the three Rs (Replace the model, Reduce the number of animals, Refine the experimental methods to improve animal welfare) (Richmond, 2000). Finally, the entire hardware design as well as the associated software are open-source and deeply detailed with tutorials and step-by-step instructions. The BEATBox does not require any deep technical knowledge to build and operate it. This makes our system very user-friendly, easily adaptable and improvable by the scientific community at a much lower cost than commercial alternatives (White et al., 2019).

## RESULTS

### **The BEATBox is a highly customizable and user-friendly system**

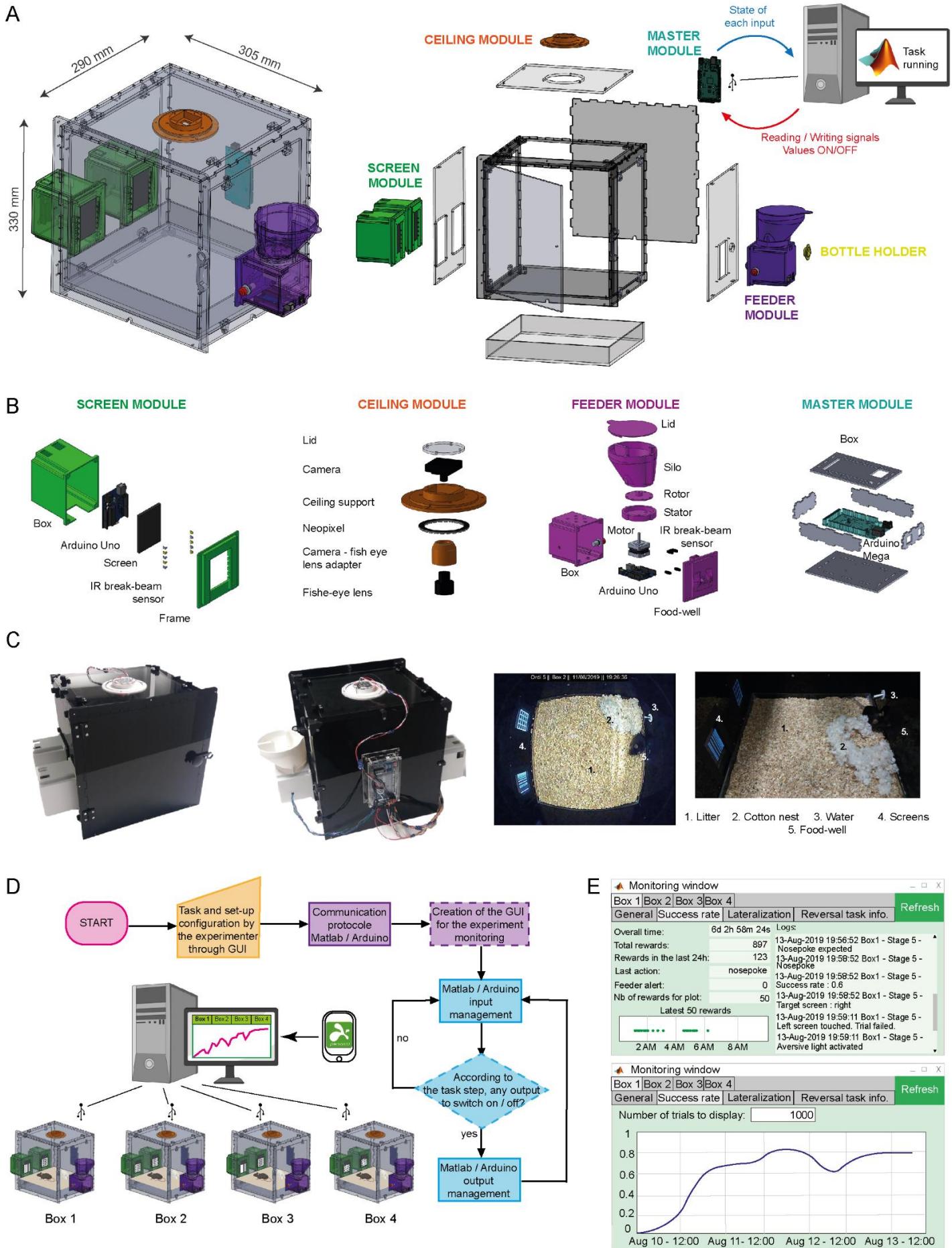
The BEATBox is an automated operant chamber that can be tailored to any behavioural tasks and any animal model. In our study, the dimensions have specially been selected for mice (**Figure 1A, left picture**). The overall resulting operant chamber is highly compact (330 mm \* 290 mm \* 305 mm, height\*width\*length) and light.

The BEATBox main structure consists in customizable screwable laser cut Plexiglas ® panels (**Figure 1A right picture, Video S1**) on which a large variety of modules can be added depending on the experimenter's needs. We display here the modules we used for our reversal task (screens, feeder, source of light/camera) (**Figure 1B**) but any other thinkable modules might be developed and improved later by the scientific community as our project is open source on our website <https://nerb.team/>.

The administration of behavioural tasks and the data recording in the BEATBox is ensured by a Matlab code running on a computer and interfacing through serial communication with a master Arduino in charge of the input / output management (**Figure 1A, right picture, Figure 1D, right picture, Figure S1A**). The timing performance of the BEATBox remains compatible with behavioural tasks administration with approximately 15 ms to gather the state of each input in every operant chamber (e.g, are the infrared break-beam sensors active?) and 50 ms to switch on/off the outputs in one box (e.g, switch on the screens) (**Figure S1B**).

The BEATBox software is also highly customizable with a main backbone code to initiate the tasks and to ensure the communication between the computer and the master Arduino. The only requirement for the experimenter is to add the specific functions relative to his / her task that basically consist in the output commands to execute depending on the active inputs (e.g, during the second training stage, if the mouse nosepokes in the food-well, a pellet must be delivered) (**Figure 1D, Figure S1, Figure S2, Figure S3**). Besides, the BEATBox software provides a personalized user-friendly monitoring interface (**Figure 1E, Figure S2**). This interface live-tracks the mice activity all along the task and displays graphic overviews of the evolution over time of any variable of interest (performance, reaction time, lateralization...) (**Figure 1E**). It is possible to directly monitor the experiments from the experimental area or remotely with smartphones thanks to the use of free commercial softwares or VPN access (**Figure 1D**).

Our system has been specifically designed for the animals to live 24h/day, 7d/week for several weeks in the BEATBox. The animals, here mice, can perform the behavioural task without any human intervention as they can trigger the trials on their own to get rewards. The mice are provided with *ad libitum* water and nesting material and still have access to the circadian light rhythm of the animal facility thanks to the transparent roof (**Figure 1C**). Several boxes can be run in parallel on the same computer at the same time with the same task or different tasks running in each (**Figure 1D**). This made the BEATBox system even more flexible as it can be used by one experimenter who needs to quickly collect data on several animals or by different users who can share the system as it is commonly the case in laboratories.



## **Figure 1 The BEATBox is a highly customizable and user-friendly system**

**(A)** 3D representation of the BEATBox. This operant chamber is highly modular thanks to customizable removable panels and modules. A single computer connected through serial communication to the master Arduino is required to run behavioural tasks within the BEATBox.

**(B)** Examples of the 3D-printed / laser-cut modules with all the electronic components used during the study (the Arduino shields are not shown).

**(C)** Pictures of a finalized BEATBox. From the left to the right: front view, back view, top and lateral views of the inside with one mouse performing a trial.

**(D)** Overview of the software management of the BEATBox. Several cages can be run in parallel and independently thanks to a simple loop in the Matlab code. The dashed lines in the flow chart indicate the parts of the code that can be customized by the experimenter. It is possible to remotely take control of the computer to check the monitoring interfaces and the progress of the mice in the tasks.

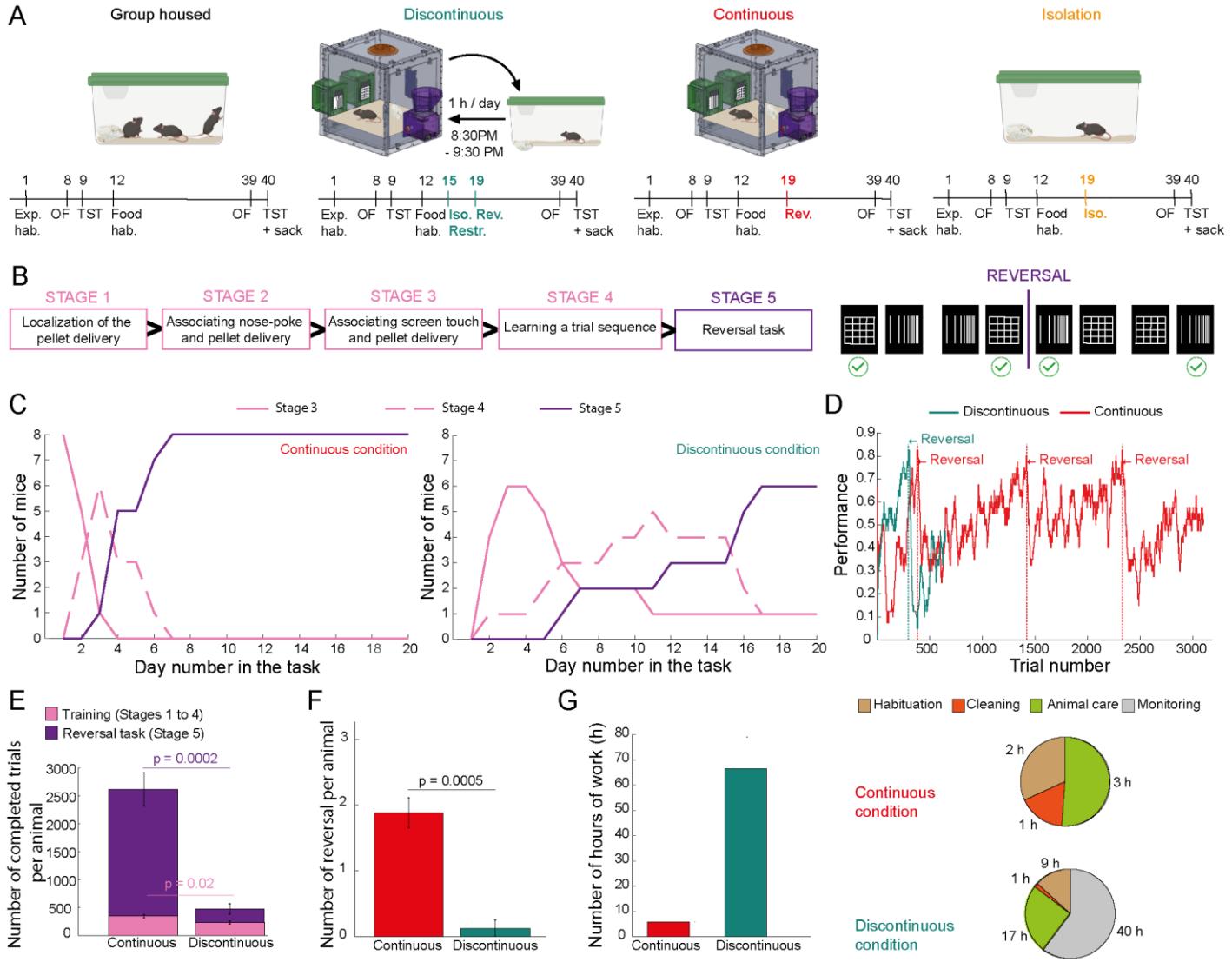
**(E)** Monitoring interface to live monitor the mouse activity during the task (top picture) and different parameters of interest (e.g. the performance, bottom picture).

## **The BEATBox ensures a high throughput data acquisition with little experimental effort**

To evaluate the efficiency and performance of the BEATBox in acquiring massive amounts of data, we compared two ways of administrating a behavioural task to mice. In our lab, we are implementing a continuous protocol during which the mice remain constantly in the BEATBox, 24h/d, 7d/week during the whole duration of the task (**Figure 2A**, ‘Continuous’ panel). We compared this protocol to the ones found in the literature (Horner et al., 2013; Hvoslef-Eide et al., 2016; Mar et al., 2013; Marquardt et al., 2017; Turner et al., 2017). We called these protocols ‘discontinuous’ as the mice can only perform the task for one hour a day in the operant chambers and then are isolated back to their home cage until the next session (**Figure 2A**, ‘Discontinuous’ panel). To control for the isolation effect, we added two controls: one mouse living in a group (**Figure 2A**, ‘Group’ panel) and one mouse constantly isolated in its homecage and not performing any behavioural task (**Figure 2A**, ‘Isolation’ panel).

In this study, we decided to use a touch-screen based reversal learning task as it has been already well documented (Izquierdo et al., 2017) and successfully implemented in our lab in a commercial system (Benzina et al., 2021). After the shaping stages that allow the mice to get used to the operant chamber (stages 1 to 4, **Figure 2B**, **Figure S3**), the mice were exposed to two visual stimuli (grid on one screen and vertical bars on the other screen, spatially pseudorandomized on the two screens over the trials) (**Video S2**, **Figure S4**). The mice had to answer on the screen associated with the rewarded stimulus thanks to the presence of infrared break beam sensors in front of each screen. When they reached 80 % of correct answers over the 40 latest trials, the contingencies were reversed (**Figure 2B**).

We first noticed that the mice in the continuous condition were faster to succeed in the training stages than the ones in the discontinuous condition. After seven days, all the mice in the continuous condition had reached the reversal learning phase (stage 5) whereas only two mice in the discontinuous condition did on the same day. Even after 20 days in the task, only 6 mice out of 8 had reached the final stage (**Figure 2C**).



**Figure 2 The BEATBox ensures a high throughput data acquisition with little experimental effort**

(A) Presentation of the four conditions used during the protocol and the associated timelines. OF: Open field. TST: Tail suspension Test. Food hab.: Food habituation. Exp hab.: Experimenter habituation. 2 test conditions (discontinuous and continuous) and 2 control conditions (group-housed and isolation) were used.

(B) Task protocol. The mice first successively underwent four training stages to get used to the BEATBox and finally reached the stage 5 during which they performed a deterministic stimulus-based reversal learning task (left picture). During this task, the mice had to learn which stimulus from the pair was rewarding and when they reached a 80% performance threshold over forty trials, the contingencies were reversed (right picture).

(C) Comparison of the progression in the different stages for the mice in the continuous condition (left) and in the discontinuous condition (right). The mice in the continuous condition reached the reversal learning task faster.

(D) Comparison of the performance of two mice (continuous versus discontinuous conditions) during the reversal learning task and localization of the performed reversals (data are smoothed according to a moving average window of 40 trials).

(E) Comparison of the number of completed trials during the training stages (stage 1 to 4) and the reversal learning task (stage 5) in the continuous versus discontinuous conditions ( $n = 8$  per condition, data are shown as mean  $\pm$  SEM, Wilcoxon rank sum test).

(F) Comparison of the number of reversals during the reversal learning task in the continuous versus discontinuous conditions ( $n = 8$  per condition, data are shown as mean  $\pm$  SEM, Wilcoxon rank sum test).

(G) Comparison of the estimated experimental time in the continuous versus discontinuous conditions. We consider here a total number of 4 mice per condition running the task.

The number of completed trials in the training phases significantly differ between the continuous and discontinuous condition ( $341.75 \pm 31.65$  trials for the mice in the continuous condition,  $231.5 \pm 227.80$  trials for the mice in the discontinuous condition,  $p = 0.0207$ , Wilcoxon rank sum test,  $n = 8$  for each condition, **Figure 2E**) and we collected 10 times more trials during the reversal stage (stage 5) for the mice in the continuous condition ( $2,272.37 \pm 297.12$  trials for the mice in the continuous condition,  $240.5 \pm 88.90$  trials for the mice in the discontinuous condition,  $p = 1.6e-04$ , Wilcoxon rank sum test,  $n = 8$  for each condition, **Figure 2E**). Besides, during our 20-day experiment, we were able to collect a mean of two reversals per mouse in the continuous condition whereas only one mouse reached the criterion in the discontinuous condition ( $1.875 \pm 0.2266$  reversal for the mice in the continuous condition,  $0.1250 \pm 0.1250$  reversal for the mice in the discontinuous condition,  $p = 4.7e-04$ , Wilcoxon rank sum test,  $n = 8$  for each condition, **Figure 2F**). Nevertheless, while this mouse reached one reversal, his paired littermate in the continuous condition performed three of them (**Figure 2D**). This massive amount of collected data in the continuous condition was possible even if the experimental time was reduced 10 times due to the complete automaticity of the protocol and the reduced need for monitoring (**Figure 2G**). We estimated these durations based on the time spent by the experimenter each day all along the procedure. The continuous BEATBox protocol is thus less labor-intensive and less time-consuming.

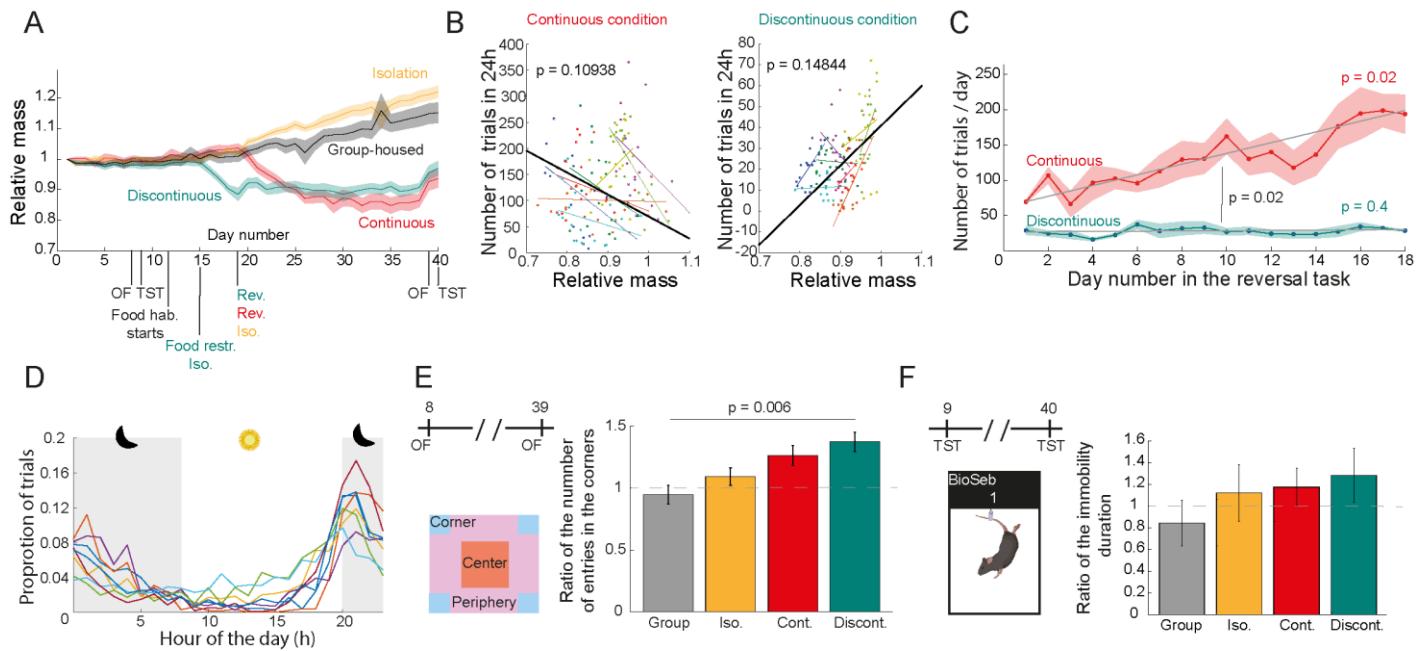
### **The BEATBox respects the natural rhythm of the animals**

We evaluated different parameters to assess the mice well-being all along the protocol to check whether a continuous protocol in the BEATBox might induce a stress in the animals.

We notably weighted the mice every day. For the mice in the continuous condition, we show that the animals did lose some weight as soon as they reached the most difficult stages in the task (stage 4 and stage 5), but their relative weight remained around or above the 85% of the free-feeding weight threshold (**Figure 3A**, **Figure 3B**). As the weight reached this plateau until the end of the reversal learning protocol, it is likely that the mice can self-regulate their weight and their involvement in the task accordingly to their metabolic need. We noticed indeed that for some mice in the continuous condition, there was a significant relationship between the relative mass and the number of performed trials: for three mice out of eight, the lower the relative mass, the more numerous the trials were (**Figure 3B right**, **Table S1 left**). Except for one mouse, the slopes are negative indicating that the mice are likely to work more if they need food. On the contrary, we noticed an opposite tendency for the mice in the discontinuous condition (**Figure 3B left**, **Table S1 right**) as if they gave up working when being underweight.

When looking at the number of completed trials, as a proxy of the mouse activity, we also noticed an increase over time of the number of completed trials per day for the mice in the continuous condition. This indicates that even after two weeks of continuous isolation and task resolution in the BEATBox,

these mice were even more active and involved (mean slope of activity,  $6.9227 \pm 1.5454$ ,  $p = 0.0156$ , Wilcoxon sign rank test for the slopes of all the  $n = 8$  mice, **Figure 3C, Table S2 left for individual data**). The mice in the discontinuous condition maintained the same level of activity (mean slope of activity,  $0.0461 \pm 0.3648$ ,  $p = 0.7422$ , Wilcoxon sign rank test for the slopes of all the  $n = 8$  mice, **Figure 3C, Table S2 right for individual data**). We can conclude that the mice in both conditions did not display apathetic behaviours despite the isolation and task training (**Figure 3C**).



**Figure 3 The BEATBox respects the natural rhythm of the animals**

(A) Evolution of the weight for the mice in the different conditions as a ratio with the initial weight on day 1 ( $n = 8$  per condition, data are shown as mean  $\pm$  SEM).

(B) Evaluation of the mice activity measured by the number of performed trials every 24h depending on their relative weight. Each colour represents one mouse in the continuous condition (left panel,  $n = 8$ ) and its paired littermate in the discontinuous condition (right panel,  $n = 8$ ). The dots represent one measure for each day and the lines represent the polynomial fit (degree 1) for each mouse. The black lines represent the mean of the individual fits (Wilcoxon rank-sum test on the slopes).

(C) Evolution of the number of completed trials per day over the training and the reversal task for the mice in the continuous and discontinuous conditions ( $n = 8$  per condition, data are shown as mean  $\pm$  SEM, Wilcoxon rank sum test on the slopes after polynomial fitting of degree 1).

(D) Distribution of all the trials performed during the reversal learning task along a nycthemeral rhythm for each mouse in the continuous condition.

(E) Evaluation of the mice anxiety with the open field test. A ratio superior to 1 indicates that the mice spent more time in the corners after the protocol (post data, acquired on day 39) than before (pre data, acquired on day 8); ( $n = 8$  per condition, data are shown as mean  $\pm$  SEM, Kruskal-Wallis test).

(F) Evaluation of the mice depression-like state with the tail suspension test. A ratio superior to 1 indicates that the mice were more immobile and struggled less after the protocol (post data, acquired on day 40) than before (pre data, acquired on day 9). ( $n = 8$  per condition, data are shown as mean  $\pm$  SEM, Kruskal-Wallis test).

To further investigate the animal wellness, we evaluated anxiety and depressed-like behaviours with the open-field test and tail suspension test. We noticed that the mice in the discontinuous condition tended to enter more the open field corners after the 20-day reversal learning task than the mice in the ‘group

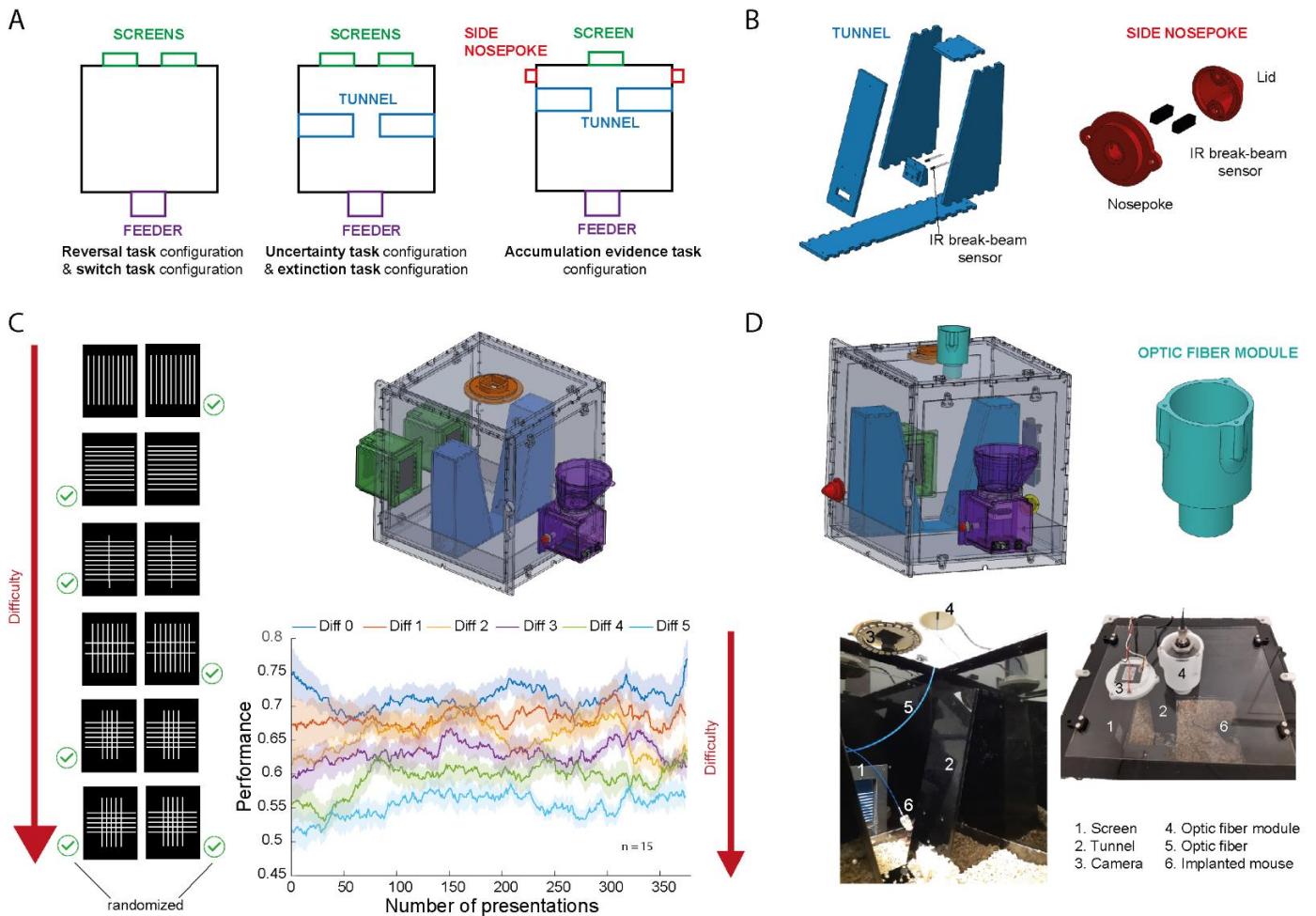
housing' control (ratio of the number of entries in the corner after the protocol versus before the protocol,  $0.9464 \pm 0.0747$  for the mice in the group condition,  $1.3739 \pm 0.0785$  for the mice in the discontinuous condition,  $p = 0.006$ , Kruskal Wallis test followed by a multicomparison test,  $n = 8$  mice per each condition), suggesting that a discontinuous protocol might induce a significant anxiety (**Figure 3E**). However, we did not replicate this tendency when looking at the other parameters of the open field (number of entries, time and distance travelled in the centre). All the groups did not differ in the tail suspension test ( $p = 0.4722$ , Kruskal Wallis test,  $n = 8$  for each condition, **Figure 3F**, **Figure S5**).

Last but not least, the BEATBox continuous protocol respects the natural rhythm of the mice. The mice could self-pace their own activity in the task and we clearly showed that they worked accordingly to their natural nocturnal rhythm as they mostly started trials during the dark phase (**Figure 3D**, **Figure S6A**). Thus, the BEATBox continuous protocol increases the ecological validity of the implemented tasks without inducing more anxiety than what lab routine protocols such as weighting would do.

### **The BEATBox allows the implementation of a large variety of innovative tasks**

In our lab, we have implemented a large diversity of tasks to answer various questions. Among these tasks, some are usual tasks, already implemented and validated in other labs such as the reversal learning task or extinction protocols, but three other ones have been designed from scratch (**Figure 4A**). For some of our tasks there was a need to better monitor the reaction time, that is why we developed a tunnel module with two infrared break beams to track the mouse position in the operant chamber (**Figure 4A, middle and right panel, Figure 4B, left**). When the mouse crosses this tunnel, the screens switch on and this ensures a better reaction time evaluation contrary to the tasks where the mouse had to nosepoke in the food-well to start a trial. For other tasks, a nosepoke module has been designed to collect the mouse answer when facing a stimulus on a unique screen (**Figure 4B, right**). Our innovative and flexible set-up was essential to develop these new behavioural paradigms that would have been impossible to implement if we did not have the BEATBox to collect enough data. (**Figure 4C**). For example, in the uncertainty task (Schreiweis et al., 2019), the mice were exposed to a set of 6 stimuli with different ratio of horizontal and vertical bars (from the least ambiguous stimulus with only vertical or horizontal bars to the most ambiguous one with an equal number of horizontal and vertical bars). The mice had to answer on the correct screen (e.g, more horizontal bars, answer on the left and vice-versa). 15 mice have been assessed in three months, with around 400 trials for each difficulty, resulting in an approximate total number of 2,400 trials per mouse. The results clearly showed that the more difficult the displayed stimuli were, the more chance-level the mice answered (each curve represents the mean for 15 mice of their smoothed performance over 50 answered trials, **Figure 4C**). We are currently working on a modified version of the uncertainty task, with a new configuration (one screen, two gates, two side nosepokes to collect the mice

answers) while implementing calcium imaging during the behavioural tasks thanks to an optic fiber module (**Figure 4D**).



**Figure 4 The BEATBox allows the implementation of a large variety of innovative tasks**

(A) Examples of the three main configurations of the BEATBox used in our team for 5 different behavioural tasks (schemes of the top views).

(B) 3D-printed / laser-cut tunnels and side nosepoke with the infrared break beam sensors.

(C) Example of the uncertainty task developed by Schreiweis and collaborators (Schreiweis et al., 2019). Stimuli with a various ratio of horizontal and vertical bars are displayed on the two screens once the mouse has crossed the tunnel. The mouse has to answer on the left screen if there are more horizontal bars (and vice-versa). The more ambiguous the stimulus (ratio tends to 1), the lower the performance ( $n = 15$  mice, data are shown as mean  $\pm$  SEM).

(D) The BEATBox has been designed to allow the recording and modulation of neuronal activity during behavioural tasks. We show here the BEATBox adapted to calcium imaging with a specific optic fiber module.

## DISCUSSION

**The BEATBox is an automated, sharable, flexible, user-friendly, and cost-effective operant chamber.**

With the BEATBox, we provide here a sophisticated, innovative and cost-effective automated operant chamber with a high degree of flexibility and customizability (**Figure 1**, **Figure 4**, **Figure S2**, **Video S1**).

Our open-source system answers the scientific needs for data and tools sharing, transparency, creativity and collaboration between all-around-the-world teams (White et al., 2019). The BEATBox also fits all the advantages of automated behavioural devices (Schaefer and Claridge-Chang, 2012). The experimental possibilities are endless thanks to the flexibility of the BEATBox. Any new or already validated psychology tasks can be implemented or refined (**Figure 4**). Providing open-source and sharable hardware and software will also guaranty a higher standardization of experiments in the scientific community as many teams will be able to use the same consistent devices thereby reducing any potential confound from employing different equipment. The BEATBox also ensures a high-throughput acquisition of multiple, various, and complex behavioural data (reaction time, intertrial interval, pellet recovery time, duration of self-exposure to visual stimuli...) over weeks for multiple animals self-training in parallel, saving time for the experimenter to perform cutting-edge quantitative analysis and modelling during this period (**Figure 2**, **Figure 4**). Besides, the limitation of the involvement of the experimenter has two other main advantages. It increases reproducibility between labs by increasing animal wellness. We get ride off the experimental variability induced by the experimenter as well as the stress bias on experimental behavioural data (Balcombe et al., 2004) (**Figure 3**, **Figure S3**) as the experimenter presence is only required at the beginning of the experiment to put the mice in the operant chambers. The rest of the task can be administrated and supervised remotely (**Figure 1D**, **Figure 1E**, **Figure S2**). Last but not least, our BEATBox protocol of 24h/d and 7d/w training ensures the access to naturalistic observations such as the circadian rhythm (**Figure 3C**, **Figure 3D**), the recording of physiological parameters such as the neuronal activity (**Figure 4D**) and the possibility to perform experiments at the very proper moment of the animal activity (for example, the administration of pharmacological compounds).

We have validated the BEATBox for the mouse specie but its flexibility (**Figure 1**) and cheapness easily allow its adaptation to other species as it has previously been done by Steurer and collaborators in order to cross-validate result in various animal species with translatable tasks (Steurer et al., 2012). The BEATBox costs around 900 \$, which is 5 times less than commercial devices. It remains more expensive than other open-source devices (Devarakonda et al., 2016; Gurley, 2019; O’Leary et al., 2018; Pineño, 2014) but the sophistication of our system implies such a cost. Yet, our system is easy and quick to build and very less complex than other devices (Erskine et al., 2019). The material is easy to find, and even if the BEATBox building requires laser cut and 3D printing machines, the development of Fablabs in many laboratories or universities palliates this issue.

The maintenance of the system is also extremely easy and quick; the overall effort for the experimenter to set-up and take care of the system is low. Yet, it is to be noticed, that the rewards we used during our task were particularly dusty and often blocked the feeder. This can lead to unintended extinction training affecting future behaviours and experimental data. As a result, we added in our code a checkpoint to verify that the pellet is correctly distributed (infrared sensor just above the rotor-stator, **Figure 1B**, **feeder**

**module)** and in case of repeated non distribution, the software displays a feeder alert in the monitoring window (**Figure 1E**). Since the end of the protocol of this article, we also modified the feeder distribution system to avoid the pellet dust blocking it. Nevertheless, it is highly recommended to regularly clean the pellet dispenser with air duster, but this step remains quick.

To have a better overview of the animal activity in the box, it could be possible to adapt already-existing open-source tracking system (Matikainen-Ankney et al., 2019) or video-tracking systems. Adding a system to monitor the liquid intake (Godynyuk et al., 2019) could also provide useful information, notably for pharmacological or pharmacogenetics studies. Sound-proofing the overall BEATBox could also be an interesting improvement to prevent the mice from stressful noises from the animal facility.

One limitation of our system remains. For historical reasons in the lab, we have developed our software with Matlab but we have a will to provide a version with Python (Perkel, 2015). We could also think of using other devices instead of Arduino such as Raspberry or Teensy as they have proven efficient for other behavioural systems (Gurley, 2019; O’Leary et al., 2018; Saunders and Wehr, 2019). Nevertheless, with appropriate coding (D’Ausilio, 2012), the Arduino microcontroller time performance still remains acceptable for behavioural experiments (Balzani et al., 2018; Francis et al., 2019) (**Figure S1**). Wireless operating and data recording (Amaral-Júnior et al., 2019; Balzani et al., 2018; Francis et al., 2019) in our BEATBox could also be a way to improve our system.

For the first release of the BEATBox, we provide the software to run the reversal learning task as well as the analysis scripts. To run this task, we share a default version, the parameters of which can be adjusted (number of reversals required, performance criteria...). The other tasks we are running in the lab will be made available as soon as they are validated, and we encourage every new task or analysis scripts that will be developed to be made freely accessible.

### **The BEATBox is an operant chamber specifically designed to run experiments continuously for several weeks.**

We have validated the BEATBox with a touchscreen based reversal learning task as it has been extensively studied (Izquierdo et al., 2017) in order to compare our results to the ones found in the literature. We notably used the same kind of training developed by Mar and collaborators (Mar et al., 2013) or Horner and collaborators (Horner et al., 2013) in order to gradually shape the screen-touching behaviour. Nevertheless, we implemented a few changes (self-pace intertrial interval, no light or sound cue, different criteria to change stage...) as previously validated in our team (Benzina et al., 2021). For the mice in the discontinuous condition (**Figure 2A, third picture**), we found that a mean number of 8/9 one-hour sessions was necessary to reach the reversal learning stage (**Figure 2B, Figure 2C**). This is in line with the results of Marquardt and collaborators who indicate a mean number of four weeks to go through the training (around 15 sessions of thirty minutes, if we remove the weight reduction) and the

results of Turner and collaborator (Turner et al., 2017) who report a mean of 17 sessions for the mice to succeed in the training stages.

It is to be noticed that for the mice in the discontinuous conditions, 3 mice successfully passed all the training stages reaching the criteria. They could properly perform in the reversal learning stage. We must report 1 mouse that never reached the fourth stage of training despite the easiness of the first three stages. We also report 4 mice that were slow to pass the training stages. For 3 of them, we decided to let them go manually to the reversal learning stages. But after analyzing the preliminary data, we noticed that 2 mice out of these 3 could not perform the reversal learning task properly. This was correlated with a bad performance over the 10 latest completed trials in stage 4 (10% and 20% contrary to 60% for the slow learner mouse that finally performed in the reversal task). We also have to indicate that one mouse in the continuous condition passed all the training stages but poorly performed in the reversal learning stage. It was taken back manually to the beginning of the training and after completing all the stages in less than one night, it performed well in the reversal learning stage. All the previous encountered issues clearly indicate that a proper shaping is essential before implementing the task of interest. A too quick training or an incomplete training are likely to lead to bad performances in the task. Since then, we have added in the code different conditions based on the number of trials and “give-up trials”. In the case they are met, they will bring back the mice to previous stages in the task.

For the visual acquisition phase, Mar and collaborators found that the C57BL6/J mice reached the 80% performance in about 12 sessions of 30 minutes (Mar et al., 2013). In Marquardt’s paper (Marquardt et al., 2017), the C57BL6/J mice achieved the discrimination phase in about 14 sessions of 30 minutes and a mean number of 8 sessions was required for Turner’s study with C57BL6/J mice (Turner et al., 2017). In our study, the C57BL6/J mouse that was able to perform one reversal achieved the performance criterion in 6 sessions of 1 hour. Yet, even the results seem consistent, they have to be considered with caution as the stimuli used were different and Turner clearly showed that depending on the stimulus, the acquisition phase can be faster or slower (Turner et al., 2017).

At the very opposite of the discontinuous protocol, the strength of our continuous experiments lies in the fact that we let the mice in the operant chambers 24h/d, 7d/w contrary to what we can find in the literature (Gurley, 2019; Horner et al., 2013; Mar et al., 2013; Marquardt et al., 2017; O’Leary et al., 2018; Pineño, 2014; Turner et al., 2017) even if more and more scientists get aware of the need to perform long lasting experiments continuously over days (Balzani et al., 2018; Erskine et al., 2019; Francis et al., 2019; Remmeliink et al., 2016; Torquet et al., 2018). We notably clearly show that implementing a continuous protocol fastens the training stages contrary to the discontinuous protocol (**Figure 2C**). Besides thanks to the continuous protocol, we can gather ten times more data in the same amount of time (**Figure 2E**, **Figure 2F**), with little involvement of the experimenter (**Figure 2G**) and a complete respect of the animal physiology (**Figure 3C**, **Figure 3D**). During this protocol we weighted the mice everyday but daily

remote checking of the number of rewards obtained over the 24 latest hours, the mouse activity, performance, and the mouse overall look with the top camera can be sufficient. By monitoring complex behavioural features with batteries of cognitive tasks as well as observing naturalistic behaviours (food intake, circadian rythm, activity, nest building....) we can easily extract endophenotypes and better characterize animals behaviour in both normal and pathological conditions (Balzani et al., 2018).

### **The BEATBox ensures a high-throughput data acquisition allowing an accurate data screening.**

We showed that we were able to collect 10 times more data in the reversal learning stage compared to classical discontinuous experiments (**Figure 2E**). This massive data acquisition embodies one of the strengths of our system. Yet, one should always remain cautious about the collected data. Mice are nocturnal animals and mostly perform the tasks during the nights (**Figure 2D**). Yet, some trials are still performed during daytime. One important question arises: are the data collected during the day equivalent as the data collected during the nights? One should always consider comparing the data from the nights and from the days to assess if they are different in terms of performance, type of trials (omissions of the answer, omissions of the pellets recovery) ... Once the answers assessed the experimenter can chose either to include all the data, filter them, or take into consideration the data variability in some modelling analysis. To that extent, the collection of thousands of data per animal is a major advantage as filtering the day trials would still leave a huge number of data and the analysis of different periods over 24 hours could be interesting to decipher the mice involvement in the task (Ashwood et al.)

### **The BEATBox continuous protocol is not as stressful as discontinuous protocols and ensures a better monitoring of the animals.**

With this study, we also wanted to evaluate to what extent the continuous BEATBox protocol was less stressful for the mice contrary to the discontinuous protocol. We compared these groups to two control groups, in order to have a baseline and a control for the isolation parameter (**Figure 2A**).

Prior to any testing, the mice were habituated for seven days to the only female experimenter with a strict protocol to ensure the best habituation and reduce the animal anxiety towards the experimenter (Hurst and West, 2010). Before and after the reversal learning task (**Figure 2A**, see the timelines), we implemented two classical tests to assess the anxiety level (open field test, **Figure 3E**) and depressive-like behaviours (tails suspension test, **Figure 3F**) to investigate the effect of the procedures on the mice. Since these tests still remain controversial as for their neuropsychological value (Fonio et al., 2012), all along the task, we also checked the weight (**Figure 3A**, **Figure 3B**, **Table S1**), scored the nest building (**Figure S6B**) (Deacon, 2006; Jirkof, 2014), and analyzed the mice activity in the task for the continuous and discontinuous groups (**Figure 3B**, **Figure S6A**). We could have also measured the corticosterone or cortisol level since it there are good indicators for the stress level (Gong et al., 2015), but we did not want

to overstress the animal with a blood sampling procedure (Harikrishnan et al., 2018). For better recovery of the animals and to induce the minimal bias possible on our data, we began the behavioural tests procedures by the least stressful test (open field) and the day after we undertook the tail suspension test. Besides, we made sure that the mice had enough time to rest (at least 8 days) after the tail suspension test before implementing the cognitive behavioural task as it has been shown that the tail suspension test can lead to decreased locomotion, exploration and increased anxiety level (Kale et al., 2013). We also habituated the mice to the novel food when they were still house-grouped in their home cage to reduce hyponeophagia (Deacon, 2011; Morozov, 2018). The mouse in the continuous condition were not voluntary food-deprived by the experimenter (no prior isolation to give 2 g of food per day). Even if we saw a decrease in the weight of these mice as observed for the mouse in the discontinuous protocol that are food deprived (**Figure 3A**), they rarely went under 85% of their initial weight. They worked in the task accordingly to their metabolic need (**Figure 3B, Table S1**) and self-pace their access to food when they needed contrary to a food-restriction protocol that might trigger stress and a reduction of activity in the operant chamber (Guarnieri et al., 2012; Levay et al., 2010). Another explanation for the reduction of activity, could be related to the fact that the mice in the discontinuous condition received a minimal amount of food to complement the pellets they got performing the task: so why working if they would get food anyway in an effortless manner? As for the mice in the continuous condition we can imagine distributing randomly a minimal amount of food (1 g to 2 g of pellet over 24 hours) for the mice that do not get enough pellets during the most difficult stages of the tasks without discouraging them from performing the task while strictly maintaining a weight above 85 %.

Contrary to the AutonoMouse system (Erskine et al., 2019), or the Souris City system (Torquet et al., 2018) one limitation of our protocol might be the isolation of the animals. Social isolation can increase anxiety and more importantly, it may disrupt cognitive performances (Võikar et al., 2005) even if it still discussed depending on the duration of the social isolation (Buckinx et al., 2019). In this line, in our own study, we notably showed that there was no significant difference in the open-field and tail suspension test between the mice in the continuous condition, the group-housed mice and the isolated mice. The mice in the discontinuous protocol however were more anxious (**Figure 3E, Figure 3F, Figure S3**). The overall results might suggest that the repeated handling was more aversive than social isolation. Despite these conclusions, we encourage the experimenters to provide a minimal enrichment in the home cages and the operant chambers for all the mice all along the procedure by adding a cotton pad. We notably noticed that the isolated mice, especially the ones in the discontinuous condition, tended to significantly nest more quickly than the mice in the group condition (**Figure S6B**). The fact that the mice in the group condition nest less might be explained by the group maintaining a higher temperature in the home cages (Gaskill et al., 2013). For isolated mice, providing nesting material can be a proper way to alleviate anxiety as it is a way to enrich a cage (Sherwin and Olsson, 2004). For future experiments, we may think

about adding a little house in the BEATBox or a wheel attached at the back panel or the door. We might also consider training the mice in groups prior to the cognitive tasks as it has been successfully done in recent studies (Francis et al., 2019). We could have also implemented a protocol during which the animals are trained during the nights and put back in their home cage with their littermates during the day. Yet, Poddar and collaborators showed there was no difference in their task between the isolated home-cage trained rats and the socially housed rats (Poddar et al., 2013). Besides, we are not in favor for this kind of procedure as the animals are food deprived and some mice can show a circadian rhythm lag, implying they might work during a part of the day. Such a procedure would imply a default in data collection.

It is to be noticed that the open field and tail suspension test results did not display clear results which might be related to the little number of animals, the validity of these tests as well as the parameters used. For the open field notably, one might say that 40 lux is not aversive enough. However, very few studies precise the illuminating protocols they use in the open field test and based on Martin-Arenas and collaborators' work we set the light at 40 lux (Martin-Arenas and Pintado, 2014).

Last but not least, it is important to underline that these results are conditioned by the strains we are using (C57BL6/J) and the mice gender. The operant chambers are not odours/pheromones hermetic and in this respect, we can't run both male and females at the same time. At the time of the experiments we only had C57BL6/J males but the same protocols should be validated with females and other strains as it has previously been shown that results from behavioural studies vary depending on the mice strain (Turner et al., 2017; Võikar et al., 2005) as well as gender (Harris et al., 2020).

### **The BEATBox enables the implementation of various innovative behavioural tasks.**

We provide here a BEATBox model tailored for visual cognitive tasks (**Figure 1A**, **Figure 4A**, **Figure 4C**, **Figure 4D**). Nevertheless, the customizability of our system is an advantage for scientists that will chose not to implement visual tasks. They could indeed adapt the BEATBox for other sensory modalities such as sounds (Poddar et al., 2013; Saunders and Wehr, 2019) or odours (Erskine et al., 2019). The flexibility of the BEATBox allows an uninterrupted improvement of the scientific questions we can ask thanks to the refinement of the implemented behavioural tasks (**Figure 4A**). The BEATBox system allowed us to develop tasks that would have been impossible to set-up because of their complexity and the number of trials they require to properly analyze the data (**Figure 4C**). For this study, the training and reversal learning task lasted 20 days, but we are currently running experiments lasing 30 days (**Figure 4C**) up to 60 days (data not shown) without any negative consequences on the animal (stable weight, proper nesting, no sign of apathy or dolor, excellent activity and performance in the task). The BEATBox allows the development of new modules to better document some behavioural parameters such as the reaction time (**Figure 4B**) or to record and manipulate the neuronal basis of the observed behaviours (**Figure 4D**). Importantly, the BEATBox can be associated with devices to record neuronal activity in

freely-moving mice which has been proven less stressful than head-fixed methods (Juczewski et al., 2020). Switching to wireless recording or neuromodulation (Hasegawa et al., 2015; Park and Carmel, 2016; Szuts et al., 2011; Wentz et al., 2011) could also embody an excellent improvement for the animal well-being and the quality of the behavioural data in the BEATBox.

## STAR METHODS

### Key Resources Table

REAGENT OR RESOURCE	SOURCE	IDENTIFIER
<b>Experimental Models: Organism/Strain</b>		
Mouse; C57BL6J	Brain and Spine Institute	
<b>Softwares and Algorithms</b>		
Matlab 2017b	MathWorks	<a href="https://fr.mathworks.com/products/matlab.html">https://fr.mathworks.com/products/matlab.html</a>
Arduino 1.8.5	Arduino	<a href="https://www.arduino.cc/">https://www.arduino.cc/</a>
Solidworks	Dassault systèmes	<a href="https://www.solidworks.com/fr">https://www.solidworks.com/fr</a>
Open field	CleverSys	<a href="http://cleversysinc.com/CleverSysInc/csi_products/open-field/">http://cleversysinc.com/CleverSysInc/csi_products/open-field/</a>
BIO-TST	Bioseb	<a href="https://www.bioseb.com/fr/anxiete-depression/2-tail-suspension-test-wireless.html">https://www.bioseb.com/fr/anxiete-depression/2-tail-suspension-test-wireless.html</a>

### Lead Contact

Further information and request for resources should be directed to and will be fulfilled by the Lead Contact, Eric Burguière ([eric.burguiere@icm-institute.org](mailto:eric.burguiere@icm-institute.org)).

### Materials Availability

This study did not generate new unique reagents.

### Data and Code Availability

Data and code supporting main results and the task software are available at Github.

## BEATBox Hardware

Roughly, the BEATBox hardware consists of one operant chamber controlled by a master Arduino, connected through serial communication to a computer running Matlab.

The whole structure of the box is made out of laser-cut Plexiglas ® and consists of a main cubic backbone with 4 adjustable and removable screwable lateral panels and a door (**Figure 1A, Video S1**). We used a 305 x 290 x 330 mm (length x width x height) operant chamber, but all the dimensions can be tailored according to the experimenters needs depending on the behavioural experiments and the animal models. The users can adjust each panel according to the task they want to run and the need to perform optogenetics / recording by laser cutting (GBOS GH960 70 mW, GBOS laser) the appropriate slots and adding the desired modules (**Figure 1A**). Indoor modules can be added such as gates to track the animal position within the cage (**Figure 4A**). Every panels and 3D modules were designed using Solidworks (Dassault Systèmes SE, 2017).

For our experiment, we decided to administrate the mice with a simple perceptual task that was food-rewarded in case of correct answers and light punished in case of wrong answers. We set-up our behavioural cages with different modules that we 3D-printed (Form 3, Formlabs and Creality CR10 Pro V2) and equipped with the appropriate electronic components (**Figure 1B**). To display a large variety of visual stimuli on a millisecond scale, we used two screens (2.8" TFT LCD with Touchscreen Breakout Board w/MicroSD Socket - ILI9341, Adafruit) controlled by two Uno Arduino (Adafruit) thanks to 8-bit wiring. We collected the mouse answers thanks to infrared red beams (IR Break Beam Sensor - 3mm LEDs, Adafruit) that were placed just in front of the screen within a 3D-printed frame allowing a more accurate detection of the mouse answers than capacitive detection. To administrate the white-flash punishment, we used a Neopixel ring (NeoPixel Ring - 24 x 5050 RGB LED with Integrated Drivers, Adafruit). We also used this LED ring as a source of red light to be able to videotape our animals even during the nights thanks to the mini camera (color camera mini-HD 700TVL 1/3" CMOS NTSC 3.6mm MTV Board Lens Mini CCTV Security Video FPV) integrated within the ceiling module. To deliver food rewards during the tasks, we used a feeder module that was adapted from Fed (Nguyen et al., 2016). Briefly, we used a stepper motor (14HM11-0404S) controlled by a Uno Arduino combined to a motor shield (Adafruit Motor/Stepper/Servo Shield for Arduino v2 Kit - v2.3, Adafruit) to trigger the rotation of a rotor to deliver the pellets stored in the silo (**Figure 1B, Feeder module**). One infrared break beam sensor is placed directly above the food-well to detect the mice nosepokes and another just below the rotor to detect if the pellets are correctly distributed. All the electronic components are controlled by a master Arduino (Mega Arduino, Adafruit) that directly communicates with the computer and the Matlab software through serial communication (**Figure 1A, Figure S1A**). This master Arduino manages all the inputs and outputs by reading the state of each input pin or writing on or off the output pins according to the Matlab instructions.

The behavioural apparatus is entirely black and non-transparent so that the mice are not disturbed by the surrounding environment except for the transparent roof. This roof lets the light shine through so that the mice are exposed to the animal facility light-dark cycle as in their homecage. The cage is equipped with a Plexiglas ® watertight tank filled with litter and we provide the mice with a cotton square so that they can build their nest. The mice also have access to *ad libitum* water (**Figure 1C**).

It is possible to connect multiple cages to one computer through USB ports and run different tasks in parallel in the different boxes, which allow multiple users to perform different experiments at the same time (**Figure 1D**).

### **BEATBox software**

We provide a software to run the tasks and to have access to a custom GUI (Graphical User Interface) such as a launcher interface, a monitoring window to live-track the performances of the animals (**Figure 1D, Figure S2**) and a dialog interface where the experimenter can record various information (weight, nest score...) or start different kind of protocols (DREADD, electrophysiology recordings...). The software was written in Matlab R2017b (MathWorks) on a Windows 10 PC but will be released in Python soon. Another part of the code was developed using the Arduino software and the proper libraries (Adafruit GFX, Adafruit TFT LCD, Adafruit Motor Shield V2, Adafruit NeoMatrix, Adafruit NeoPixel, Adafruit ILI9341) associated to the electronic components.

Briefly, after the experimenter sets the parameters of the task through the GUI (**Figure S2**), the task starts running: the communication between Matlab and the Arduino is initiated, and the monitoring graphic interfaces created (**Figure 1D**). Then the code enters a loop during which Matlab communicates continuously with the master Arduino through serial communication in order to collect the state of each input and to switch on / off the different outputs when it is required (**Figure S1A, Figure 1D**). This communication is possible as the address of each input / output is registered during the settings of the task (**Figure S2B**). Matlab has thus a specific decoding / encoding strategy when receiving / sending information from / to the Master Arduino (**Figure S1 A**).

Importantly we aimed at developing a highly flexible code. All the previous parts can be common to any task and the experimenters only have to code the commands required for the task they want to implement (e.g. “if the mouse is nosepoking, switch on the screens”). Different tasks can be run at the same time in different boxes connected to the same computer (**Figure 1D**).

### **Animals**

A total of 47 specific pathogen-free (SPF) wild-type male C57BL/6J mice were born, weaned at 21 days, and raised in groups (6 mice per cage, and one cage with 5) in our own animal facility in individually ventilated cages on a 12 h light-dark cycle (lights on at 08:00 AM and off at 8:00 PM) at 21°C and 40%

humidity. They had unlimited access to water and food and were provided with nesting material. The experiment started when the animals reached the mean age of 4 months. All the procedures were in accordance with European and French laws and conducted in agreement with institutional guidelines.

We set-up 4 conditions per batch for 4 animals coming from the same cage. In each batch, one mouse underwent the continuous protocol, i.e., when it started the behavioural task, it remained constantly in the cage with permanent access to the task until the experimenter stopped the protocol (Continuous). Another mouse underwent a discontinuous protocol as described in the literature. This mouse was isolated and food-restricted to 85% of its initial weight. It performed the task for 1 hour each day during the dark phase (8:30 PM – 9:30 PM), (Discontinuous). We used a socially isolated mouse (Isolation) and a group-housed mouse (Group-housed) that did not perform any task to control for isolation (**Figure 2A**).

During the whole experiment, all the mice had access to the same food (LabTab AIN-76A Rodent 20 mg, Test Diet), to nesting materials and the litter was changed every week or as soon as the mice were isolated or started the tasks. They had unlimited access to water. All along the tests and procedures, the same female experimenter handled the animals and she always picked them up with open hands (Hurst and West, 2010). The weight was monitored every day, the nest score regularly assessed when changing the cage and we had access to the food intake for all the isolated mice. The overall wellness of the animals was assessed by monitoring the skin turgor, and looking at the animal activity, posture, face, fur cleanliness and any sign of distress would have led to the animal exclusion of the study (Spangenberg and Keeling, 2016). None of them reached a limit point.

The experiments were conducted according to a precise schedule (**Figure 2A**). The animals first got used to the experimenter with a daily interaction for seven days. The habituation protocol occurred between 6:30 PM and 8:30 PM, before the beginning of the dark phase but at a moment during which we know from previous experiments in the lab that the mice are active. The habituation included 10 minutes of observation of the mice in their home cage with the lid open for the mice to get used to the experimenter and her scent; then the experimenter put her gloved hand in the cage for the mice to freely explore it; finally, each mouse was handled for 5 minutes using the cup method (Hurst and West, 2010). They were assessed for anxiety and depression-like behaviours before undergoing the reversal task (or the control housing) after getting used to the food. After the task, the mice were assessed once again for anxiety and depression-like behaviours and sacrificed. All the experiments occurred during the dark phase, around 8:30 PM.

### **Open-field test**

Anxiety-like behaviour was assessed with the open-field (Crawley, 2007; Seibenhener and Wooten, 2015). The apparatus consisted of a square-shaped arena (5000x500x500mm, length x width x height), made of white and grey Plexiglas ® and illuminated at 40 lux. Subjects were placed facing the wall of the

top right corner and could explore the apparatus for 15 min. The open field apparatus was wiped with Aniospray after each mouse to remove odours. The open field was subdivided into a central square (200x200 mm) and 4 corners (100x100 mm) (**Figure 3E**). The following parameters were recorded: number of entries in the central zone, time spent in central zone, number of entries in the corners, time spent in the corners, overall travelled distance in the open-field, distance travelled in the central zone. The data were acquired and analysed through the TopScan software (CleverSys). Mice were habituated to the testing room by bringing their home cage into the testing room for a minimum of one hour prior to the test.

### **Tail suspension test**

Depression-like behaviours were assessed with the tail suspension test (Powell et al., 2012). During the 6-minute TST, the mice were suspended by their tails from a metal hook (Bioseb, Model: BIO-TST4) using a climbstopper device (Can et al., 2012). The experiment was conducted with a light of 5 lux. The distance between the floor and the tail was about 10 cm. The data (immobility, energy and power) were automatically acquired and analysed through the Bioseb software. Mice were habituated to the testing room by bringing their home cage into the testing room for a minimum of one hour prior to the test.

### **Task**

The mice in the discontinuous condition were habituated to the testing room by bringing their home cage into the room for a minimum of one hour prior to the test. The mice in the continuous condition were habituated for one hour to the testing room on day 1 and then remained in the BEATBox.

### **Reversal learning training**

The pre-training phases were necessary for the mice to get used to the behavioural set-up. They were divided into 4 steps (**Figure 2B**, **Figure S3**). First, the mice had to learn where the pellets were delivered. To achieve that, a pellet was automatically distributed in the food well every minute for 30 times (**Figure S3A**). Then the mice had to nosepoke 30 times in the food well to get a pellet (**Figure S3B**). After that, the mice had to learn how to give an answer. When exploring the cage, if they crossed the infrared break beams before one of the two screens, the screens started blinking and a pellet was automatically delivered. The mice had 15 s to retrieve the reward otherwise the trial was considered as wrong and the countdown was set back to 10 trials to complete anew (**Figure S3C**). Finally, the animals had to understand the global structure of the task. When they nosepoked in the food well, both screens turned with a big white square on and the animal had one minute to give an answer by breaking one infrared beam and they had 15 s to retrieve the reward. The mice had to complete 10 successive complete trials within the imposed delays otherwise the countdown was set back to 10 trials to complete anew (**Figure**

**S3D).** For the mice in the discontinuous task, we moved them from Stage 4 to the reversal task if they had more than 10 rewards during 3 consecutive sessions.

### Reversal learning task

The mice were exposed to a deterministic stimulus-based reversal task (**Figure 2B**, **Figure S4**, **Video S2**). After starting the trial by nosepoking in the food-well, the mice were exposed at the same time to two visual stimuli with equivalent brightness (grid and vertical bars) spatially pseudorandomized and they had to pick the rewarded stimulus. If they answered correctly, the screens started blinking and they had 15 s to nosepoke in the food-well to get a pellet. If they gave the wrong answer, a 5 s bright flash was turned on and the following trials were corrective trials, meaning that the locations of the stimuli were repeated until the mouse gave the correct answer as previously described (Mar et al., 2013). When the mice reached a performance of 80 % over the past 40 trials, the contingencies were reversed, and the previous unrewarded stimulus become the new rewarded one (**Figure 2B**). The initial correct stimulus (grid or vertical bars) was balanced across the pairs of mice.

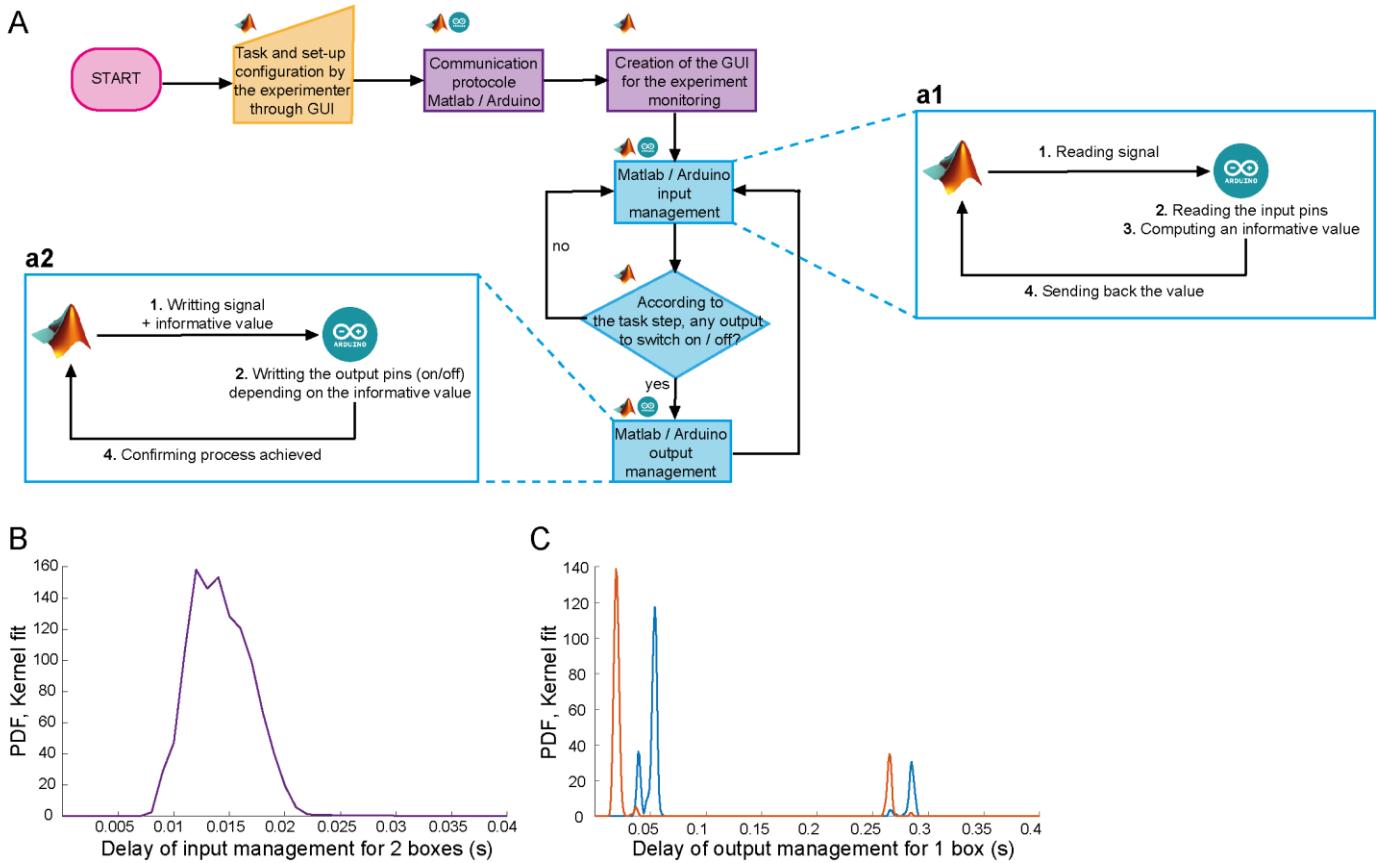
### Analysis and statistics

All the analysis, statistics and figures of this article were produced with custom-made Matlab codes (R2017b). Due to the small number of animals, we used non-parametric tests. All the analysis codes are provided with extended comments. For the OF, TST, weight, nest building, number of trials, number of completed trials reversals (trials for which the mice gave an answer), we gathered the values of each variables of interest recorded in the experiments for each animal. The results are plotted as the mean  $\pm$  SEM. For the curves fitting, the Matlab polyfit function (degree 1) was used and we calculated the correlation coefficient and the p-value using the corrcoef function (R was then squared and the p-values test the null hypothesis that there is no relationship between the variables). The appropriate tests were used depending on the samples at stake (Wilcoxon sign-rank test for paired samples, Wilcoxon rank-sum test for independent variables, Kruskal-Wallis test followed by a multiple comparison test for multiple data sets). For the performance curves, the correctness of each answer (1 for correct, 0 for wrong) was collected and the overall vector was smoothed with the Matlab smoothdata function, using a moving average of 40 trials (**Figure 2D**) or 50 trials (**Figure 4C**).

The actograms are based on the calculation of an intertrial interval ITI threshold for each mouse above which two trials are considered as belonging to two different bursts of activity (red lines). Briefly for each trial n ( $n > 2$ ), we calculate the following ratio R and studied its distribution. The first value superior to 95<sup>th</sup> quantile of the distribution is considered as the threshold value.

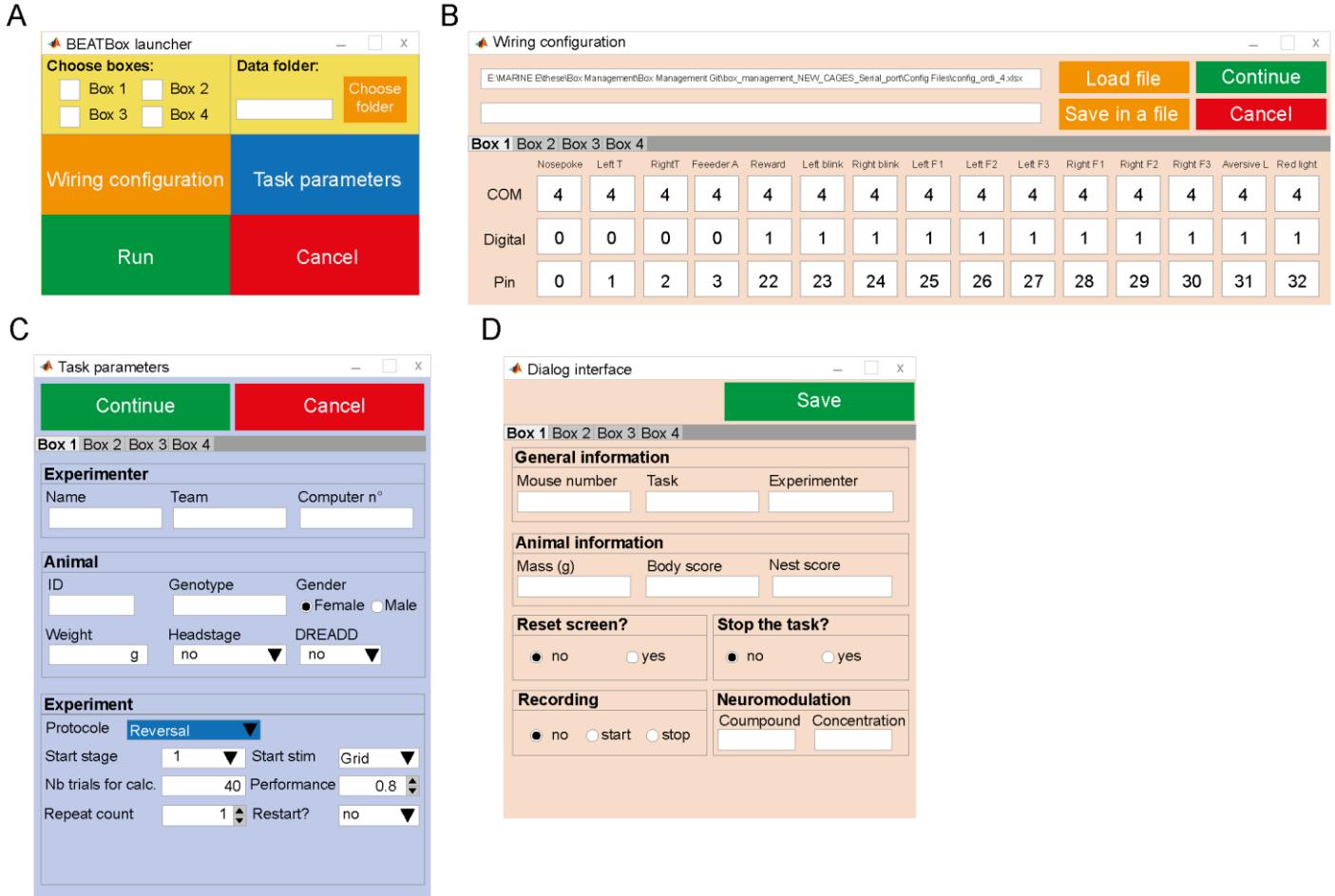
$$R(n) = ITI(n) + \log\left(\frac{ITI(n-2) + ITI(n-1)}{2}\right)$$

## SUPPLEMENTARY FIGURES AND LEGENDS



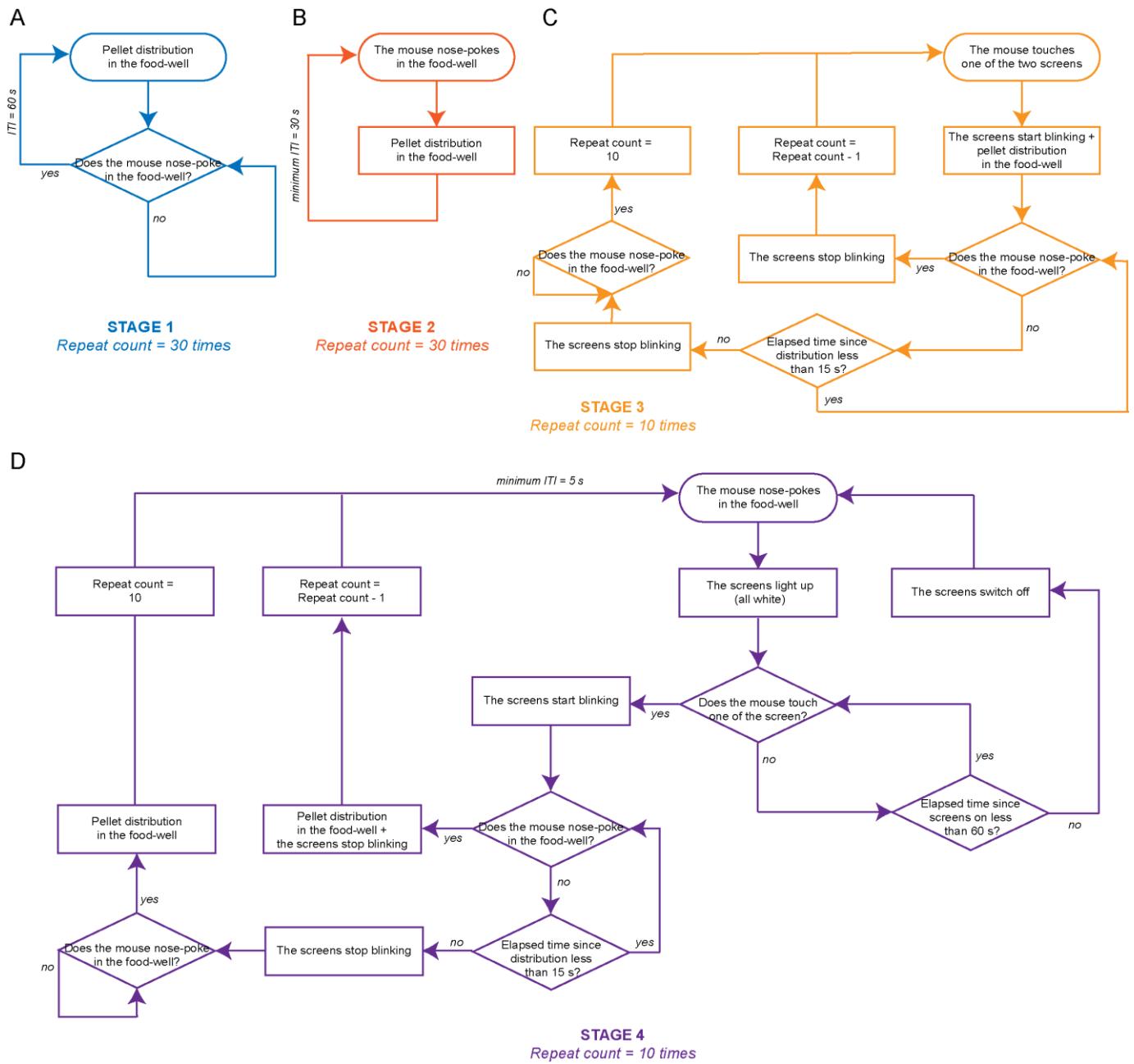
**Figure S1 BEATBox temporal resolution.**

- (A) Overview of the software management of the BEATBox with the detailed processes of communication between Matlab and the master Arduino.
- (B) Delay for the input state detection for 2 boxes connected to the computer (process pictured in a1, panel a).
- (C) Delay for the output activation/deactivation in one box of interest (process pictured in a2, panel a). Each colour represents one box.



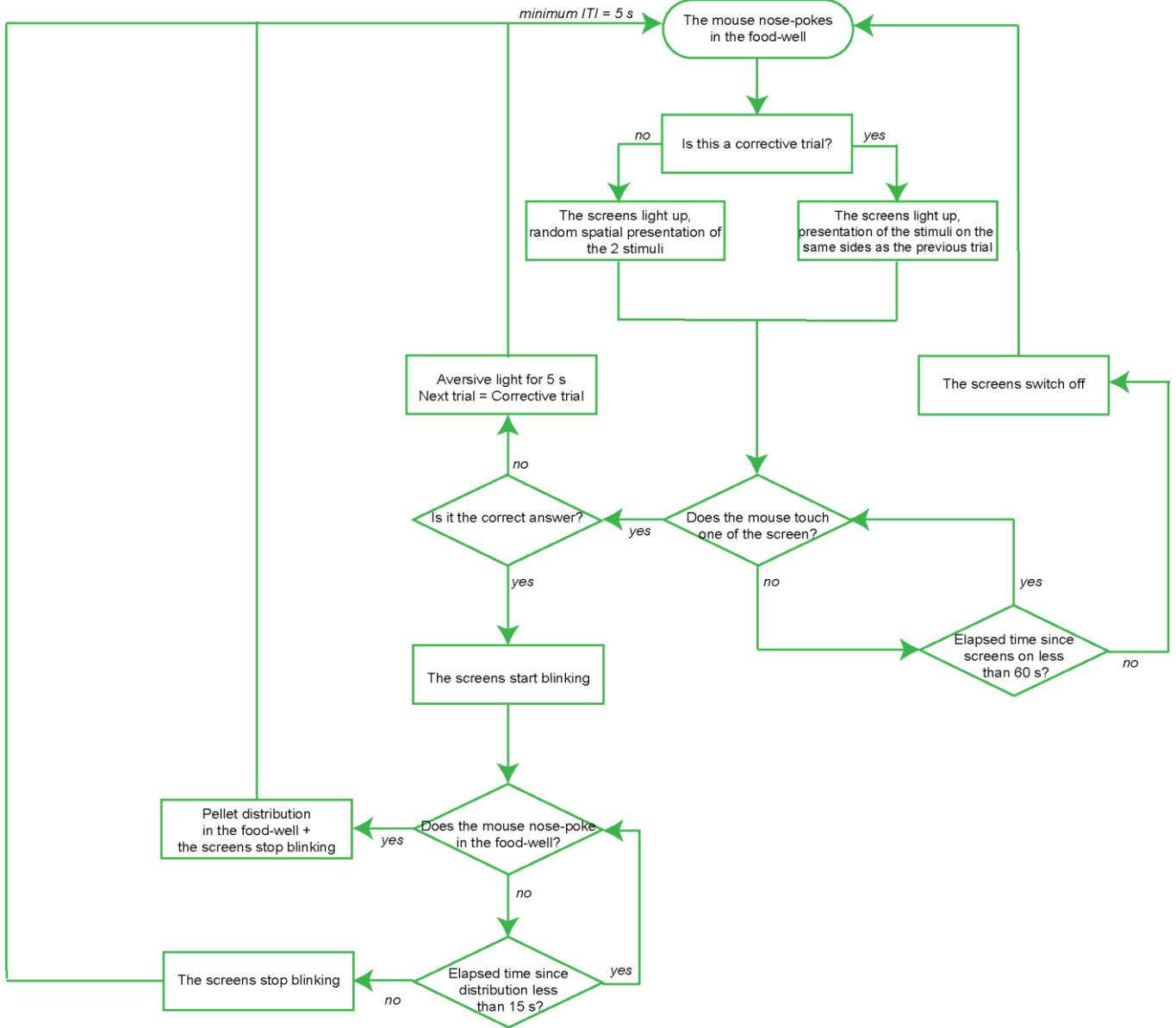
**Figure S2 Overview on the interfaces provided with the BEATBox software.**

- (A) BEATBox launcher to navigate between the different settings to update before running the task.
- (B) Interface to load the address of each input/output to ensure a proper communication between Matlab and the master Arduino.
- (C) Interface to select the different parameters of the task and to fill in various experimental information.
- (D) Dialog interface allowing the savings of different information during the task and the timestamp of the neuromodulation/recording protocols.



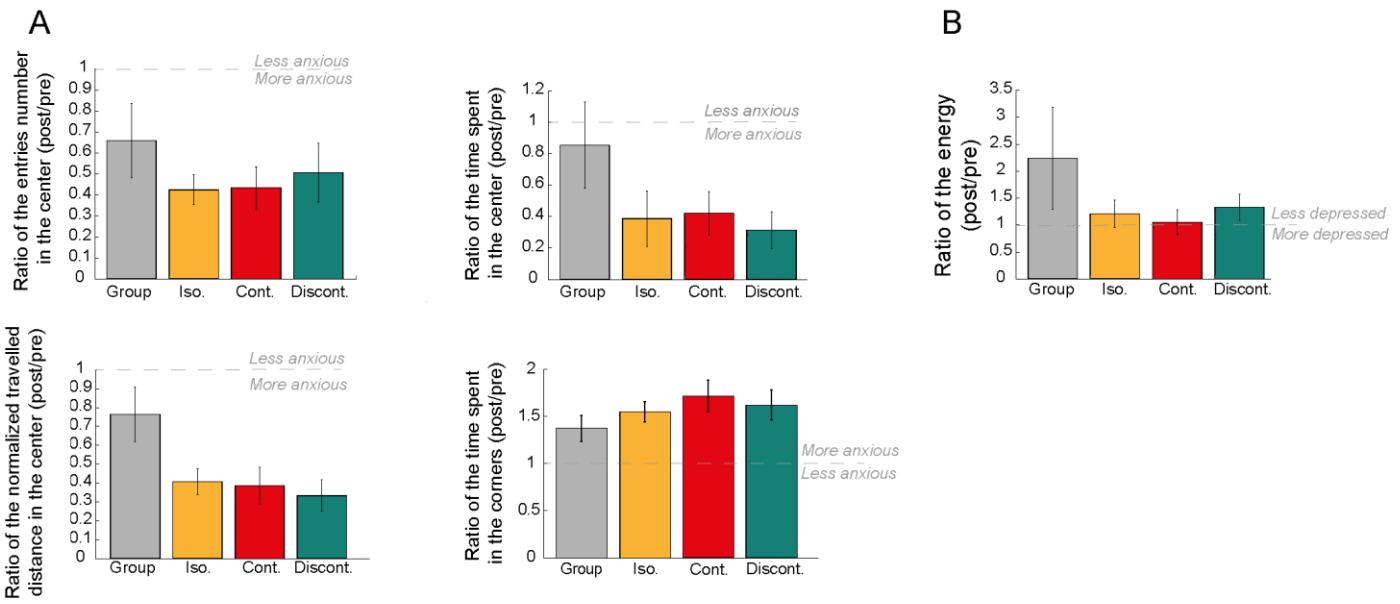
**Figure S3 Flow chart of the training stages.**

Four training stages were required to shape the mouse behaviour towards the final stage. These stages helped the mice to understand how the operant chamber works from the pellet delivery system (Stages 1 and 2) to the answers on the screen (Stage 3) and the way to start and complete a trial (Stage 4). In the most difficult stages (3 and 4) the mice had to complete 10 successive trials (proper actions and timings) in a row to go to the next stages.



**Figure S4 Flow chart of the reversal learning task.**

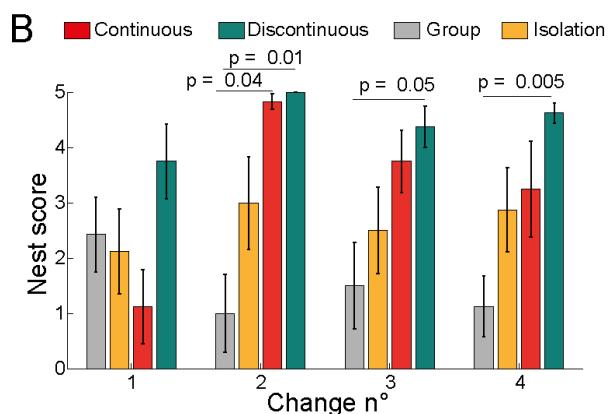
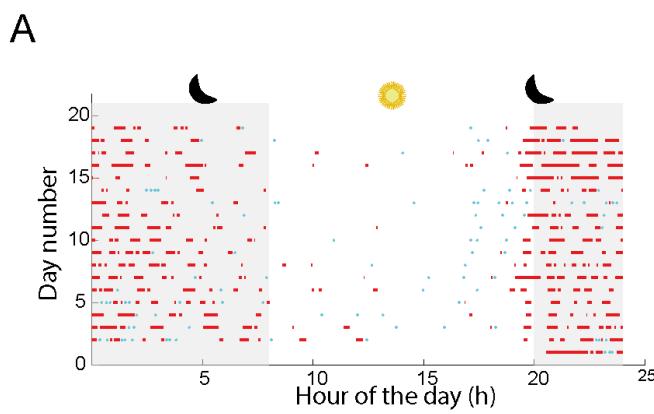
During the reversal learning part, the mice had to start a trial by nosepoking in the food-well and pick the side with the rewarded stimulus before 60 seconds. If they gave the correct answer, the screen started to blink, and they had 15 seconds to go to the food-well to get their pellet. If they gave an incorrect answer, a flashlight switched on for 5 seconds. In the latest case, during the next trials the stimuli will be presented on the same sides as the previous trials, otherwise their right/left localisation is random.



**Figure S5 Complementary results of the open field and tail suspension test.**

(A) Evaluation of the mice anxiety with the open field test. The travelled distance in the center was normalized by the overall distance travelled. For the center data, a ratio inferior to 1 indicates that the mice tend to avoid more the center after the protocol (post data, acquired on day 39) than before (pre data, acquired on day 8) and vice-versa for the corners data. ( $n = 8$  per condition, data are shown as mean  $\pm$  SEM, Kruskal-Wallis test, Multiple comparison test if  $p \leq 0.05$  in the Kruskal-Wallis test).

(B) Evaluation of the mice depression-like state with the tail suspension test. A ratio superior to 1 indicates that the mice struggled more after the protocol (post data, acquired on day 40) than before (pre data, acquired on day 9). ( $n = 8$  per condition, data are shown as mean  $\pm$  SEM, Kruskal-Wallis test).



**Figure S6 Supplementary assessments of the mice well-being.**

(A) Example of the activity of one mouse in the continuous condition, along the training and the reversal learning task. Each red line indicates a burst of activity (high number of consecutive trials), the light blue dots indicate isolated trials.

(B) After each litter change, a new cotton pad was given to the mice and the nest score was assessed 24h afterwards (Deacon, 2006) ( $n = 8$  per condition, data are shown as mean  $\pm$  SEM, Kruskal-Wallis test, Multiple comparison test if  $p \leq 0.05$  in the Kruskal-Wallis test).

## Video S1 Flexibility and customizability of the BEATBox

Each panel and modules are screwable and the BEATBox configuration can be adjusted to the user needs.

## Video S2 Mouse performing the stimulus-based reversal task

When the mouse nosepokes in the food-well, the screens switch on with the 2 visual stimuli. If the mouse answers correctly, the screens start blinking and the mouse has 15 seconds to get its reward. If the answer is incorrect, there is a strong flashlight. The mouse is being videotaped thanks to the camera from the ceiling and the red light from the Neopixel.

### Continuous condition

Mouse n°	R <sup>2</sup>	p	Significant relationship	Slope
797	<b>0,31</b>	<b>0,02</b>	*	<b>-623,47</b>
789	0,10	0,19		-535,95
896	0,12	0,18		-306,20
871	0,08	0,27		-236,43
988	<b>0,24</b>	<b>0,04</b>	*	<b>-816,57</b>
939	0,15	0,12		796,93
996	0,00	0,88		-19,68
24	<b>0,51</b>	<b>8,32E-04</b>	*	<b>-1273,38</b>

### Discontinuous condition

Mouse n°	R <sup>2</sup>	p	Significant relationship	Slope
802	<b>0,45</b>	<b>0,002</b>	*	<b>293,7575</b>
803	0,01	0,753		-18,31844
897	0,16	0,125		224,95211
872	0,00	0,851		9,5056912
989	0,09	0,227		-234,1853
957	0,04	0,400		157,34783
999	<b>0,72</b>	<b>9,03E-06</b>	*	<b>436,78988</b>
26	0,23	0,046		573,12828

**Table S1**

Results (correlation coefficient, p-value and slope) for the individual linear fits of the number of trials against the relative mass of the mice.

### Continuous condition

Mouse n°	R <sup>2</sup>	p	Significant relationship	Slope
797	<b>0,48</b>	<b>0,0015</b>	*	<b>9,30</b>
789	<b>0,35</b>	<b>0,0103</b>	*	<b>8,06</b>
896	<b>0,43</b>	<b>0,0032</b>	*	<b>7,12</b>
871	0,01	0,6901		-0,87
988	<b>0,69</b>	<b>1,9586E-05</b>	*	<b>13,15</b>
939	0,16	0,1007		4,82
996	0,16	0,0959		3,53
24	<b>0,63</b>	<b>0,0001</b>	*	<b>10,27</b>

### Discontinuous condition

Mouse n°	R <sup>2</sup>	p	Significant relationship	Slope
802	0,22	0,051		-0,74
803	0,02	0,606		-0,17
897	0,01	0,722		0,13
872	<b>0,24</b>	<b>0,048</b>	*	<b>0,46</b>
989	<b>0,37</b>	<b>0,008</b>	*	<b>1,42</b>
957	0,07	0,280		0,99
999	<b>0,53</b>	<b>0,001</b>	*	<b>-1,90</b>
26	0,01	0,776		0,18

**Table S2**

Results (correlation coefficient, p-value and slope) for the individual linear fits of the number of trials against the day number in the task.

## AUTHOR CONTRIBUTIONS

Conceptualization, E.B, N.B and M.E; Methodology, E.B, N.B, M.E; Hardware, M.E, P.T, Y.T, M.B on an initial basis developed by N.B for MedAssociate. Software, M.E and Y.J.T on an initial basis developed by G.P for MedAssociate. Resources, M.E; Investigation, M.E; Formal analysis, M.E; Data curation, M.E; Writing – original draft preparation, M.E; Writing – review and editing, E.B and N.B. Website management: S.L. MG, Funding acquisition, E.B.

## DECLARATION OF INTEREST

The authors declare no competing interests.

## ACKNOWLEDGMENTS

All animal work was conducted at the PHENO-ICMice facility. The Core is supported by 2 “Investissements d’avenir” (ANR-10- IAIHU-06 and ANR-11-INBS-0011-NeurATRIS) and the “Fondation pour la Recherche Médicale”.

The research leading to these results has received funding from the program “Investissements d’avenir” ANR-10- IAIHU-06.

We would like to thank Clément BOTREL for the BEATBox logo. You can check his website to see more of his work:

[https://clementbotrel.wordpress.com/?fbclid=IwAR1TANRnkRy2EXb5N8l9WrZMT3NZr06t1WohrxchDblxuNvUmY4XUdSz\\_IU](https://clementbotrel.wordpress.com/?fbclid=IwAR1TANRnkRy2EXb5N8l9WrZMT3NZr06t1WohrxchDblxuNvUmY4XUdSz_IU)

## REFERENCES

- Amaral-Júnior, P.A., Mourão, F.A.G., Amancio, M.C.L., Pinto, H.P.P., Carvalho, V.R., Guarnieri, L. de O., Magalhães, H.A., and Moraes, M.F.D. (2019). A Custom Microcontrolled and Wireless-Operated Chamber for Auditory Fear Conditioning. *Front. Neurosci.* *13*, 1193.
- Ashwood, Z.C., Roy, N.A., Stone, I.R., Churchland, A.K., Pouget, A., and Pillow, J.W. Mice alternate between discrete strategies during perceptual decision-making. *45*.
- Balcombe, J.P., Barnard, N.D., and Sandusky, C. (2004). Laboratory routines cause animal stress. *Contemp. Top. Lab. Anim. Sci.* *43*, 42–51.
- Balzani, E., Falappa, M., Balci, F., and Tucci, V. (2018). An approach to monitoring home-cage behavior in mice that facilitates data sharing. *Nat. Protoc.* *13*, 1331–1347.
- Benzina, N., N'Diaye, K., Pelissolo, A., Mallet, L., and Burguière, E. (2021). A cross-species assessment of behavioral flexibility in compulsive disorders. *Commun. Biol.* *4*, 1–12.
- Buckinx, A., Van Schuerbeek, A., Van Den Herrewegen, Y., Smolders, I., and Dimitri, D.B. (2019). The effect of housing conditions on working memory and anxiety in male C57BL/6 mice.
- Bussey, T.J., Padain, T.L., Skillings, E.A., Winters, B.D., Morton, A.J., and Saksida, L.M. (2008). The touchscreen cognitive testing method for rodents: how to get the best out of your rat. *Learn. Mem. Cold Spring Harb. N* *15*, 516–523.
- Can, A., Dao, D.T., Terrillion, C.E., Piantadosi, S.C., Bhat, S., and Gould, T.D. (2012). The tail suspension test. *J. Vis. Exp. JoVE* e3769.
- Crawley, J.N. (2007). Social Behavior Tests for Mice. *8*.
- D'Ausilio, A. (2012). Arduino: A low-cost multipurpose lab equipment. *Behav. Res. Methods* *44*, 305–313.
- Deacon, R.M.J. (2006). Assessing nest building in mice. *Nat. Protoc.* *1*, 1117–1119.
- Deacon, R.M.J. (2011). Hyponeophagia: A Measure of Anxiety in the Mouse. *J. Vis. Exp. JoVE*.
- Devarakonda, K., Nguyen, K.P., and Kravitz, A.V. (2016). ROBucket: A low cost operant chamber based on the Arduino microcontroller. *Behav. Res. Methods* *48*, 503–509.
- Erskine, A., Bus, T., Herb, J.T., and Schaefer, A.T. (2019). AutonoMouse: High throughput operant conditioning reveals progressive impairment with graded olfactory bulb lesions. *PLOS ONE* *14*, e0211571.
- Fonio, E., Golani, I., and Benjamini, Y. (2012). Measuring behavior of animal models: faults and remedies. *Nat. Methods* *9*, 1167–1170.
- Francis, N.A., Bohlke, K., and Kanold, P.O. (2019). Automated Behavioral Experiments in Mice Reveal Periodic Cycles of Task Engagement within Circadian Rhythms. *ENeuro* *6*.
- Gaskill, B.N., Karas, A.Z., Garner, J.P., and Pritchett-Corning, K.R. (2013). Nest Building as an Indicator of Health and Welfare in Laboratory Mice. *J. Vis. Exp. JoVE*.
- Godynuk, E., Bluitt, M.N., Tooley, J.R., Kravitz, A.V., and Creed, M.C. (2019). An Open-Source, Automated Home-Cage Sipper Device for Monitoring Liquid Ingestive Behavior in Rodents. *ENeuro* *6*.
- Gomez-Marin, A., and Mainen, Z.F. (2016). Expanding perspectives on cognition in humans, animals, and machines. *Curr. Opin. Neurobiol.* *37*, 85–91.
- Gong, S., Miao, Y.-L., Jiao, G.-Z., Sun, M.-J., Li, H., Lin, J., Luo, M.-J., and Tan, J.-H. (2015). Dynamics and correlation of serum cortisol and corticosterone under different physiological or stressful conditions in mice. *PloS One* *10*, e0117503.
- Gouveia, K., and Hurst, J.L. (2017). Optimising reliability of mouse performance in behavioural testing: the major role of non-aversive handling. *Sci. Rep.* *7*, 44999.

Guarnieri, D.J., Brayton, C.E., Richards, S.M., Maldonado-Aviles, J., Trinko, J.R., Nelson, J., Taylor, J.R., Gourley, S.L., and DiLeone, R.J. (2012). Gene Profiling Reveals a Role for Stress Hormones in the Molecular and Behavioral Response to Food Restriction. *Biol. Psychiatry* *71*, 358–365.

Gurley, K. (2019). Two open source designs for a low-cost operant chamber using Raspberry Pi™. *J. Exp. Anal. Behav.* *111*, 508–518.

Harikrishnan, V.S., Hansen, A.K., Abelson, K.S., and Sørensen, D.B. (2018). A comparison of various methods of blood sampling in mice and rats: Effects on animal welfare. *Lab. Anim.* *52*, 253–264.

Harris, C., Aguirre, C., Kolli, S., Das, K., Izquierdo, A., and Soltani, A. (2020). Unique features of stimulus-based probabilistic reversal learning. *BioRxiv* 2020.09.24.310771.

Hasegawa, T., Fujimoto, H., Tashiro, K., Nonomura, M., Tsuchiya, A., and Watanabe, D. (2015). A wireless neural recording system with a precision motorized microdrive for freely behaving animals. *Sci. Rep.* *5*, 1–8.

Horner, A.E., Heath, C.J., Hvoslef-Eide, M., Kent, B.A., Kim, C.H., Nilsson, S.R.O., Alsiö, J., Oomen, C.A., Holmes, A., Saksida, L.M., et al. (2013). The touchscreen operant platform for testing learning and memory in rats and mice. *Nat. Protoc.* *8*, 1961–1984.

Hurst, J.L., and West, R.S. (2010). Taming anxiety in laboratory mice. *Nat. Methods* *7*, 825–826.

Hvoslef-Eide, M., Nilsson, S.R.O., Saksida, L.M., and Bussey, T.J. (2016). Cognitive Translation Using the Rodent Touchscreen Testing Approach. *Curr. Top. Behav. Neurosci.* *28*, 423–447.

Izquierdo, A., Brigman, J.L., Radke, A.K., Rudebeck, P.H., and Holmes, A. (2017). The neural basis of reversal learning: An updated perspective. *Neuroscience* *345*, 12–26.

Jirkof, P. (2014). Burrowing and nest building behavior as indicators of well-being in mice. *J. Neurosci. Methods* *234*, 139–146.

Juczewski, K., Koussa, J.A., Kesner, A.J., Lee, J.O., and Lovinger, D.M. (2020). Stress and behavioral correlates in the head-fixed method. *BioRxiv* 2020.02.24.963371.

Kale, P.P., Addepalli, V., and Ghadawale, S.R. (2013). Impact of pre-exposure of tail suspension on behavioural parameters like locomotion, exploration, and anxiety in mice. *Indian J. Exp. Biol.* *51*, 732–738.

Levay, E.A., Tammer, A.H., Penman, J., Kent, S., and Paolini, A.G. (2010). Calorie restriction at increasing levels leads to augmented concentrations of corticosterone and decreasing concentrations of testosterone in rats. *Nutr. Res.* *30*, 366–373.

Mainen, Z.F., Häusser, M., and Pouget, A. (2016). A better way to crack the brain. *Nature* *539*, 159–161.

Mar, A.C., Horner, A.E., Nilsson, S.R.O., Alsiö, J., Kent, B.A., Kim, C.H., Holmes, A., Saksida, L.M., and Bussey, T.J. (2013). The touchscreen operant platform for assessing executive function in rats and mice. *Nat. Protoc.* *8*, 1985–2005.

Marquardt, K., Sigdel, R., and Brigman, J.L. (2017). Touch-screen visual reversal learning is mediated by value encoding and signal propagation in the orbitofrontal cortex. *Neurobiol. Learn. Mem.* *139*, 179–188.

Martin-Arenas, F.J., and Pintado, C.O. (2014). Results of the Open Field Test at different light intensities in C57 mice. *5*.

Matikainen-Ankney, B.A., Garmendia-Cedillos, M., Ali, M., Kryniotsky, J., Salem, G., Miyazaki, N.L., Pohida, T., and Kravitz, A.V. (2019). Rodent Activity Detector (RAD), an Open Source Device for Measuring Activity in Rodent Home Cages. *ENeuro* *6*.

Morozov, A. (2018). Behavioral modulation by social experiences in rodent models. *Curr. Protoc. Neurosci.* *84*, e50.

Neely, C., Lane, C., Torres, J., and Flinn, J. (2018). The Effect of Gentle Handling on Depressive-Like Behavior in Adult Male Mice: Considerations for Human and Rodent Interactions in the Laboratory.

Nguyen, K.P., O'Neal, T.J., Bolonduro, O.A., White, E., and Kravitz, A.V. (2016). Feeding Experimentation Device (FED): A flexible open-source device for measuring feeding behavior. *J. Neurosci. Methods* *267*, 108–114.

Nithianantharajah, J., McKechnie, A.G., Stewart, T.J., Johnstone, M., Blackwood, D.H., St Clair, D., Grant, S.G.N., Bussey, T.J., and Saksida, L.M. (2015). Bridging the translational divide: identical cognitive touchscreen testing in mice and humans carrying mutations in a disease-relevant homologous gene. *Sci. Rep.* *5*, 14613.

Niv, Y. (2020). The primacy of behavioral research for understanding the brain.

O'Leary, J.D., O'Leary, O.F., Cryan, J.F., and Nolan, Y.M. (2018). A low-cost touchscreen operant chamber using a Raspberry Pi<sup>TM</sup>. *Behav. Res. Methods* *50*, 2523–2530.

Park, H.G., and Carmel, J.B. (2016). Selective Manipulation of Neural Circuits. *Neurother. J. Am. Soc. Exp. Neurother.* *13*, 311–324.

Perkel, J.M. (2015). Programming: Pick up Python. *Nature* *518*, 125–126.

Pineño, O. (2014). ArduiPod Box: A low-cost and open-source Skinner box using an iPod Touch and an Arduino microcontroller. *Behav. Res. Methods* *46*, 196–205.

Poddar, R., Kawai, R., and Ölveczky, B.P. (2013). A fully automated high-throughput training system for rodents. *PLoS One* *8*, e83171.

Powell, T.R., Fernandes, C., and Schalkwyk, L.C. (2012). Depression-Related Behavioral Tests. *Curr. Protoc. Mouse Biol.* *2*, 119–127.

Remmelman, E., Smit, A.B., Verhage, M., and Loos, M. (2016). Measuring discrimination- and reversal learning in mouse models within 4 days and without prior food deprivation. *Learn. Mem.* *23*, 660–667.

Richmond, J. (2000). The 3Rs - Past, Present and Future. *Scand. J. Lab. Anim. Sci.* *27*, 84–92.

Saunders, J.L., and Wehr, M. (2019). Autopilot: Automating behavioral experiments with lots of Raspberry Pis. *BioRxiv* 807693.

Schaefer, A.T., and Claridge-Chang, A. (2012). The surveillance state of behavioral automation. *Curr. Opin. Neurobiol.* *22*, 170–176.

Schreiweis, C., Feltgen, Q., Euvrard, M., Daunizeau, J., and Burguière, E. (2019). Motivational control of decision making in wild-type mice and a mouse model for compulsive-like behaviours.

Seibenhener, M.L., and Wooten, M.C. (2015). Use of the Open Field Maze to Measure Locomotor and Anxiety-like Behavior in Mice. *J. Vis. Exp. JoVE*.

Sherwin, C.M., and Olsson, I.A.S. (2004). Housing conditions affect self-administration of anxiolytic by laboratory mice. *Anim. Welf.* *13*, 33–38.

Siegler, J.H., López, A.C., Patel, Y.A., Abramov, K., Ohayon, S., and Voigts, J. (2017). Open Ephys: an open-source, plugin-based platform for multichannel electrophysiology. *J. Neural Eng.* *14*, 045003.

Sorge, R.E., Martin, L.J., Isbester, K.A., Sotocinal, S.G., Rosen, S., Tuttle, A.H., Wieskopf, J.S., Acland, E.L., Dokova, A., Kadoura, B., et al. (2014). Olfactory exposure to males, including men, causes stress and related analgesia in rodents. *Nat. Methods* *11*, 629–632.

Spangenberg, E.M.F., and Keeling, L.J. (2016). Assessing the welfare of laboratory mice in their home environment using animal-based measures--a benchmarking tool. *Lab. Anim.* *50*, 30–38.

Steinmetz, N.A., Koch, C., Harris, K.D., and Carandini, M. (2018). Challenges and opportunities for large-scale electrophysiology with Neuropixels probes. *Curr. Opin. Neurobiol.* *50*, 92–100.

Steurer, M.M., Aust, U., and Huber, L. (2012). The Vienna comparative cognition technology (VCCT): an innovative operant conditioning system for various species and experimental procedures. *Behav. Res. Methods* *44*, 909–918.

Szuts, T.A., Fadeev, V., Kachiguine, S., Sher, A., Grivich, M.V., Agrochão, M., Hottowy, P., Dabrowski, W., Lubenov, E.V., Siapas, A.G., et al. (2011). A wireless multi-channel neural amplifier for freely moving animals. *Nat. Neurosci.* *14*, 263–269.

Torquet, N., Marti, F., Campart, C., Tolu, S., Nguyen, C., Oberto, V., Benallaoua, M., Naudé, J., Didienne, S., Debray, N., et al. (2018). Social interactions impact on the dopaminergic system and drive individuality. *Nat. Commun.* 9, 3081.

Tucci, V., Hardy, A., and Nolan, P.M. (2006). A comparison of physiological and behavioural parameters in C57BL/6J mice undergoing food or water restriction regimes. *Behav. Brain Res.* 173, 22–29.

Turner, K.M., Simpson, C.G., and Burne, T.H.J. (2017). BALB/c Mice Can Learn Touchscreen Visual Discrimination and Reversal Tasks Faster than C57BL/6 Mice. *Front. Behav. Neurosci.* 11.

Vissers, K., De Jongh, R., Hoffmann, V., Heylen, R., Crul, B., and Meert, T. (2003). Internal and external factors affecting the development of neuropathic pain in rodents. Is it all about pain? *Pain Pract. Off. J. World Inst. Pain* 3, 326–342.

Võikar, V., Polus, A., Vasar, E., and Rauvala, H. (2005). Long-term individual housing in C57BL/6J and DBA/2 mice: assessment of behavioral consequences. *Genes Brain Behav.* 4, 240–252.

Wentz, C.T., Bernstein, J.G., Monahan, P., Guerra, A., Rodriguez, A., and Boyden, E.S. (2011). A wirelessly powered and controlled device for optical neural control of freely-behaving animals. *J. Neural Eng.* 8, 046021.

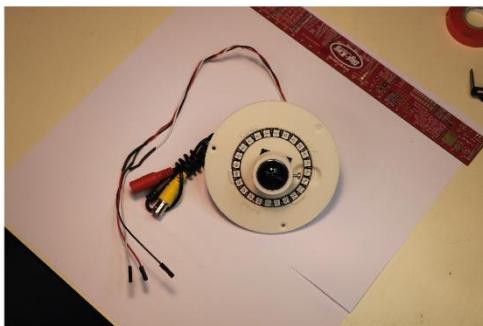
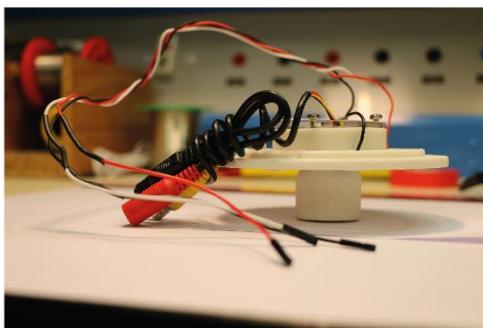
Whissell, P.D., Tohyama, S., and Martin, L.J. (2016). <sup>2</sup>. *Front. Genet.* 7, 70.

White, S.R., Amarante, L.M., Kravitz, A.V., and Laubach, M. (2019). The Future Is Open: Open-Source Tools for Behavioral Neuroscience Research. *ENEURO* 6, ENEURO.0223-19.2019.



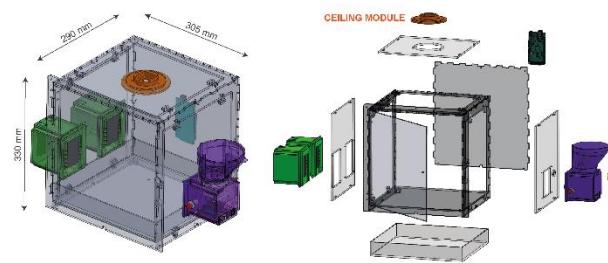


## BEATBOX - HOW TO? CEILING MODULE



Latest update: 05/11/2020

## OVERVIEW



### CEILING MODULE



The ceiling module is a BEATBox module attached to the roof of the operant chamber.

This module includes two main components:

1. A Neopixel ring with 24 LEDs that will be used in the behavioural tasks to display various lights (strong white lights as an aversive stimulus, red lights to enable night videotaping...)
2. A HD camera coupled with a fish-eye lens to videotape the mouse in the BEATBOX

Euvrard et al., 2021

1/14

Euvrard et al., 2021

2/14

## REQUIRED MATERIAL

### Printed parts



1. Ceiling support (~ 1.8 cm height, Ø 11.5 cm)
2. Camera - fish eye lens adapter (~ 2.8 cm height, Ø 3.2 cm)

### Laser cut parts



1. Camera lid (3 mm height, Ø 5.4 cm)

### Electronic components



1. Neopixel Ring 24 LED
2. 3 wire jumpers with at least one male side: Black - White - Red
3. Monofilament wires: Black - White - Red

### Hardware and supplies



1. 3 heat shrink tubings (~ 2 cm length, Ø 3 mm --> Ø 1 mm)
2. 5 M3 x 8 mm screws

### Optical instruments



1. Mini camera HD
2. Fish eye lens

## 3D PRINTING



**Files:** ceiling.stl  
fisheye.stl



**Material:** Resin or filament



**Estimated quantity of material:**

- ceiling.stl, ~ 70 g
- fisheye.stl, ~ 10 g



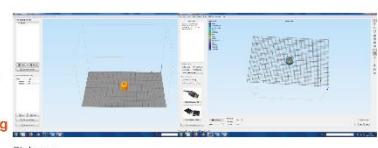
**Estimated time of printing:**

- ceiling.stl, ~ 3h
- fisheye.stl, ~ 1h30

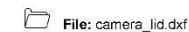


**Parameters:**

With the 3D printer of our FabLab, we printed the files in the following configurations, and the following support structures. **Adjust the printing files as required according to your printer and software.**



## LASER CUTTING



**File:** camera\_lid.dxf



**Material:** PMMA transparent Plexiglas ® 3 mm



**Estimated quantity of material:**

- 70 mm x 70 mm x 3 mm (L x I x W) of Plexiglas ® panel



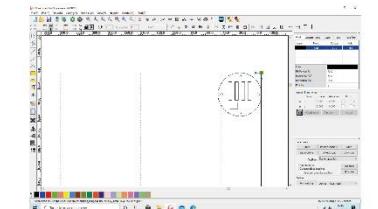
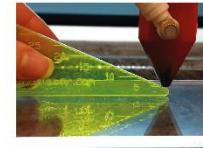
**Estimated time of cutting:** 40 s



**Parameters :**

On the laser cutter machine of our FabLab, we used the following parameters:

- Laser distance: 5 mm
- Laser power: 70 %
- Laser speed: 18 mm/s



Euvrard et al., 2021

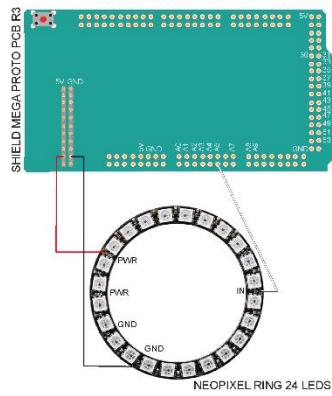
3/14

Euvrard et al., 2021

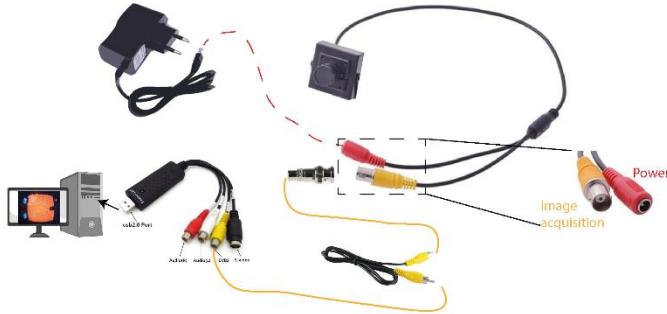
5/14

## CIRCUIT DIAGRAM

The Neopixel ring is plugged to the Arduino Mega for its **ground**, its **power** supply and **control pin** (A5).



The **camera** must be plugged to electrical **power** with a 12 V power adapter. For videos acquisition, the camera must be plugged to the computer. The USB-RCA and the RCA-BNC must be used.



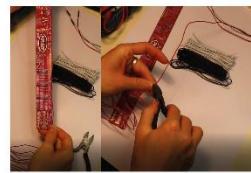
NB: This provides a visual reference for the components wiring (not to scale).

Euvrard et al., 2021

## ASSEMBLY

⌚ 20 min

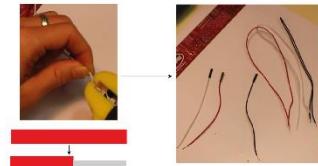
1. Measure 30 cm for each monofilament wire and cut with the wirecutter.



2. Measure 10 cm for each jumper, keep the male side and cut the other side.



3. Strip each wire. (2 sides for the monofilaments, 1 side for the jumpers).



4. Thread a heat shrink tube on each monofilament wire.



## SOFTWARE

- Softwares:**
  - Arduino
  - Matlab
- Associated library:**
  - Adafruit\_NeoPixel (for Arduino)

### ARDUINO

Every coding information for the Neopixel is located in the `Serial_Communication_Arduino_Mega` code which will be loaded in the Arduino Mega. If you want to modify things, don't forget they key elements below:

1. The Neopixel library must be included.



2. Define the Arduino Mega control pin for the Neopixel related to the pin where the Neopixel has been plugged (cf Circuit diagram) and set which LEDs will be dedicated to each color. **Define colors, LEDs number and position as you wish.**

```
#define NEOPIXEL A5
Adafruit_NeoPixel strip = Adafruit_NeoPixel(24, NEOPIXEL, NEO_GRB + NEO_KHZ800);

//Coordinates for the LEDs on Neopixel
int redLED[] = {0, 6, 12, 18};
int whiteLED[] = {1, 2, 4, 5, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23};
int numRedColor = 4;
int numWhite = 16;
```

3. Even if the control pin for the Neopixel is A5 and that there is no other specific wiring, other pins dedicated to specific outputs (red lights, white lights...) must be defined to be switched on (1) and off (0) when the task is running and the Arduino is communicating with Matlab. **Add as many outputs as you wish corresponding to the colors you chose.**

```
#define LIGNE_32
#define PERRON_32
int perronPin[] = {REVERSE, REVERSE, FORWARD, FORWARD, FORWARD, FORWARD, FORWARD, FORWARD, FORWARD, FORWARD, FORWARD};
```

4. In the void set-up, the Neopixel must be initialized as an output and the characteristics of the LEDs are defined. **Define the brightness as you wish.** Set the Neopixel off for the very beginning of the task.

```
void setup()
{
    // Neopixel initialization
    pixel.begin(NEOPIXEL, OUTPORT);
    digitalWrite(NEOPIXEL, LOW);
    strip.setBrightness(1200); //adjust brightness here
    for (int i = 0; i < 24; i++) {
        strip.setPixelColor(i, 0, 0, 0);
    }
    strip.show();
}
```

Euvrard et al., 2021

12/14

# **CHAPITRE 2. Adaptation de la souris en environnement volatil : investigation des stratégies comportementales mises en jeu et des mécanismes computationnels sous-jacents**

---

## **1. RESUME INTRODUCTIF**

### **1.1. ENJEUX DE L'ETUDE**

Toutes les espèces animales dotées d'un système nerveux, du plus simple au plus complexe, sont capables d'apprendre à maximiser des avantages vitaux à et à minimiser des pertes qui pourraient être nuisibles à leur survie. La forme la plus simple d'apprentissage dont est capable toute espèce animale est l'apprentissage par renforcement. Cette forme d'apprentissage consiste à apprendre des tables de valeurs pour chaque association (état - action) afin de déterminer le choix à faire pour maximiser les gains à long terme (Sutton and Barto, 2018). Cependant, l'environnement dans lequel le vivant évolue n'est pas stationnaire et les règles apprises à un moment donné peuvent devenir invalides : il est donc nécessaire de s'adapter à ces changements brusques. Même si l'apprentissage par renforcement permet d'expliquer l'apprentissage non seulement au niveau comportemental, mais aussi au niveau neuronal (Rmus et al., 2021), il présente deux inconvénients majeurs dans le cadre d'un environnement *volatile*. D'une part, il ne permet pas une adaptation rapide car il nécessite dans un premier temps de désapprendre progressivement les règles précédemment apprises puis d'en réapprendre de nouvelles. D'autre part, c'est un processus sans mémoire qui ne permet pas de réutiliser les tables de valeurs apprises dans des situations similaires rencontrées par le passé (Koechlin, 2016). L'apprentissage par renforcement dans sa forme la plus simple ne permet donc pas d'expliquer comment les êtres vivants peuvent apprendre et s'adapter efficacement et rapidement dans des environnements complexes.

Dans la littérature, plusieurs hypothèses existent concernant les différentes stratégies impliquées pour une adaptation efficace. Alors que certains suggèrent que l'adaptation repose sur l'utilisation d'approximations simples et d'heuristiques (Findling et al., 2019; Hermoso-Mendizabal et al., 2020; Wyart and Koechlin, 2016; Yu and Cohen, 2009), d'autres proposent

que l'adaptation repose sur un processus de méta-apprentissage (Behrens et al., 2007; Farashahi et al., 2017; Khamassi et al., 2013; McGuire et al., 2014; Schweighofer and Doya, 2003; Wilson et al., 2014a), quand d'autres encore travaillent sur la capacité à détecter des changements brusques et structurés (Bartolo and Averbeck, 2020; Collins and Frank, 2013; Domenech and Koechlin, 2015; Domenech et al., 2020; Donoso et al., 2014; Vertechi et al., 2020).

D'un point de vue évolutif, peu d'éléments sont connus en ce qui concerne la capacité du vivant à utiliser ces stratégies, dans quelle mesure certaines d'entre elles sont spécifiques à certaines espèces et pourquoi. Pourtant, les modèles animaux tels que les rongeurs sont souvent utilisés pour étudier avec des techniques invasives l'activité du cerveau qui sous-tend les processus comportementaux. Il devient donc essentiel de comprendre quelles stratégies computationnelles sont impliquées dans le processus comportemental d'adaptation à des changements brusques de règles déjà apprises dans le modèle de la souris et de comprendre dans quelle mesure ces animaux peuvent être des modèles adéquats pour mieux caractériser les bases neurales des processus cognitifs lors de l'adaptation chez l'être humain.

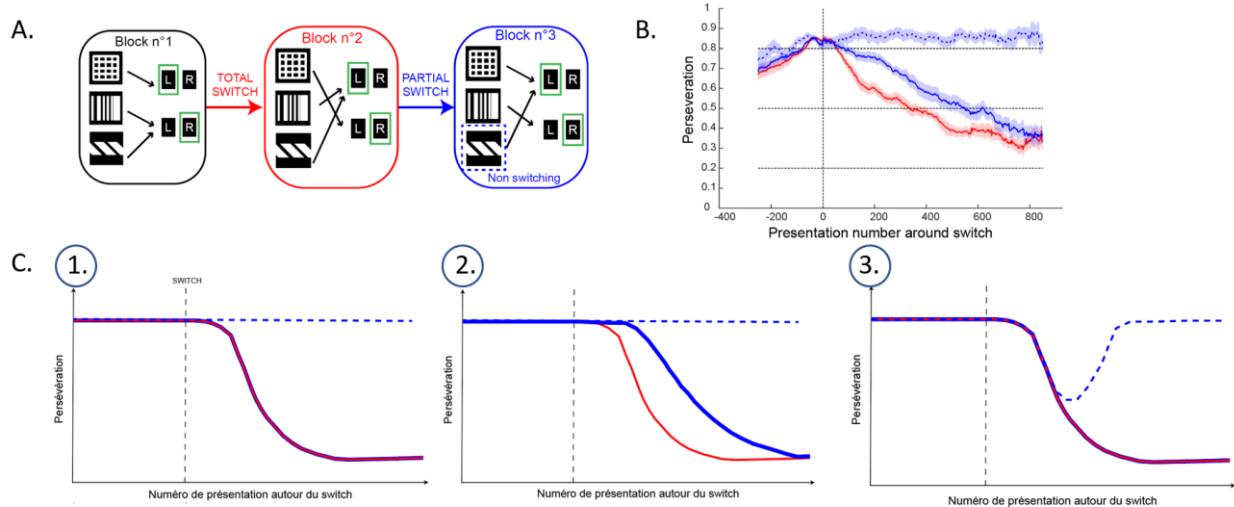
## **1.2. LA TACHE DE SWITCH : UNE NOUVELLE TACHE DE PSYCHOLOGIE EXPERIMENTALE POUR INVESTIGUER LES MECANISMES COMPUTATIONNELS DES COMPORTEMENTS ADAPTATIFS CHEZ LA SOURIS**

Pour répondre aux questions posées, une nouvelle tâche comportementale, dérivée de la tâche de renversement d'apprentissage, a été développée afin de distinguer les différentes classes d'algorithmes d'adaptation.

Dans le cadre de cette tâche, un stimulus visuel parmi trois est présenté aux souris à chaque essai sur deux écrans. Au fur et à mesure des essais, les souris doivent apprendre par essai-erreur, la réponse associée à chacun des trois stimuli (réponse sur l'écran de gauche ou sur l'écran de droite). Lorsque les souris ont appris l'ensemble des règles, un changement abrupt et non indicé des règles est effectué. Deux types de changements de règles, appelés *switch* sont réalisés successivement au cours de la tâche : (1) un renversement complet (*switch total*) au cours duquel toutes les associations (stimulus-action) sont inversées et (2) un renversement partiel (*switch partiel*) durant lequel seulement deux des associations (stimulus-action) changent et une reste identique (**Figure 1A**). Le fait qu'une des associations ne change pas lors de certains *switchs* nous permet de déterminer si le processus d'adaptation est fondé

Stratégies d'adaptation de la souris face à un environnement volatil  
**TRAVAIL EXPERIMENTAL. Chapitre 2. Projet SWITCH**

sur une stratégie de réapprentissage par renforcement simple, sur un méta-apprentissage ou sur l'utilisation d'un *task-set* (**Figure 1C, 1., 2., 3.**).



**Figure 1 : La tâche de switch permet d'identifier les différentes classes d'algorithme d'adaptation mises en jeu à la suite de changements abrupts de contingences expérimentales.**

**A.** Design expérimental. Les souris doivent apprendre l'association entre chaque stimulus visuel et le côté de réponse associé (L = côté gauche. R = côté droit). Le bloc n°1 correspond à la phase d'acquisition. Lorsque les critères de performance sont atteints, intervient un *switch total* durant lequel toutes les associations sont inversées. Puis, lorsque les critères sont de nouveaux atteints, le deuxième *switch*, appelé *switch partiel*, intervient et consiste en une modification partielle des associations. Un stimulus reste associé au même côté de réponse que précédemment dans le bloc n°2 : ce stimulus est appelé le *non switching* stimulus. Les deux autres stimuli ont leur côté de réponse associé inversé.

**B.** L'évolution du comportement de persévération des souris autour des deux *switchs* suit les prédictions d'un modèle de méta-apprentissage par renforcement. **C.** Prédictions de l'évolution de la persévération au cours de la tâche suivant différents modèles. **1.** Prédictions d'un modèle d'apprentissage par renforcement simple. **2.** Prédictions d'un modèle de méta-apprentissage par renforcement. **3.** Prédiction d'un modèle utilisant des *task-sets*. Pour plus de détails, se référer à la section « Hypothèses de travail ».

Pour les panels B et C. Courbe rouge pleine : évolution de la persévération pour les trois stimuli dont le côté de réponse change lors du *switch total* (ici, la grille, les barres verticales et les diagonales, voir 1A, Block n°2). Courbe bleue pleine : évolution de la persévération pour les stimuli dont le côté de réponse change lors du *switch partiel* (ici, la grille, les barres verticales, voir 1A, Block n°3). Courbe bleue pointillée : évolution de la persévération pour le stimulus dont le côté de réponse ne change PAS lors du *switch partiel* (dans cet exemple, les barres diagonales, voir 1A, Block n°3).

Les résultats montrent que les souris sont capables d'apprendre correctement chacune des trois associations, de détecter les changements de règles qui interviennent au cours de la tâche et de s'y adapter. L'étude du comportement de persévération autour des *switchs* montre que les souris s'adaptent aux changements brusques grâce à un mécanisme de méta-apprentissage impliquant une variation du taux d'apprentissage à la suite des renversements de contingences expérimentales, une fois le changement de règle détecté (**Figure 1B en comparaison de 1C, 2.**). Ces résultats mettent également en évidence l'absence d'utilisation de *task-set* chez la

souris dans le cadre de cette tâche (**Figure 1B en comparaison de 1C, 3.**), contrairement à ce que l'on aurait pu attendre si celle-ci avait été résolue par des sujets humains. Par ailleurs, la stratégie comportementale d'adaptation des souris dépend d'une heuristique d'exploration exercée dans l'espace des actions : à la suite des *switchs* en effet, les souris présentent une augmentation transitoire du comportement de latéralisation (enchaînement privilégié des réponses sur un côté, indépendamment du stimulus présenté). L'étude plus en détails de ces résultats permet ainsi de montrer que la variation du taux d'apprentissage est dépendante du type de switch (*total* ou *partiel*) et de la manière dont les stimuli sont groupés entre eux (par design, un côté est toujours associé à deux stimuli, tandis que l'autre est associé à un seul stimulus, **Figure 1A**).

Cette étude apporte ainsi des résultats nouveaux quant aux capacités d'ajustement des prises de décision dans un environnement volatil chez la souris et offre des perspectives intéressantes quant à l'investigation des bases neuronales de ce méta-apprentissage par renforcement ancré dans l'espace des actions.

## **2. DEVELOPPEMENT D'UNE NOUVELLE TACHE COMPORTEMENTALE, VALIDATION EXPERIMENTALE ET IDENTIFICATION DES STRATEGIES COMPORTEMENTALES ET MECANISMES COMPUTATIONNELS DE L'ADAPTATION DE LA SOURIS EN ENVIRONNEMENT VOLATIL : LE PROJET SWITCH**

*Article en cours de préparation : page 161 à 194.*

---

## ARTICLE 2

Stratégies d'adaptation de la souris face à un environnement volatil  
*TRAVAIL EXPERIMENTAL. Chapitre 2. Projet SWITCH*

# Mice rely on a meta-learning strategy when facing volatile environments

M. Euvrard<sup>1</sup>, P. Domenech<sup>1,2, \*, ☰</sup>, E. Burguiere<sup>1, \*, ☰</sup>

1. Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Inserm, CNRS, AP-HP, Hôpital de la Pitié Salpêtrière, Paris, France

2. APHP, Groupe Hospitalier Henri Mondor, DMU Psychiatry, Department of Neurosurgery, Université Paris Est Créteil, Créteil, France.

✉ Corresponding authors: [eric.burguiere@icm-institute.org](mailto:eric.burguiere@icm-institute.org), [philippe.domenech@icm-institute.org](mailto:philippe.domenech@icm-institute.org)

## SUMMARY

Flexible adaptation to abrupt environmental changes is critical for survival. To identify the computational strategy onto which mice rely for adaptation to uncertain and volatile environments, we developed a new “reversal learning”-like decision task that was performed over thousands of trials by 17 mice thanks to the use of innovative and automated high-throughput operant chambers. In this task, the mice learnt by trial-and-errors, stimulus-action-outcome associations that were periodically reversed without cues. Our results show that mice adaptation to abrupt changes probably relied on a meta-reinforcement learning strategy, modifying their learning rate when rule changes were detected. More interestingly, that meta-learning strategy further depended on an action-based heuristic for exploration. Unlike humans, mice did not use task-set to solve the task as shown by the absence of behavioural switches and counterfactual learning. Altogether, our results shed a new light on the evolution of computational decision-making capabilities and demonstrate the relevance of our setup to carry out causal investigation of the neural basis of meta-reinforcement learning.

**Keywords:** Adaptation, flexibility, volatility, mouse model, behavioural task, automated operant chamber, reinforcement learning, meta-reinforcement learning, task-set, exploration, exploitation

## INTRODUCTION

All animal species with a nervous system, from the simplest to the most complex, are capable of learning to maximize vital benefits and minimize losses that could be harmful to their survival. The simplest form of learning that any animal species is capable of is reinforcement learning. This consists of learning value tables for each (state - action) association in order to determine which choice to make in order to maximize gains in the long term (Sutton and Barto, 2018). However, the environment in which we evolve is not stationary and the rules learnt at a certain point in time can become invalid: it is therefore necessary to adapt to these abrupt changes. Even if reinforcement learning has proven to be effective over millions of years of evolution, it has two major disadvantages. On the one hand, it does not allow for rapid adaptation because it requires unlearning previously learnt rules and then relearning new ones. On the other hand, it is a memoryless process that does not allow for the reuse of tables of values learnt in similar situations encountered in the past (Koechlin, 2016). In the literature, several hypotheses exist regarding the different strategies involved in effective adaptation. While some suggest that adaptation relies on the use of simple approximations and heuristics (Findling et al., 2019; Hermoso-Mendizabal et al., 2020; Wyart and Koechlin, 2016; Yu and Cohen, 2009), others propose that adaptation relies on a meta-learning process (Behrens et al., 2007; Farashahi et al., 2017; Khamassi et al., 2013; McGuire et al., 2014; Schweighofer and Doya, 2003; Wilson et al., 2014), when some others work on the ability to detect abrupt and structured changes (Bartolo and Averbeck, 2020; Collins and Frank, 2012; Domenech and Koechlin, 2015; Domenech et al., 2020; Donoso et al., 2014; Vertechi et al., 2020). From an evolutionary point of view, it is still unclear whether all species can use these strategies, to what extent some of these strategies are specific to some species and why. Yet, animal models such as rodents are often used to study with invasive techniques the activity of the brain underlying behavioural processes. Therefore, we wish to understand which computational strategies are involved in the behavioural process of adaptation to abrupt changes of already learnt rules in the mouse model and understand to what extent these animals can be proper models to better characterize the neural bases of cognitive processes during adaptation in humans. To address this question, we developed a new behavioural task to disentangle from the different class of adaptation algorithms. In this task, the mice are presented with one (out of three) visual stimulus associated with a left or right response. As the trials progress, they must learn the correct response for each of the three stimuli. When the rules have been learnt, they are changed abruptly without cues. We perform two types of rule changes: (1) a complete change (*total switch*) in which all (stimulus-action) associations are reversed and (2) a partial change (*partial switch*) in which only two of the (stimulus-action) association change and one remains stable. The fact that one of the associations does not change during some switches allows us to disambiguate whether the adaptation process is based on a simple reinforcement relearning strategy, heuristics (lateralization, win-stay/lose shift...), meta-learning, or on the use of task-sets.

## RESULTS

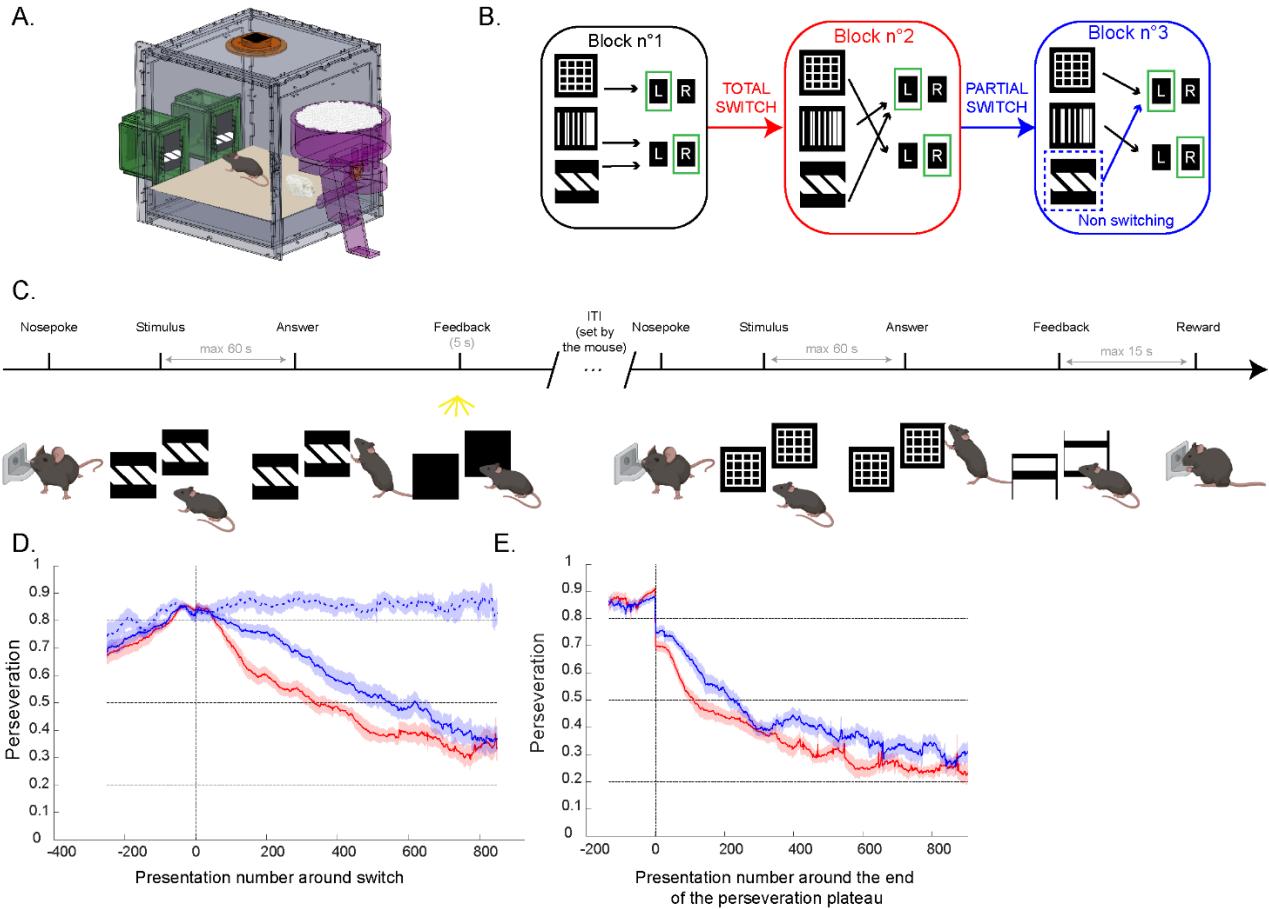
### Mice can achieve a complex learning task at their own pace

We administered the switch task to 17 male mice placed in fully automated operant chambers in which the mice could work according to their circadian rhythm, 24 hours a day, 7 days a week. These operant chambers were equipped to ensure both the life of the animal (bedding, cotton to build a nest, *ad libitum* water) and task performance (2 screens equipped with infrared beams allowing the presentation of visual stimuli and the collection of answers, a feeder in which the mice could collect their rewards) (**Figure 1A**, **Supplementary figure 1**).

The mice performed the training stages in the operant chamber prior to the task for approximately 5 days and 650 trials (**Supplementary figure 2 and 3A-B**). Then they could perform the switch task for about 45 days, completing an average of 9,000 trials (**Supplementary figure 5A, 5B**) and at least 2 switches (**Supplementary figure 5C**).

The switch task is a learning task where successive changes in the learnt rules occur. The mice had to learn by trial and errors which response was rewarding depending on the stimulus on display. Consequently, they learnt three different (stimulus-action) pairs. When the mice reached the learning criteria, a switch was triggered after 100 additional trials. Two types of switches occurred during the task: (1) a *total switch*, during which all the contingencies were reversed, and (2) a *partial switch* during which, one (stimulus-action) association remained the same (the *non-switching stimulus*) while the other two were reversed (**Figure 1B**). In this task, when the mice triggered a trial by breaking the infrared beam present within the feeder, one of the three visual patterns that comprises the task was displayed simultaneously on both screens in the operant chamber. The mouse then had 60 seconds to respond by approaching either the left or right screen. Correct responses were indicated by screens blinking and rewarded with a cereal kibble that the mouse got only if it returned to the feeder within 15 seconds. Wrong answers were indicated by a five-second flashing light (**Figure 1C, Supplementary figure 4**).

The task performance in the operant chambers was continuous, free and self-paced, which raised some questions about the homogeneity of the collected data. To tackle this issue, we first detected the mice activity bursts by grouping trials based on an adjusted maximum time interval (**Supplementary Figure 6**). Unsurprisingly, the mice were more active during the night (**Supplementary figure 5D, 5E**), while during daytime, they were less active and tended not to finish the trials they started (no response or no reward recovery when the responses were correct, **Supplementary figure 5F**). There was also a higher proportion of corrective trials which were triggered in case of an excessive lateralized behaviour (**Supplementary figure 7**). Nevertheless, the mice retained and even seemed to consolidate what they had learnt between two consecutive nights (**Supplementary figure 8**). Based on these observations, we decided to exclude daytime trials from subsequent behavioural analyses (about 15% of initial data set).



**Figure 1** The switch task paradigm implemented in an automated operant chamber allowed the disentanglement of several computational models underlying mice behaviour

(A) Experimental set-up. The mouse lived 24/24, 7/7 in an automated operant chamber equipped with 2 screens on which pictures could be displayed. The screens were equipped with an infrared break-beam system to detect the mouse answer: when the mouse was close to the screen ( $< 0.5$  cm), the break beams were broken, and the answer was detected. The same picture was displayed on the two screens at the same time. The mouse self-paced the trials and could get food from the pellet dispenser when giving the correct answer.

(B) The switch task paradigm. The mice had to learn the association between three visual stimuli and their related side of answer. L = left side. R = right side. Block n°1 corresponds to the acquisition phase. When the mice had reached 80 % of performance over the 40 latest presentations of each stimulus and 80 % of performance over the 40 latest trials, the rule associating each stimulus with one side of answer was modified after 100 trials. The first rule change (the first switch) consisted in a total reversal of all the (stimulus – side of answer) associations. The stimuli that were previously associated with the right side were then associated with the left side and vice-versa (Block n°2). The second switch consisted in a partial modification of the associations. One stimulus remained associated to the same side of answer as in block n°2: this stimulus was called the *non-switching stimulus*. The two other stimuli had their associated side of answer reversed.

(C) Sequence of two trials. The mouse started a trial by nose poking in the food well. The screens lighted on with the same picture. The mouse had 60 seconds to give an answer by coming close to the screen. If the mouse gave a correct answer, the screen started blinking and the mouse had 15 seconds to get its reward. If the mouse gave an incorrect answer, the stimulus on screens disappeared and there was a five-second flashlight. The mouse could start a trial whenever it wanted: the intertrial intervals were not fixed by the experimenter.

(D) Evolution of the perseveration around the two switches as a function of the number of presentations for each stimulus. Red line: perseveration for the three switching stimuli around the total switch. Blue line: perseveration for the two switching stimuli around the partial switch. Blue dashed line: perseveration for the non-switching stimulus around the partial switch. Each curve corresponds to the mean of the data from the 17 mice  $\pm$  SEM.

(E) Evolution of the perseveration around the end of the perseveration plateau as a function of the number of presentations for each stimulus. Red line: perseveration for the three switching stimuli during the total switch. Blue line: perseveration for the two switching stimuli during the partial switch. Blue dashed line: perseveration for the non-switching stimulus around the partial switch. Each curve corresponds to the mean of the data from the 17 mice  $\pm$  SEM.

## Mice adapt to abrupt changes using meta-reinforcement learning

We first focused on perseverative responses around switches. Perseverative responses correspond to correct responses given by the mice in the preceding episode that are no longer rewarded after the switch. We saw that for the switching stimuli, the perseveration decreased over the number of presentations of each stimulus meaning that the mice could detect the changes and adapt to them. However, it took more than eight hundred presentations for each switching stimulus to reach a perseveration level that indicated that they no longer answered according to the previous rules but rather answered according to the current ones (**Figure 1D, red and blue solid lines**). However, in the partial switch, the perseveration of the non-switching stimulus did not decrease meaning that the mice kept answering on the same side, which was still the correct side after the switch (**Figure 1D, blue dashed lines**). Even if it seemed coherent since the mice still got rewards, these results were not predicted by a model in which the animals would rely on the use of task-sets. Had the mice relied on task-sets, they would have considered the different (stimulus - side of answers) associations as a whole set of information and the observation of negative outcomes for some of the associations (for the switching stimuli) would have led to the consideration that all the rules might have changed at once. In this case we would have expected to observe a decrease in the perseveration for the non-switching stimulus. Therefore, the absence of such decrease showed that the mice did not rely on the use of task-set or the use of counterfactual learning to solve the task.

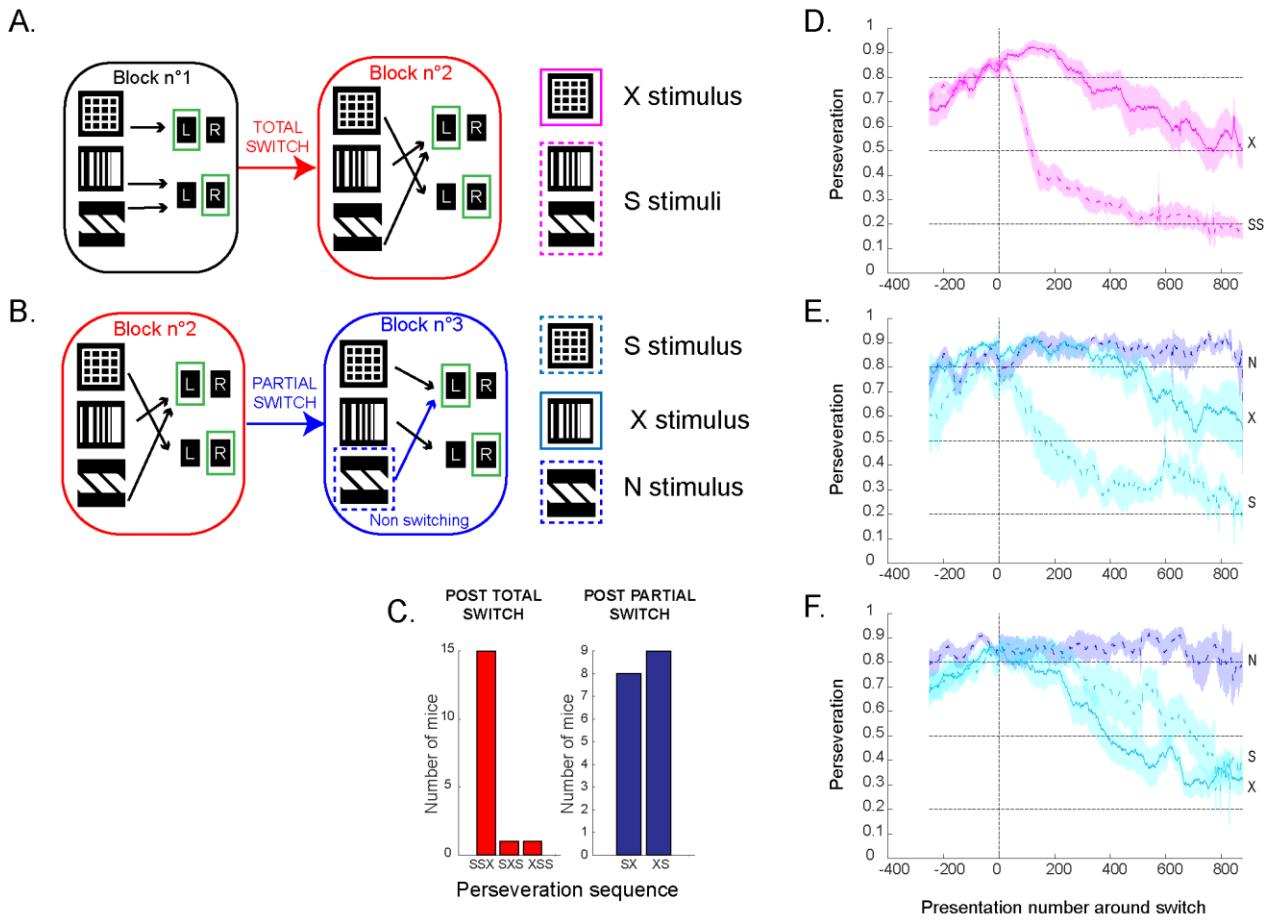
Furthermore, when we compared the perseveration curves for the switching stimuli around the total and the partial switches, we could clearly see that the curves did not overlap, and that the perseveration decreased faster for the switching stimuli after the total switch (**Figure 1D, red solid line**) than for the switching stimuli after the partial switch (**Figure 1D, blue solid line**). These results were not predicted by a simple model of reinforcement learning with constant parameters: if the mice relied on such a computational model, we would have expected to see the same decrease for the switching stimuli after the total switch and the partial switch. Indeed, if the learning parameters were identical throughout the task and the stimuli, the tables of (stimulus – action) association would be updated in a similar manner and the learning of each of the association would occur at the same rate. Consequently, these results suggested that the learning parameters were different depending on the context in which the mice evolved (e.g after a total switch versus after a partial switch). These results would suggest that the mice relied on a meta-reinforcement learning computational process with adaptive learning parameters. Particularly, the fact that both the perseveration curves tended to the same asymptote but did not overlap suggested that the learning rate was different between the total switch and the partial switch, with a higher learning rate after the total than the partial. The causes of such a difference in the learning rate had then to be investigated. Finally, it is to be noticed that the perseveration reduction for all the switching stimuli after the total and partial switches seemed progressive and continuous (**Figure 1D, red and blue solid lines**). However, this effect could just be related to an averaging effect between the stimuli and the mice, erasing thus a

dropping effect of the perseveration. To investigate this question, we characterized the perseveration dynamics with a hidden Markov chain model (Smith et al., 2004, 2007) that allowed us to determine the length of the perseveration plateau right after the switch and just before the perseveration started decreasing significantly (**Supplementary Figure 10**). When perseverative responses for each stimulus and each mouse were realigned on the appropriate stimulus presentation that marked the end of the mouse perseveration, we showed that there was no sharp drop of the perseveration as it would have been observable if the mice were able to switch their behaviour. The end of the perseveration plateau was not abrupt, and the probability of giving a correct answer gradually shifted to a new equilibrium therefore ruling out a switching mechanism.

### The learning rate variation depends on action space

The hidden Markov chain model previously used to identify the end of the perseveration plateau was also applied to determine the number of presentations required to learn the associated correct side of answer for each stimulus (**Supplementary Figure 10**). The analysis revealed that there was a clear-cut distinction on the perseveration and learning dynamics between the different stimuli depending on how they were associated to the sides of answer after the two switches (**Figure 2A-B, Supplementary Figure 10**). To investigate this aspect more thoroughly, we established the perseveration sequence (**Figure 2C**) and the learning sequence (**Supplementary Figure 10**) for each mouse. After the total switch, 11 mice out of the 17 first stopped perseverating on the stimuli associated to the same side (S stimuli) and then on the stimulus alone on its side (X stimulus). These 11 mice then learnt first the S stimuli before learning the X stimulus. Thus, most of the mice displayed a SSX perseveration and learning sequences after the total switch (SSX subgroup). After the partial switch however two subgroups of mice were identified. 7 mice out of 17 first stopped perseverating on the stimulus associated to the same response side as the non-switching stimulus (S stimulus) and then on the stimulus alone on its side (X stimulus). These 7 mice then learnt first the S stimulus and then the X stimulus (SX subgroup). 7 mice out of 17 presented the opposite sequence, namely XS perseveration and learning sequences (XS subgroup). Because of the existence of subgroups of mice as well as a difference in the processing of the different stimuli, we repeated the analyses of perseveration taking these aspects into account (**Figure 2D-E-F**).

For the total switch (**Figure 2D**), we could observe for the stimuli associated to the same response side (S stimuli): (1) a short plateau of perseveration, (2) then an abrupt drop in perseveration to reach chance level around 175 presentations after the switch, (3) finally, a slowing down in the perseveration speed reduction until a relearning of the associations around 800 presentations after the switch. For the stimulus alone on its response side (X stimulus), we observed on the contrary: (1) an excessively long plateau of perseveration, suggesting that the learning for this stimulus was blocked during 300 presentations, (2) then a slow decrease of the perseveration.



**Figure 2 The mice seemed to rely on a meta-reinforcement learning strategy associated to the action space and modulated by the type of switch**

(A) Illustration of the categories of stimuli around the total switch. By design, the S stimuli corresponded to the two stimuli associated to the same side of answer both before and after the total switch. The X stimulus corresponded to the stimulus alone on its side of answer both before and after the total switch.

(B) Illustration of the categories of stimuli around the partial switch. The S stimulus corresponded to the stimulus associated to the same side of answer as the non-switching stimulus (N) *after* the partial switch. The X stimulus corresponded to the stimulus alone on its side of answer *after* the partial switch.

(C) Bar charts of the perseveration sequences observed after the two switches. These histograms were established with the analysis of the duration of the perseveration plateau for each stimulus. S: stimuli associated to the same side of answer. X: stimulus alone on its side of answer. For example, after the *total switch* 15 mice stopped perseverating on one of the stimuli associated to the same side of answer, then the second one, and finally stopped perseverating on the stimulus alone on its side of answer.

(D) Evolution of the perseveration around the total switch when splitting the data of the stimuli associated to the same side of answer (S, pink dashed line) from the stimulus alone on its side of answer (X, solid pink line) for the mice that first stopped perseverating and learnt the 2 stimuli associated to the same side (SSX subgroup). Each curve corresponds to the mean of the data from 11 mice  $\pm$  SEM.

(E) Evolution of the perseveration around the partial switch when splitting the data of the stimulus associated to the same response side as the non-switching stimulus (S, light blue dashed line) from the stimulus alone on its side of answer (X, solid light blue line) for the mice that first stopped perseverating and learnt the stimulus associated to the same side as the non-switching stimulus (SX subgroup). The data from the non-switching stimulus are represented with the dark blue dashed line (N). Each curve corresponds to the mean of the data from 7 mice  $\pm$  SEM.

(F) Evolution of the perseveration around the partial switch when splitting the data of the stimulus associated to the non-switching stimulus (S, light blue dashed line) from the stimulus alone on its side of answer (X, solid light blue line) for the mice that first stopped perseverating and learnt the stimulus alone on its side of answer (XS subgroup). The data from the non-switching stimulus are represented with the dark blue dashed line (N). Each curve corresponds to the mean of the data from 7 mice  $\pm$  SEM.

After 800 presentations of this stimulus, the animals still persevered up to 50% for this visual stimulus. The overall results seemed to imply that: (1) the learning rate was variable, pointing at the existence of a meta-learning phenomenon as previously suggested, (2) the variation of the learning rate depended on the way the stimuli were associated with each other for their response side.

However, after the partial switch, the evolution of the perseveration was not the same depending on the subgroup of mice. For the SX subgroup (**Figure 2E**), the same tendency as for the SSX subgroup of mice after the total switch was observed, even if the blockade of the learning for the stimulus alone on its side of answer seemed longer (around 500 presentations). However, for the XS subgroup (**Figure 2F**), the perseveration reduction for the X stimulus tended to be more like the S stimulus perseveration reduction. Results after the partial switch compared to results after the total switch clearly indicated that the learning rate did not evolve similarly after the different categories of switches.

Altogether, these results indicate that the mice were unable to learn all the three (stimulus – side of answer) associations at the same time and had to prioritize the learning of one or two associations before learning the other(s). This selective learning could be explained by a variable learning rate adjusted relatively to the action space in which the mouse solved the task and the type of switch the mouse had just faced. The mice may have thus relied on a lateralization heuristic to solve the task and adjust their learning.

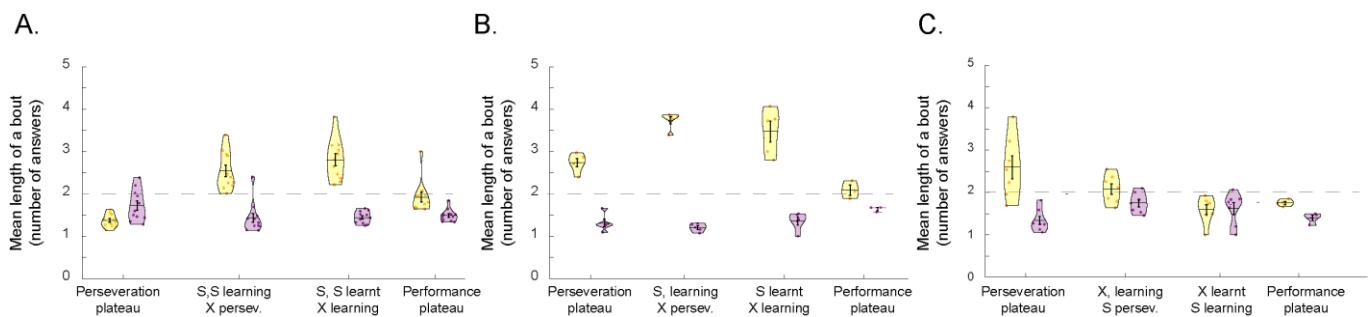
### **The mice rely on a lateralization strategy to explore their environment**

After the switches, we notably noticed a surge in the proportion of corrective trials for the stimulus alone on its side of answer (**Supplementary figure 11B-C**), suggesting that the mice kept on answering on the side associated to the two other stimuli. To properly quantify the mice lateralization, we calculated for different periods the mean length of a sequence of consecutive answers towards the side associated to 2 stimuli versus the mean length of a sequence of consecutive answers towards the side associated to one stimulus (**Figure 3**).

After the total switch, the mean length of a sequence of consecutive answers towards the side associated with two stimuli was longer than the one towards the side of answer associated with one stimulus, suggesting that the mice from the SSX subgroup tended to lateralize more towards the side associated to two stimuli (**Figure 3A**). After the partial switch, for the SX subgroup of mice, the same tendency was observed (**Figure 3B**), suggesting consequently that the mice from the SX subgroup also tended to lateralize towards the side associated with two stimuli. After the partial switch and following this line of thought, for the XS subgroup of mice, we would have expected to see longer sequences of answer towards the side associated with the X stimulus. However, no such result could be seen: the length of the sequences of consecutive answers towards the two sides were not different (**Figure 3C**). The fact that the mice tended not to lateralize much in this condition might also be related to the non-switching stimulus.

Indeed, even if the mice would tend to answer more on the side associated to the X stimulus, the recurrent presentation of the non-switching stimulus for which they still gave good answers made them answer on the counter-side, breaking thus the lateralization bouts.

Altogether, these results seemed to indicate that: (1) the non-switching stimulus is considered as an independent stimulus, (2) the lateralization bias of the mice might be synergic to the meta-reinforcement learning mechanism, explaining thus the gating of the learning for the stimulus alone on its side of answer and the dynamic variation of the learning rate for the stimuli associated to the same response side, (3) mice evolving thus in the action space of the task, one might assume that lateralizing is a way of exploring new unknown contingencies. From this perspective, lateralization could consist in a form of directed exploration.



**Figure 3 The mice tended to lateralize after the switches**

The different plots represent the mean length of sequences of answer towards the side associated to two stimuli (yellow violin plots) and towards the side associated to one stimulus (purple violin plots). In other words, we have calculated how many times in a row the mice would answer towards the two sides depending on the different switches (total A., partial B. and C.), the subgroup of mice (A. SSX, B. SX, C. XS). The different phases in the task were determined based on the learning dynamics of each stimulus.

## DISCUSSION

The combined approach of a new behavioural task with the use of high throughput automated operant chambers helped us disentangle the strategies used by mice when facing a volatile environment. We notably showed that the learning occurs sequentially due to mice lateralization strategy to explore their environment, depending on the type of switch and the categories of stimuli. These results would suggest that mice adaptation to abrupt changes probably relied on a meta-reinforcement learning strategy, based on a heuristic for exploration in action-space.

## Cutting edges methods to study the mice behaviour: quality, quantity and generalization of the data

The strength of our behavioural system is related to the fact that the mice can self-pace their activity and that we can collect a huge number of trials for a long period of time. In this respect, our behavioural system was particularly adapted to our needs as we were able to study the adaptive strategies of mice over

long experimental periods, in a complex task without any interruption. Contrary to experimental designs where the animals work in blocks of trials, imposed by the experimenter, and are put back and forth every day in the experimental system, in our case, the animals always remained in the same context. There is thus no cofounding factors or disruptive element that could have distracted the animals in their adaptive strategies: the mice only needed to adapt to the changes in the task but did not have to adapt to the recurrent manipulation of the experimenter and the daily changes of environment (home-cage to experimental system and vice-versa). Besides, considering that we needed around 10,000 trials per animal to observe the results we show, collecting such a large amount of data would have taken around 200 days if we consider that the mice perform around 50 trials within an hour (Mar et al., 2013). On top of that, such a large data collection provides useful material for further computational approach.

Nevertheless, some issues must be addressed, notably during the training stages. We notably showed that during the two final training stages, when the mice could answer correctly on both sides, some mice tended to lateralize on one side of answer (**Supplementary Figure 3C**) and this could be related to the natural handedness of the animals (Signore et al., 1991) . Yet, for some mice, this lateralization in the training stages had some consequences for the early learning in the task. Indeed, if the mice had a favourite side of answer, they would naturally tend to answer on the same side during the early phase of acquisition leading to the immediate learning of one or two associations and the delayed learning of the other(s) (**Supplementary Figure 9A**). In our lab, we are thus trying to improve the previously established shaping protocols (Benzina et al., 2021) in order to reduce the propensity of the mice to lateralize during the training stages.

Furthermore, our set-up did not allow us to properly record the reaction time. Indeed, the mouse triggered the trials by poking in the food-well which was located at the opposite side of the screens in the operant chamber. Depending on the trajectory of the mice as well as its speed, the reaction times could not be recorded consistently. Besides, we let the mice the possibility to answer in maximum 60 seconds which is a huge delay. Even when excluding the 10 % trials with the highest reaction time, the data remained noisy with a high variability in the reaction time (data not shown). Because of that, we could not study the evolution of the reaction time around the switches. This would have been really interesting since for humans, the abrupt changes in the contingencies of the task are followed by an increase in the reaction time (Domenech et al., 2020; Rogers and Monsell, 1995). It would have been interesting to check whether this effect was induced in our task. Hardware and software developments in the lab are already undertaken to be able to measure with a higher precision this variable.

Finally, this work has been conducted with male C57BL6 mice only due to experimental constraints. Yet the strain of the mouse has been shown to have an impact on the efficiency in behavioural task resolution (Turner et al., 2017) and the on the learning parameters (Luksys et al., 2009). The gender of the animals also seems to influence the behavioural strategies implemented to solve the task. For example, female

mice in a reversal learning task seems to lateralize more than male mice (Chen et al., 2021; Harris et al., 2020). However, in the switch task we have shown that male mice seemed to use an action-based strategy: would it be enhanced with females solving the task? Working with female mice and different strains of mice could be a particularly interesting perspective to check whether the observed results can be reproduced in different individuals or if other strategies can be highlighted.

### Mice adaptative strategies in the landscape of evolutionary behaviours

Despite the complex design of our task, the mice were able to detect the changes in the rules and to adapt to them. They progressively stopped perseverating (**Figure 1D**) and learnt the new associations. Yet they needed hundreds of presentations of each stimulus to be able to learn the new associations (**Supplementary figure 9H, I**) and their learning was sequential. Besides, after the partial switch, for the non-switching stimulus, the mice kept answering the same way as they used to do before the switches. These results have two main consequences.

First mice have more limited learning abilities than humans. This is consistent with translational studies that compare the resolution of the same task between humans and rodents. Benzina and collaborators showed for example that when it takes around 25 trials to reach the criteria to reversals in a probabilistic reversal learning task, mice need around 450 trials in the deterministic version of the same task (Benzina et al., 2021). Humans learn faster than mice, implying that the learning rates in mice are far below the ones observed in humans and that they may not be able to generalize prior structural knowledge (Bartolo and Averbeck, 2020). Secondly, we proved that contrary to humans (Collins and Koechlin, 2012; Domenech and Koechlin, 2015), mice seem unable to rely on task-sets to adapt to complex and abrupt changes in their environment. We also noticed that mice were unable to display an abrupt change of their behaviour after the switch (**Figure 1E**). When relocking the evolution of the perseveration around the end of the perseveration, we do not observe a sharp decrease of this variable. This is not consistent with studies that show that mice can quickly switch their behaviour and can solve a foraging reversal learning task with an inference-base strategy as humans would do (Vertechi et al., 2020). Yet, in these tasks, mice are overtrained to perform the successive reversals. If we had let the mice perform more switches, we might have also observed such a tendency. Yet, mice behaviour is more complex than what a simple reinforcement learning model would forecast: a meta-reinforcement learning model would be more accurate to predict the behavioural results. Thus, these analyses could be consistent with previous studies that also state that a reinforcement-learning model with meta-parameters adjusted by various modulatory factors can actually predict rodent behaviour (Funamizu et al., 2012; Luksys et al., 2009; Strösslin et al., 2006). Nevertheless, these analyses must be verified by some modelling work to check whether a meta-reinforcement learning model is indeed the most appropriate. Besides, we need to be able to fit the data and extract the values of the different learning parameters, notably the learning rate to characterize

quantitatively its variation instead of simply comparing qualitatively the reduction in the speed of the perseverative behaviours.

Finally, it is to be noticed that in our task, the mice faced first a total switch and then a partial switch so that they can be aware that they evolve in a changing environment when they first encounter the total switch. Yet, it would be interesting to test different sequences of switches (two total, one partial then one total, two partial) and check whether the observed results in this study can be strictly reproduced or if some different results would emerge, suggesting that maybe other cognitive processes could be at stake.

### **Lateralization and behavioural variability: behavioural noise or actual adaptive strategies?**

Even though we implemented a deterministic task, the mice never reach 100 % of performance even for the non-switching stimulus (**Figure 1D**, **Supplementary Figure 9A, B, C**). This suggests that the mice still made mistakes some errors even when they had properly learnt the association, as it is observed across other studies using rodents. Many reasons have been put forward to explain these lapses but a recent study suggest that they do not occur independently over time but rather arise from the animals switching between engaged and disengaged states during sessions of task resolution (Ashwood et al., 2020). Applying the models developed by these authors could be an interesting way to deeply characterize the dynamics of the strategies of the mice in our task and check whether they indeed fluctuate between different discrete states or if the errors we observe only represent random noise. Importantly these errors can consist in a way to explore one's environment as an epsilon-greedy reinforcement learning model would predict it (Sutton and Barto, 2018). Selecting from time to time another answer can give useful information about one's environment and can help to detect whether the rules have changed or not. In this perspective, lapses can reflect exploration (Pisupati et al., 2021) and consist in a useful strategy to predict changes and adapt to them. A probabilistic version of the task, where experimental noise is added to the volatility of the task could be an interesting way to analyse how mice deal with different forms of uncertainty (expected and unexpected, (Yu and Dayan, 2005)).

Mice seemed to rely on a meta-reinforcement learning modifying their learning rate when the rule changes were detected. We tried to identify what could explain this variable learning rates. Behrens and collaborators notably showed that the learning rate can be modulated as a function of the volatility of the environment (Behrens et al., 2007). In our case, after the partial switch, one of the associations remains the same which is why the perception of volatility after the partial switch may be altered. In other words, after the partial switch, the uncertainty about the fact that rules have changed is not the same as after the total switch and this could explain why the variation of the learning rate looked different after the total and partial switches (**Figure 1D and 2D-E-F**). However, we show in our task that these learning differences are due to the lateralization bias of the animals after the switches (**Figure 3, Supplementary figure 11**) towards one side of answer implying that lateralization can be a form of directed exploration

towards unknown new contingencies and an adaptive cognitive resources allocation to efficiently learn (Rmus et al., 2021). Interestingly, after the total switch most, of the mice lateralized towards the side associated to two stimuli (**Figure 2A, 2D, Supplementary figure 9E, 9H, 10, 11B**) but after the partial, two subgroups of mice clearly emerged with two different dynamics of lateralization (**Figure 2C, 2E, 2F**). This learning profile reveals that mice selected one side out of two to be able to start learning but the reasons why they selected one side of answer rather the other still needs to be investigated. One could imagine that, even though the sides of answer were balanced (50% each), mice would tend to answer more to the side associated with more stimuli associated to each. Mice are able to distinguish quantities (Panteleeva et al., 2013) and are able to count (Çavdaroglu and Balci, 2016). Selecting one side with two stimuli that bring a reward could thus be perceived by mice as a more effective strategy than selecting one side associated to only one stimulus even if the task is not designed that way. In this perspective, we could have expected that after the partial switch, mice would have behaved the same way as after the total switch. Yet, as the animals consider the stimuli as independent and they still exploit the correct association for the non-switching stimulus, there remains two sides of answer and two stimuli whose side of answer is still unknown. Picking randomly one of the two sides is thus an easy way to first learn one of the remaining associations, explaining maybe why we observe two subgroups of mice after the partial switch. The fact that mice exploit a lateral bias is quite compatible with studies focusing on the heuristics of perceptual decision making. For example, rats can exploit a transition bias to perform an auditory task and an internal estimate of the sequence repeating probability (Hermoso-Mendizabal et al., 2020). Furthermore, as previously suggested in one study on a reversal learning task (Harris et al., 2020), the lateralization process is a way to reduce the dimension of the task and can fasten the acquisition of one or two (image – side) association(s), therefore helping the animal to quickly adapt right after a change in the rules. Thus, lateralization could be a default strategy to quickly adapt that would be progressively replaced as the animals get a better insight on their environment. This exploratory bias has been notably characterized as being advantageous for learning and increasing the fitness of animals (Vallortigara and Rogers, 2005). Recent studies seem also to suggest that action space is actually a crucial and separable dimension in the context of reinforcement learning (Rmus et al., 2021). Yet, it should be checked whether there is no experimental bias related to the stimuli we used as it has been shown that stimulus similarity can have an effect on behavioural laterality (Treviño, 2014; Treviño et al., 2013). To have a better view on the exploration processes, it could have also been interesting to modify a bit our set-up. We only used an experimental set-up with two screens which might have biased towards the natural tendency of rodents to lateralize (Signore et al., 1991). Besides, it can limit the interpretation of the results (Piantadosi et al., 2019). Adding a third screen to the set-up that would not have been associated to the different stimuli would have been a proper way to study the exploration strategies of mice and disentangle the different forms of explorations (random exploration, directed exploration...).

## MATERIAL AND METHODS

### Animals

A total of 17 specific pathogen-free (SPF) wild-type male C57BL/6J mice were born, weaned at 21 days, and raised in groups in our own animal facility in individually ventilated cages on a 12 h light-dark cycle (lights on at 8:00 AM and off at 8:00 PM) at 21°C and 40 % humidity. They had unlimited access to water and food and were provided with nesting material. The experiment started when the animals reached the mean age of 4 months. When the experiments started the mice were isolated in individual operant chambers to perform the behavioural task (see Experimental set-up and Behavioural task). During this period, they did not have access to *ad libitum* food but were delivered 20 mg AIN-76 rodent tablets (Test Diet) when they gave correct answers (see Behavioural task). However, they had unlimited access to water.

The animals were always manipulated by the same female experimenter all along the protocol. Their weight was regularly monitored, and they were supplemented with additional pellets if they weighted under 80 % of their initial mass to strictly compensate the weight loss. The overall wellness of the animals was assessed by monitoring the skin turgor, and looking at the animal activity, posture, face, fur cleanliness and any sign of distress would have led to the animal exclusion of the study (Spangenberg and Keeling, 2016). All the procedures were in accordance with European and French laws and conducted in agreement with institutional guidelines.

### Experimental set-up

C57BL/6 male mice performed the behavioural task in automated operant chambers modified from the Med Associates (Vermont, USA) ENV-007CTX experimental chambers measuring  $30.5 \times 24.1 \times 29.2$  cm (**Fig.1A, Supplementary figure 1**). The stainless-steel tray was covered with bedding and the mice were provided with a cotton pad for nest building. The operant chamber was equipped as previously described (Benzina et al., 2021) apart from the touch function of the screen. The left wall was equipped with two screens as previously described but these two screens were equipped with an infrared beams frame to detect the mouse answer when the animal was close to the screens breaking the infrared bundles. This allowed a better detection of the mice answer than the touchscreen pad.

This experimental set-up allowed a more ecological on-going of the task as the animals could work whenever they wanted with a limited intervention of the experimenter to limit the stress of the animals (**Supplementary figure 6A**).

### Behavioural task

The behavioural task was divided into 2 parts: the training phases and the switch task itself. The task was coded in MatLab R2017b and administrated with a computer connected to the Med Associate system.

## Training phases

The training phases were necessary for the mice to get used to the cages and the behavioural set-up. There were divided in 4 steps. First, the mice had to learn where the pellets were delivered. To achieve this, a pellet was automatically distributed every minute in the food well, and then the mice had to nose poke 50 times in the food well to get a pellet. During the second part, the mice had to learn to answer on the screens. When exploring the cage, if they crossed the infrared beams in front of one of the screens, a pellet was delivered. The mice had 15 s to retrieve the reward otherwise the trial was considered as wrong, and the countdown was reset to 10 trials to complete anew: this short delay enables a proper association between the action and the rewarding outcome. Finally, the animals had to understand the global structure of the task. When they nose-poked in the food well, both screens turned on and the mice had one minute to answer on one of them and 15 s to retrieve the reward. They had to complete 10 successive complete trials within the imposed delays otherwise the countdown was reset to 10 trials to complete anew (**Supplementary figure 2**).

## Switch task

This new paradigm was improved from a reversal learning task. The switch task is based on the use of three visual stimuli mapping onto two responses. During this task, mice could start a trial by nose poking in the food-well (detection with infrared break beam sensors), and then one of the three visual stimuli used in the task was displayed simultaneously on both screens. The mouse had to respond within one minute to the stimulus by approaching either the left or the right screen according to an association rule (e.g. grid – answer left; diagonal or vertical bars – answer right, **Fig.1B**). Mice got a reward (food pellets) each time they answered on the screen associated with the stimulus displayed. They were exposed to an aversive stimulus (strong light) whenever they selected the wrong screen. The trial was considered as complete and correct when the mice selected the correct screen and retrieved the pellet within 15 s, otherwise it was considered as wrong (**Fig.1C, Supplementary figure 3**).

Once the animals had reached 80 % of correct answers over the 40 latest trials as well as 80 % of performance over the latest 40 presentations of each stimulus, the task rule associating the shapes to a response (left or right) changed after one hundred more on trials on the current rule. This change in rule was what we called a switch. There were two kinds of switches. In partial switches, the new rule had one (stimulus-action-outcome) association in common with the previous action set: the side of response changed for two stimuli out of three and one of them -the ‘non switching stimulus’- remained associated to the same side for the response. However, during total switches, the new rule had no (stimulus-action-outcome) association in common with the previous action set: the side of response changed for all the stimuli (**Fig.1B**). All the 17 mice underwent two switches, one total followed by one partial.

Stimuli were presented pseudo-randomly by presenting on average 50% of the time stimuli with a left response side and 50% of the time stimuli with a right response side to limit the lateralization phenomenon that some mice showed during training (**Supplementary figure 3C**). Thus, to avoid induced spatial lateralization, the two stimuli on the same side were presented 25 % of the time each, while the stimulus on the counter side was presented 50 % of the time. Besides, when lateralization on one side was detected, mice were forced to answer on the other side: these trials were called corrective trials. They were started only if the mice had a performance under 50 % over 6 trials for which they gave the answer on the same side (spatial lateralization). The stimulus/i whose response side was opposite to the lateralizing side was/were presented until the animals gave a correct response (**Supplementary figure 7**).

## Behavioural data

During the task, the following data were registered in a .mat file: the current rule, the number of trials, the time of each trial start, the stimulus displayed, the mouse answer (right or left), the correctness of the response (correct or incorrect), the reaction time (time between the nose poke in the food-well and the answer), the pellet retrieval time (time between the answer and the nose poke in the food-well to get the cereal kibble), the nature of the trial (corrective or non-corrective) and the type of trials (e.g correct, incorrect, no answer, pellet non retrieved).

## Analysis and statistics

All the analysis, statistics and figures of this article were produced with custom-made Matlab codes (R2017b, The Mathworks). Due to the small number of animals, we used non-parametric tests. We detail below the principal steps of analysis.

### Intertrial interval threshold and actograms

The actograms (**Supplementary figure 6A**) are based on the calculation of an intertrial interval (ITI) threshold for each mouse (**Supplementary figure 6B**). Two trials separated by an ITI superior to this threshold are considered as belonging to two different bursts of activity.

Briefly for each trial n ( $n > 2$ ), we calculated the following ratio R and studied its distribution. The first value superior to 95<sup>th</sup> quantile of the distribution is considered as the threshold value.

$$R(n) = ITI(n) + \log\left(\frac{ITI(n-2) + ITI(n-1)}{2}\right)$$

Trials that did not belong to any block were considered as isolated trials.

In some mice, the bursts of activity occurred predominantly during the nocturnal period of the animal facility (8 PM - 8 AM, **Supplementary Figure 6A left**) but other mice exhibited a phase shift in the

circadian rhythm (**Supplementary Figure 6A right**) suggesting that for some mice, their subjective days and nights were different from the day-night cycle imposed by the animal facility. As a consequence, to classify the trials and these blocks as day or night trials/blocks, we could not use the circadian rhythm from the animal facility (8 AM – 8 PM: day ; 8 PM – 8 AM: night).

To overcome this difficulty, we led an analysis of the activity rhythm for each animal. We subdivided the duration of the task in minutes. If at least one trial occurred during a minute, we put 1 for this minute, 0 otherwise. Calculating the cumulative sum of this vector enabled us to distinguish the active parts of the task (slope > 0) from the non-active parts of the task (slope = 0).

We were then able to establish the actograms for each animal (**Supplementary figure 6A**) and to compare the qualitative and quantitative aspects of the trials occurring during the subjective days and nights for each animal (**Fig.1D, Fig.1E, Supplementary figure 5D, 5E Supplementary figure 7B**).

### Information retention

To evaluate how the mice maintained what they learnt over time, an index called “Information retention” was calculated by comparing the performance of the mice at different periods. We compared the performance of the mice: (1) At the final third of each night compared to the first third of the same night (**Supplementary figure 8A, blue violin plot**); (2) At the final third of each night compared to the performance during the first third of the following night (**Supplementary figure 8A, purple violin plot**), (3) At the final third of each night compared to the performance during the following day (**Supplementary figure 8A, blue violin plot**); (4) At the first third of each night compared to the performance during the previous day (**Supplementary figure 8A, purple violin plot**). For each night/day, the following ratio was calculated:  $I = (P_2 - P_1)/P_1$  where  $P_2$  is the performance for the trials occurring after the trials used to calculate the  $P_1$  performance (for example  $P_2$  is the performance at the end of night  $n$  compared to the performance  $P_1$  at the beginning of night  $n$ ). For each night/day, these ratios were averaged for each animal after removing the ratio involving a day or a night where the switches occurred in order to avoid the bias related to the consequent drop of performance. A negative value indicates that  $P_1$  is superior to  $P_2$  and that there is a decrease of performance (the mice tend to do worse or forget what they knew) whereas a positive value indicates an increase in performance (the mice tend to do better and increase their knowledge on the task).

### Data filtering

The initial set of data was filtered according to different steps.

First, we removed artefactual trials. Indeed, sometimes, some cotton from the nest blocked the infrared beams from the feeder or the screens creating artefactual answers or pellet retrievals. These trials were

characterized by extremely fast answer or pellet retrievals (< 1 s) which was impossible due to the configuration and the size of the chamber.

Then we removed the trials for which the mouse did not give any answer as they were non-informative for the mice since they did not get any feedback.

We finally removed the day trials as we showed that: (1) their number was negligible compared to the number of trials during the nights, (2) the quality of the trials suggested that the mice were not much involved during the day (more non answered trials, more good answers with no retrieved pellets) (**Fig.1D**, **Fig.1E**, **Supplementary figure 5D-E-F**).

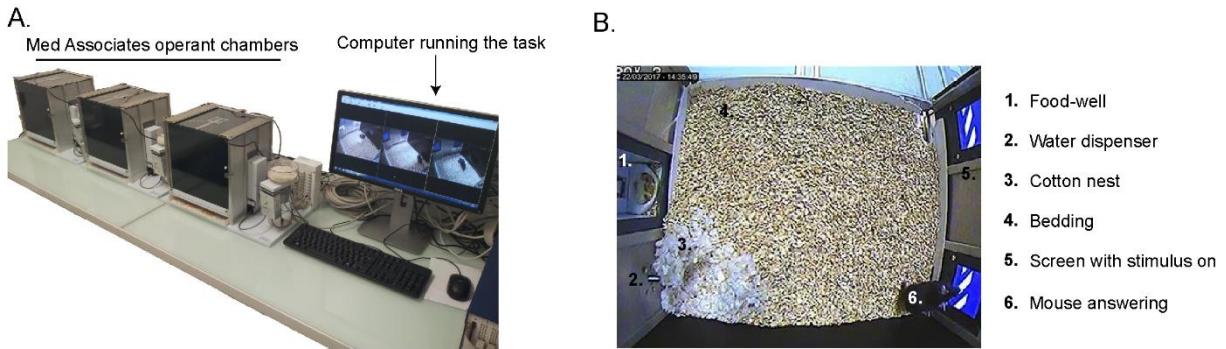
### **Learning trials and perseveration plateau**

To assess the learning trials as well as the last trial of perseveration post-switch, we adapted the codes provided by Smith and collaborators (Smith et al., 2004, 2007). The matlab- and winBUGS-based software can estimate the learning curve and the dynamics of the learning process of each single subject by analysing the correctness of the answers in the task of interest. The learning trial and the trial of the end of perseveration were defined for each animal and each stimulus by analysing the reasonable certainty dynamics, considering that from that point of the experiment this subject performs better than chance.

### **Analysis around the switch**

The analyses around the switch were performed using the filtered data (see Data filtering). The raw data were extracted and aligned for each mouse around the switch as defined by the computerized task. The mean and SEM were calculated for each of the plotted results.

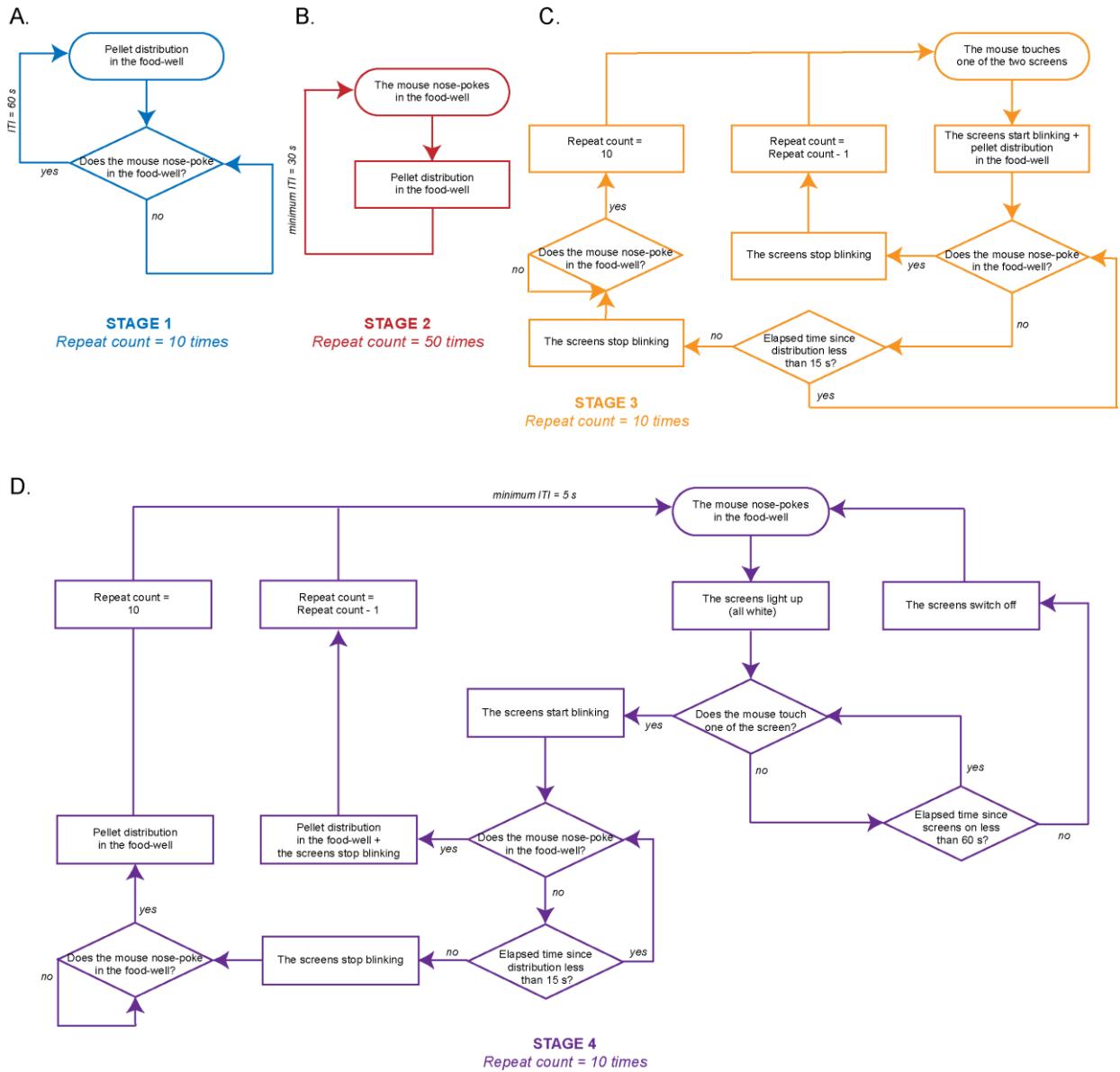
## SUPPLEMENTARY FIGURES AND LEGENDS



**Figure S1 Experimental set-up based on commercial operant chambers made automated.**

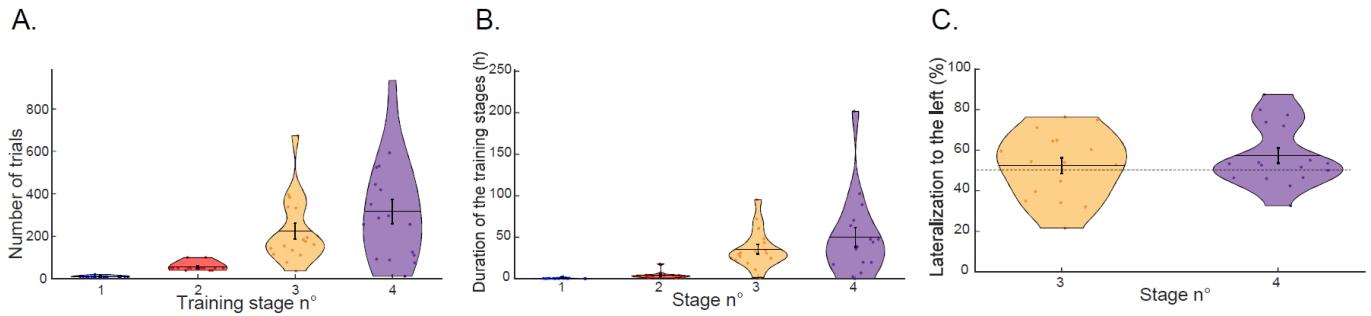
(A) Outside view of the Med associate operant chambers modified and made automated for the protocol. Several cages can be connected to the same computer and run continuously the task without the intervention of the experimenter.

(B) Inside view of the operant chamber with a mouse answering on the right screen. The mouse can live in the cage for more than a month of experiment (*ad libitum* water, bedding, nest material) and gets its food when performing properly the task.



**Figure S2 Flow chart of the training stages: shaping the mice behaviour.**

Four training stages were required to shape the mouse behaviour towards the final stage. These stages helped the mice to understand how the operant chamber works from the pellet delivery system (Stages 1 and 2) to the answers on the screen (Stage 3) and the way to start and complete a trial (Stage 4). In the most difficult stages (3 and 4) the mice had to complete 10 successive trials (correct actions and timings) in a row to go to the next stages.

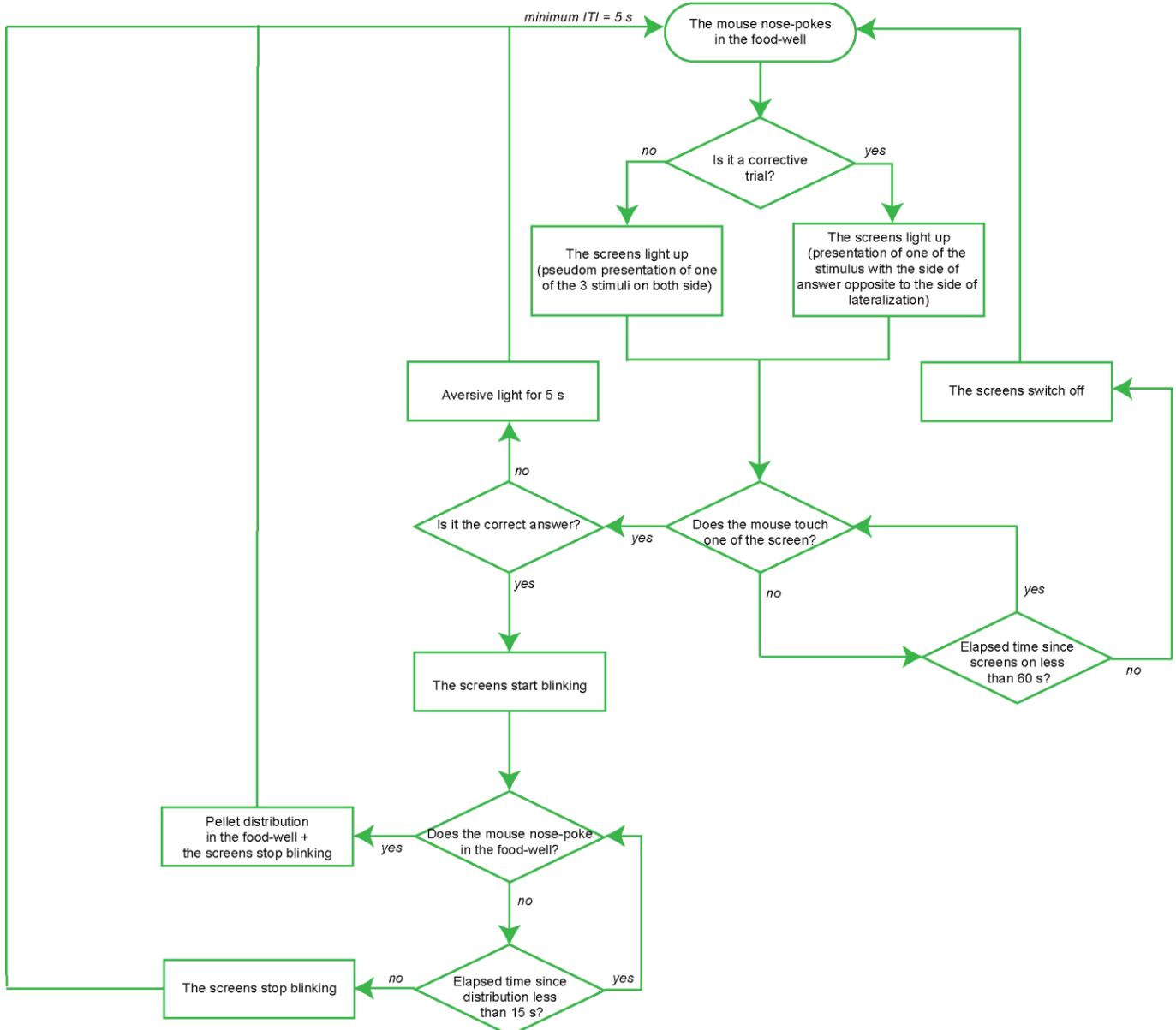


**Figure S3 Overview of the behaviour of the mice during the training stages**

(A) Number of trials performed by the mice during each stage of the training session. Each dot represents the data for one mouse. The value for the group is represented by the mean  $\pm$  SEM.

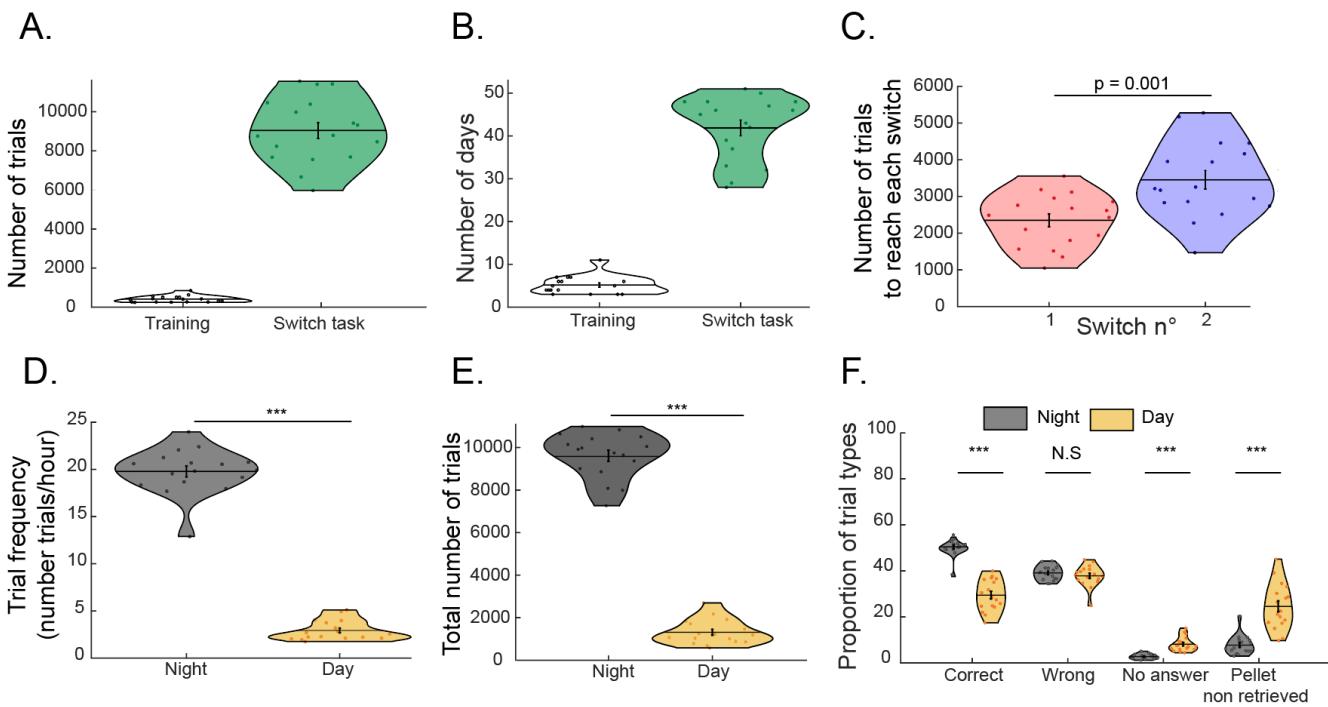
(B) Time spent by the mice in each stage of the training session. Each dot represents the data for one mouse. The value for the group is represented by the mean  $\pm$  SEM.

(C) Evaluation of the lateralization of the mice during the 2 last stages of training: quantification of the propensity of each mouse to preferentially answer on the left screen. Each dot represents the data for one mouse. The value for the group is represented by the mean  $\pm$  SEM. The dashed line represents 50 % of lateralization to the left side: in this case the mouse has no side preference. Between 0 and 50 %, the mouse preferentially answers on the right screen. Between 50 and 100 %, the mouse preferentially answers on the left screen.



**Figure S4 Flow chart of the switch task.**

During the switch task, the mice had to start a trial by nose poking in the food-well and pick the rewarded side associated with the displayed stimulus before 60 seconds. If they gave the correct answer, the screen started to blink, and they had 15 seconds to go to the food-well to get their pellet. If they gave an incorrect answer, a flashlight switched on for 5 seconds.



**Figure S5 Overview of the behaviour of the mice during the switch task**

(A) Number of trials performed by the mice during the training and the switch task. Each dot represents the data for one mouse. The value for the group is represented by the mean  $\pm$  SEM.

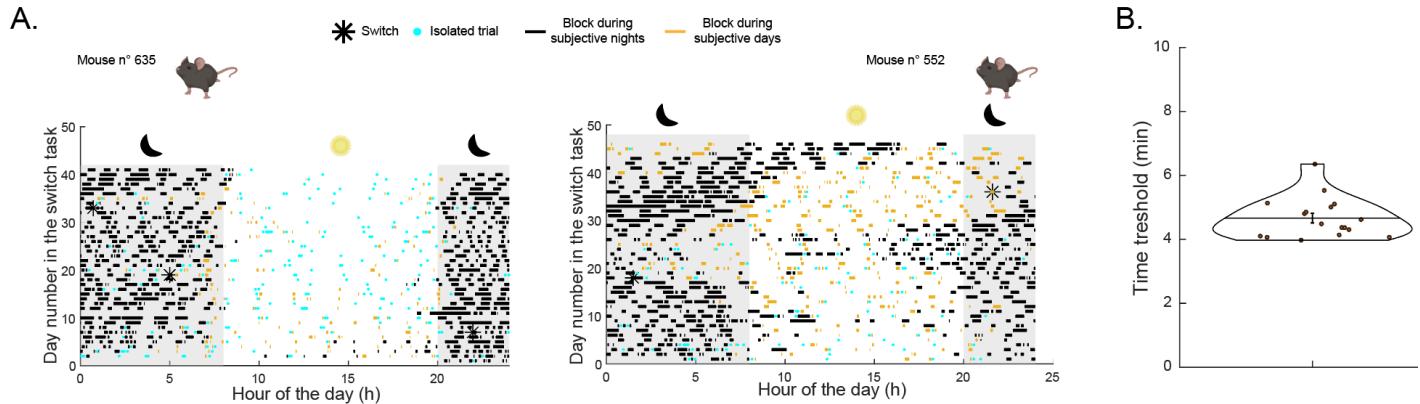
(B) Time spent by the mice in the training session and the switch task. Each dot represents the data for one mouse. The value for the group is represented by the mean  $\pm$  SEM.

(C) Comparison of the number of trials performed by the mice to reach the first switch (number of trials in block n°1) and the second switch (number of trials in block n°2). Each dot represents the data for one mouse. The value for the group is represented by the mean  $\pm$  SEM. Wilcoxon signed-rank test.

(D) Comparison of the trial frequency during daytime and night-time. Each dot represents the data for one mouse. The value for the group is represented by the mean  $\pm$  SEM. Wilcoxon signed-rank test, \*\*\*  $p = 0.0003$ .

(E) Comparison of the total number of trials performed by the mice during daytime and night-time. Each dot corresponds to the data of one mouse. Wilcoxon signed-rank test, \*\*\*  $p = 0.0003$ .

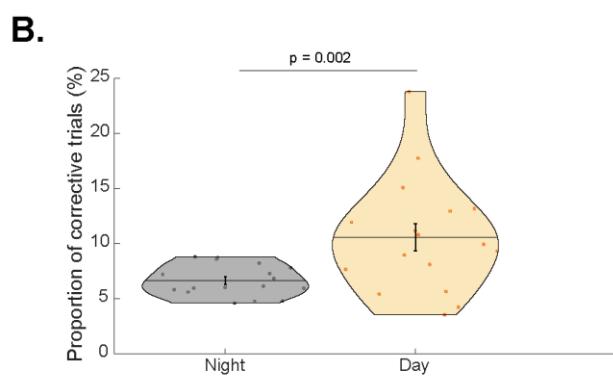
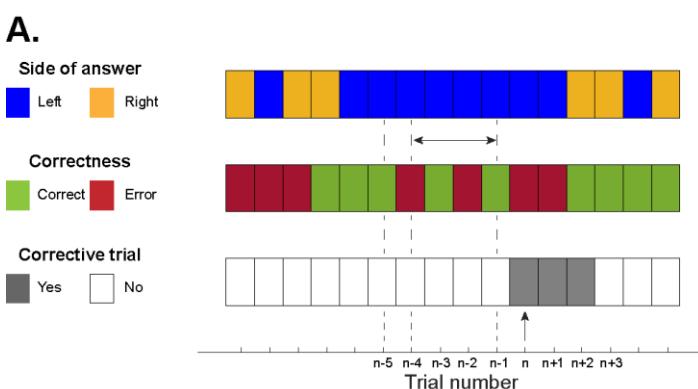
(F) Comparison of the proportion of each type of trials performed by the mice during daytime and night-time. Each dot corresponds to the data of one mouse. Wilcoxon signed-rank test, N.S: non-significant, \*\*\*  $p = 0.0003$ .



**Figure S6 The self-paced activity of the mouse imposes an innovative way of managing data**

(A) Examples of the actograms of two mice. The actograms sum up the mice activity by differentiating trials occurring within blocks of trials (trials separated by an intertrial interval inferior to a specific value computed for each animal, see S7B.) from isolated trials. The blocks of trials were then classified as “night blocks” (black lines) or “day blocks” (yellow lines) based on the frequency of the trials. Contrary to the mouse n°365, the mouse n°552 clearly showed a shift in its circadian rhythm. Consequently, it would have been incorrect to filter the trials based on the circadian rhythm of the animal facility (between 8 AM and 8 PM) and the development of this method has proven useful to correctly filter the trial occurring during the *subjective days* of the mice.

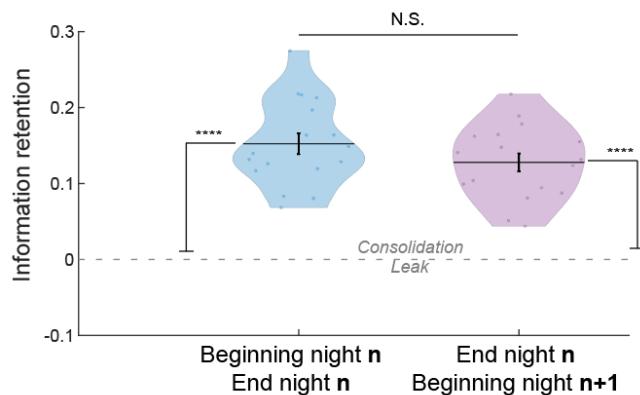
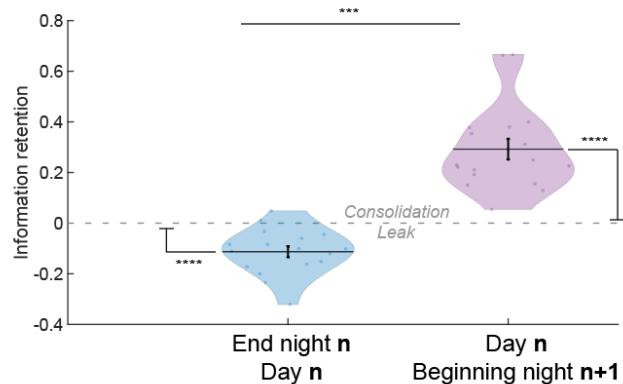
(B) Interval trial time threshold estimated for each mouse to gather trials into blocks of activity ( $n = 17$ ). Two trials with an interval trial time superior to this value belong to two different consecutive bursts of activity.



**Figure S7 The corrective trials consist in a proxy for lateralization quantification**

(A) Example of a sequence of trials that leads to corrective trials. When the mouse answers two times in a row on the same side, giving one bad answer and one correct answer (n-5, n-4), the next three trials are being rigorously evaluated. If the mouse keeps answering on the same side and that the overall performance over the four latest trials is below 50% (n-4 to n-1), corrective trials are presented (trial n). During these trials, the stimuli/stimulus associated with the other side are being presented as long as the mouse does not give the correct answer (n to n+2). Once the mouse has given an answer on the counter side (n+2), the corrective trials stop (n+3). Due to the method used to trigger these corrective trials, their presence is an indicator of the lateralization of the mouse.

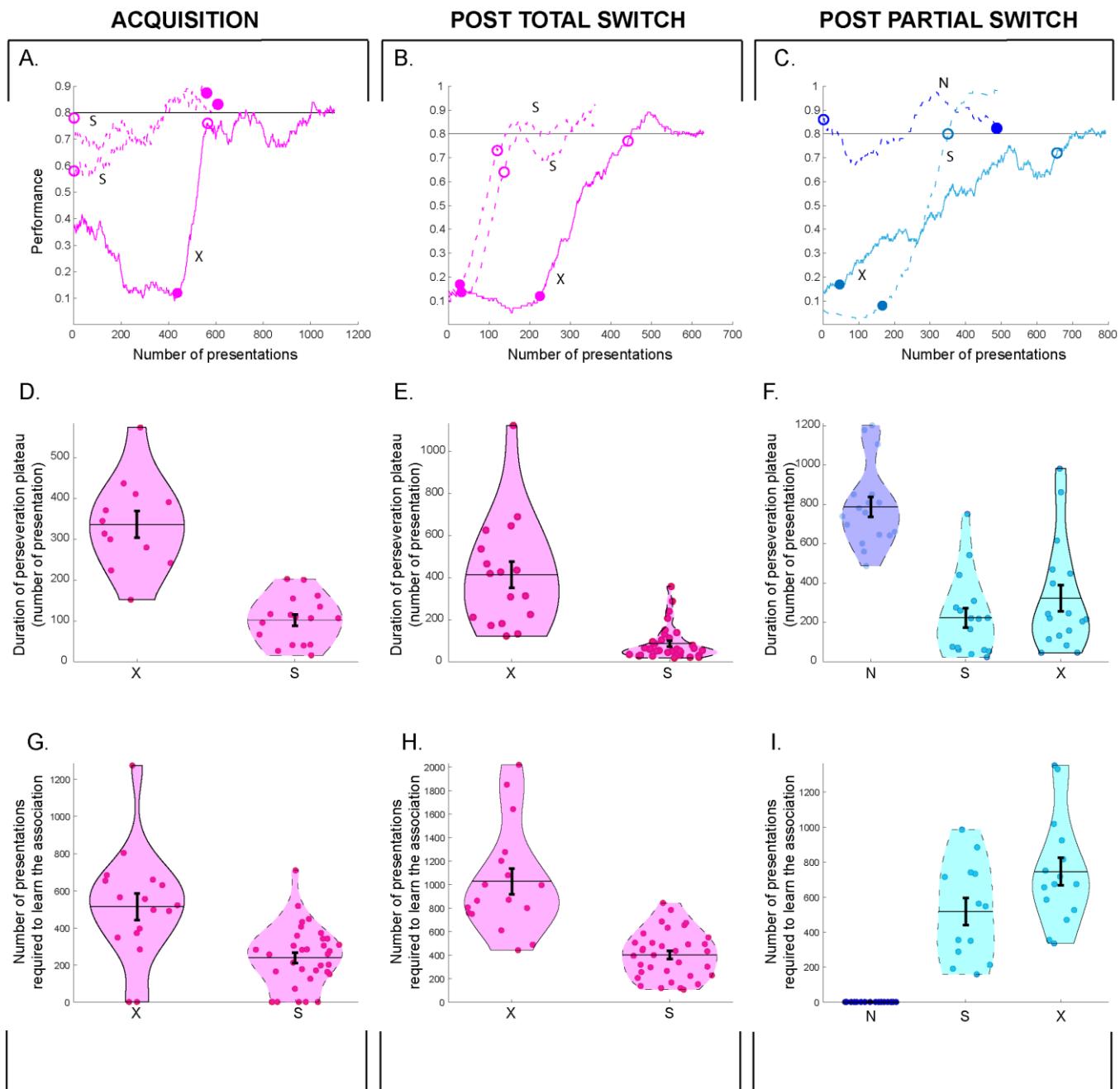
(B) Comparison of the proportion of corrective trials during daytime and night-time. The mice tend to more lateralize during the day than the nights. Each dot represents the data for one mouse ( $n = 17$ ). The value for the group is represented by the mean  $\pm$  SEM. Wilcoxon signed-rank test.

**A.****B.**

**Figure S8 The performance of the mice is altered during the day but there is no consequence on the learning for the following night**

(A) Evaluation of the information retention during the night periods. Both indexes are positive indicating that the mice tend to learn and improve their performance during a whole night of activity and that during the day they do not forget the associations they have previously learnt. On the contrary sleep might help consolidate learning. Wilcoxon signed-rank test, N.S: not significant, \*\*\*\* p < 10<sup>-5</sup>. Each dot represents the data for one mouse. The value for the group is represented by the mean ± SEM (n = 17).

(B) The mice tend to poorly perform during the day for the few trials they do (negative index) but this does not have any deleterious impact on their performance when they start again to do lots of trials during the following night. Wilcoxon signed-rank test, \*\*\* p = 0.0003, \*\*\*\* p < 10<sup>-5</sup>. Each dot represents the data for one mouse. The value for the group is represented by the mean ± SEM (n = 17).



**Figure S9 Identification of learning and perseveration phases**

**(A)** Localization of the end of the perseveration plateau (full circle) and of the learning presentation (empty circle) for one mouse during the acquisition. Each curve corresponds to the smoothed performance over 100 presentations of one stimulus. S: stimuli associated to the same side of answer. X: stimulus alone on its side of answer.

**(B)** Localization of the end of perseveration plateau (full circle) and of the learning presentation (empty circle) for one mouse after the total switch. Each curve corresponds to the smoothed performance over 100 presentations of one stimulus. S: stimuli associated to the same side of answer. X: stimulus alone on its side of answer.

**(C)** Localization of the end of perseveration plateau (full circle) and of the learning presentation (empty circle) for one mouse after the partial switch. Each curve corresponds to the smoothed performance over 100 presentations of one stimulus. N: non switching stimulus. S: stimulus associated to the same side of answer as the non-switching stimulus. X: stimulus alone on its side of answer.

**(D)** Duration of the perseveration plateau for each stimulus during acquisition. After the last training stage and at the beginning of acquisition, some mice presented a strong lateralization bias to the right side or left side explaining why some mice displayed a perseveration plateau even during acquisition for some stimuli. Each dot represents the data for one stimulus. The value for the group is represented by the mean  $\pm$  SEM. S: stimuli associated to the same side of answer. X: stimulus alone on its side of answer.

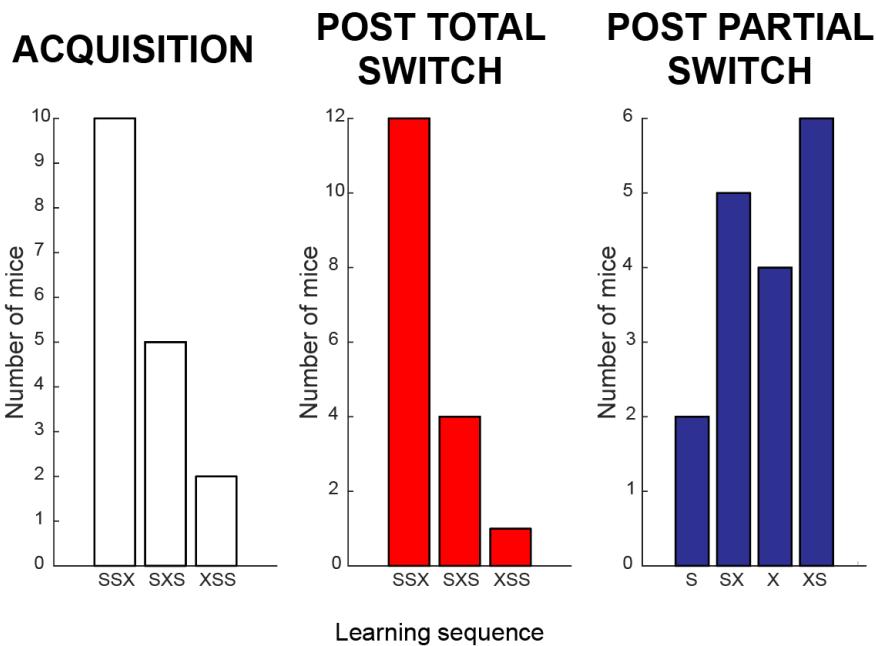
(E) Duration of the perseveration plateau for each stimulus after the total switch. Each dot represents the data for one stimulus for one mouse. The value for the group is represented by the mean  $\pm$  SEM. S: stimuli associated to the same side of answer (n = 34). X: stimulus alone on its side of answer (n = 17).

(F) Duration of the perseveration plateau for each stimulus after the partial switch. Each dot represents the data for one stimulus for one mouse. The value for the group is represented by the mean  $\pm$  SEM. N: non switching stimulus (n = 17). S: stimulus associated to the same side of answer as the non-switching stimulus (n = 17). X: stimulus alone on its side of answer (n = 17).

(G) Number of presentations for each stimulus required to learn the associated side of answer during acquisition. Each dot represents the data for one stimulus for one mouse. The value for the group is represented by the mean  $\pm$  SEM. S: stimuli associated to the same side of answer (n = 34). X: stimulus alone on its side of answer (n = 17).

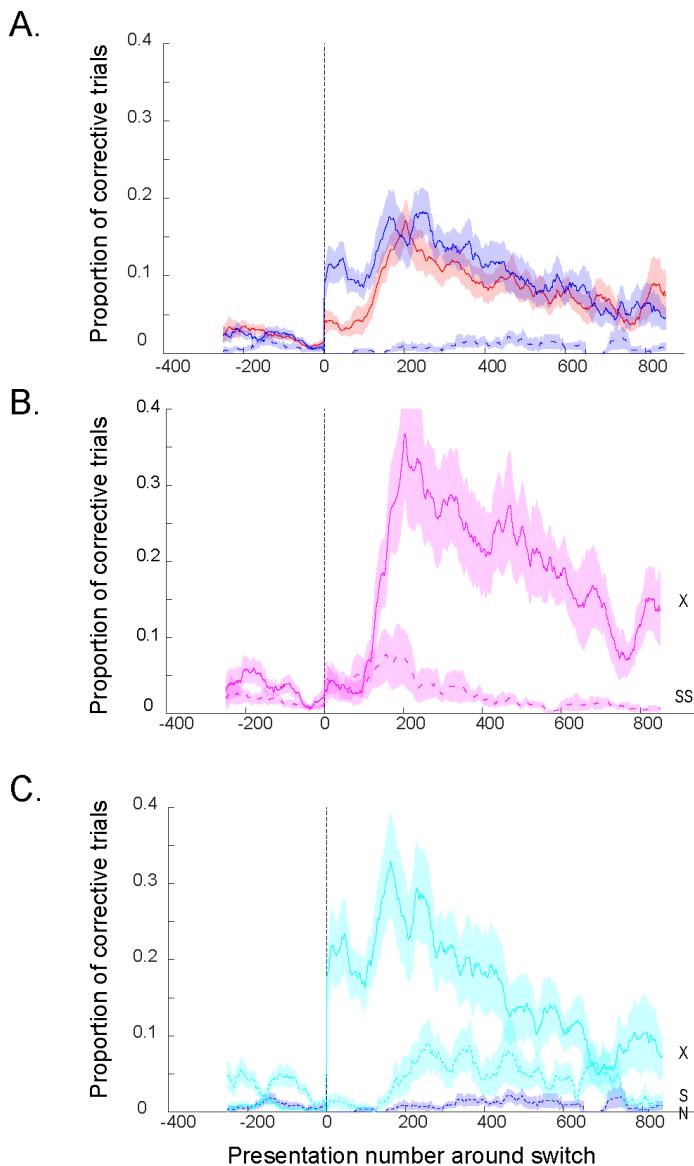
(H) Number of presentations for each stimulus required to learn the associated side of answer after the total switch. Each dot represents the data for one stimulus for one mouse. The value for the group is represented by the mean  $\pm$  SEM. S: stimuli associated to the same side of answer (n = 34). X: stimulus alone on its side of answer (n = 17).

(I) Number of presentations for each stimulus required to learn the associated side of answer after the partial switch. Each dot represents the data for one stimulus for one mouse. The value for the group is represented by the mean  $\pm$  SEM. S: stimuli associated to the same side of answer. N: non switching stimulus (n = 17). S: stimulus associated to the same side of answer as the non-switching stimulus (n = 17). X: stimulus alone on its side of answer (n = 17).



**Figure S10 Learning sequences**

Histograms of the learning sequences observed during acquisition and after the two switches. These histograms were established based on the analysis of the number of presentations required to learn the side of answer for each stimulus. Acquisition and Total Switch. S: stimuli associated to the same side of answer. X: stimulus alone on its side of answer. Partial Switch. S: stimulus associated to the same side of answer as the non-switching stimulus. X: stimulus alone on its side of answer.



**Figure S11 The rise in the number of corrective trials after the two switches indicates that the mice tend to display a lateral bias**

(A) Evolution of the proportion of corrective trials around the two switches. Red line: proportion of corrective trials for the three switching stimuli around the total switch. Blue line: proportion of corrective trials for the two switching stimuli around the partial switch. Blue dashed line: proportion of corrective trials for the non-switching stimulus around the partial switch. Each curve corresponds to the mean of the data from the 17 mice  $\pm$  SEM.

(B) Evolution of the proportion of corrective trials around the total switch when splitting the data from the stimuli associated to the same side of answer (S, pink dashed line) from the stimulus alone on its side of answer (X, pink solid line). Each curve corresponds to the mean of the data from the 17 mice  $\pm$  SEM.

(C) Evolution of the proportion of corrective trials around the partial switch when splitting the data from the stimulus associated to the same side as the non-switching stimulus (S, light blue dashed line) from the stimulus alone on its side of answer (X, light blue solid line). The data from the non-switching stimulus are represented with the dark blue dashed line (N). Each curve corresponds to the mean of the data from the 17 mice  $\pm$  SEM.

## AUTHOR CONTRIBUTIONS

Conceptualization, E.B, Ph.D; Methodology, E.B, Ph.D, M.E; Hardware, M.E. Software, M.E. Resources, M.E. Investigation, M.E. Formal analysis, M.E. Data curation, M.E. Writing – original draft preparation, M.E. Writing – review and editing, E.B, Ph.D. Funding acquisition, E.B.

## DECLARATION OF INTEREST

The authors declare no competing interests.

## ACKNOWLEDGMENTS

All animal work was conducted at the PHENO-ICMice facility. The Core is supported by 2 “Investissements d’avenir” (ANR-10- IAIHU-06 and ANR-11-INBS-0011-NeurATRIS) and the “Fondation pour la Recherche Médicale”.

The research leading to these results has received funding from the program “Investissements d’avenir” ANR-10- IAIHU-06.

We would like to thank Sami Beaumont for his brilliant and most useful feedbacks during discussions about this work. We would like to thank Pierre Tissier for the 3D picture of the operant chamber (Figure 1A). We would like to thank Guillaume Penderia and Nabil Benzina for the initial adaptation of MedAssociates codes to Matlab codes.

## REFERENCES

- Ashwood, Z.C., Roy, N.A., Stone, I.R., The International Brain Laboratory, Churchland, A.K., Pouget, A., and Pillow, J.W. (2020). Mice alternate between discrete strategies during perceptual decision-making (Neuroscience).
- Bartolo, R., and Averbeck, B.B. (2020). Prefrontal Cortex Predicts State Switches during Reversal Learning. *Neuron* *106*, 1044–1054.e4.
- Behrens, T.E.J., Woolrich, M.W., Walton, M.E., and Rushworth, M.F.S. (2007). Learning the value of information in an uncertain world. *Nat. Neurosci.* *10*, 1214–1221.
- Benzina, N., N'Diaye, K., Pelissolo, A., Mallet, L., and Burguière, E. (2021). A cross-species assessment of behavioral flexibility in compulsive disorders. *Commun. Biol.* *4*, 1–12.
- Çavdaroglu, B., and Balcı, F. (2016). Mice can count and optimize count-based decisions. *Psychon. Bull. Rev.* *23*, 871–876.
- Chen, C.S., Ebitz, R.B., Bindas, S.R., Redish, A.D., Hayden, B.Y., and Grissom, N.M. (2021). Divergent Strategies for Learning in Males and Females. *Curr. Biol. CB* *31*, 39–50.e4.
- Collins, A., and Koechlin, E. (2012). Reasoning, learning, and creativity: frontal lobe function and human decision-making. *PLoS Biol.* *10*, e1001293.
- Collins, A.G.E., and Frank, M.J. (2012). How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *Eur. J. Neurosci.* *35*, 1024–1035.
- Domenech, P., and Koechlin, E. (2015). Executive control and decision-making in the prefrontal cortex. *Curr. Opin. Behav. Sci.* *1*, 101–106.
- Domenech, P., Rheims, S., and Koechlin, E. (2020). Neural mechanisms resolving exploitation-exploration dilemmas in the medial prefrontal cortex. *Science* *369*.
- Donoso, M., Collins, A.G.E., and Koechlin, E. (2014). Foundations of human reasoning in the prefrontal cortex. *Science* *344*, 1481–1486.
- Farashahi, S., Rowe, K., Aslami, Z., Lee, D., and Soltani, A. (2017). Feature-based learning improves adaptability without compromising precision. *Nat. Commun.* *8*, 1768.
- Findling, C., Skvortsova, V., Dromnelle, R., Palminteri, S., and Wyart, V. (2019). Computational noise in reward-guided learning drives behavioral variability in volatile environments. *Nat. Neurosci.* *22*, 2066–2077.
- Funamizu, A., Ito, M., Doya, K., Kanzaki, R., and Takahashi, H. (2012). Uncertainty in action-value estimation affects both action choice and learning rate of the choice behaviors of rats. *Eur. J. Neurosci.* *35*, 1180–1189.
- Harris, C., Aguirre, C., Kolli, S., Das, K., Izquierdo, A., and Soltani, A. (2020). Unique features of stimulus-based probabilistic reversal learning. *BioRxiv* *2020.09.24.310771*.
- Hermoso-Mendizabal, A., Hyafil, A., Rueda-Orozco, P.E., Jaramillo, S., Robbe, D., and de la Rocha, J. (2020). Response outcomes gate the impact of expectations on perceptual decisions. *Nat. Commun.* *11*, 1057.
- Khamassi, M., Enel, P., Dominey, P.F., and Procyk, E. (2013). Medial prefrontal cortex and the adaptive regulation of reinforcement learning parameters. *Prog. Brain Res.* *202*, 441–464.
- Koechlin, E. (2016). Prefrontal executive function and adaptive behavior in complex environments. *Curr. Opin. Neurobiol.* *37*, 1–6.
- Luksys, G., Gerstner, W., and Sandi, C. (2009). Stress, genotype and norepinephrine in the prediction of mouse behavior using reinforcement learning. *Nat. Neurosci.* *12*, 1180–1186.
- Mar, A.C., Horner, A.E., Nilsson, S.R.O., Alsiö, J., Kent, B.A., Kim, C.H., Holmes, A., Saksida, L.M., and Bussey, T.J. (2013). The touchscreen operant platform for assessing executive function in rats and mice. *Nat. Protoc.* *8*, 1985–2005.

McGuire, J.T., Nassar, M.R., Gold, J.I., and Kable, J.W. (2014). Functionally dissociable influences on learning rate in a dynamic environment. *Neuron* *84*, 870–881.

Panteleeva, S., Reznikova, Z., and Vygonayilova, O. (2013). Quantity judgments in the context of risk/reward decision making in striped field mice: first “count,” then hunt. *Front. Psychol.* *4*.

Piantadosi, P.T., Lieberman, A.G., Pickens, C.L., Bergstrom, H.C., and Holmes, A. (2019). A novel multichoice touchscreen paradigm for assessing cognitive flexibility in mice. *Learn. Mem.* *26*, 24–30.

Pisupati, S., Chartarifsky-Lynn, L., Khanal, A., and Churchland, A.K. (2021). Lapses in perceptual decisions reflect exploration. *ELife* *10*, e55490.

Rmus, M., McDougle, S., and Collins, A. (2021). The role of executive function in shaping reinforcement learning. *Curr. Opin. Behav. Sci.* *38*, 66–73.

Rogers, R., and Monsell, S. (1995). Costs of a Predictable Switch Between Simple Cognitive Tasks. *J. Exp. Psychol. Gen.* *124*, 207–231.

Schweighofer, N., and Doya, K. (2003). Meta-learning in Reinforcement Learning. *Neural Netw.* *16*, 5–9.

Signore, P., Nosten-Bertrand, M., Chaoui, M., Roubertoux, P.L., Marchaland, C., and Perez-Diaz, F. (1991). An assessment of handedness in mice. *Physiol. Behav.* *49*, 701–704.

Smith, A.C., Frank, L.M., Wirth, S., Yanike, M., Hu, D., Kubota, Y., Graybiel, A.M., Suzuki, W.A., and Brown, E.N. (2004). Dynamic Analysis of Learning in Behavioral Experiments. *J. Neurosci.* *24*, 447–461.

Smith, A.C., Wirth, S., Suzuki, W.A., and Brown, E.N. (2007). Bayesian analysis of interleaved learning and response bias in behavioral experiments. *J. Neurophysiol.* *97*, 2516–2524.

Spangenberg, E.M.F., and Keeling, L.J. (2016). Assessing the welfare of laboratory mice in their home environment using animal-based measures--a benchmarking tool. *Lab. Anim.* *50*, 30–38.

Strösslin, T., Ito, M., and Doya, K. (2006). Estimating reinforcement learning meta-parameters from animal behavioural data. p.

Sutton, R.S., and Barto, A.G. (2018). Reinforcement Learning: An Introduction, Second edition.

Treviño, M. (2014). Stimulus similarity determines the prevalence of behavioral laterality in a visual discrimination task for mice. *Sci. Rep.* *4*, 7569.

Treviño, M., Oviedo, T., Jendritza, P., Li, S.-B., Köhr, G., and De Marco, R.J. (2013). Controlled variations in stimulus similarity during learning determine visual discrimination capacity in freely moving mice. *Sci. Rep.* *3*, 1048.

Turner, K.M., Simpson, C.G., and Burne, T.H.J. (2017). BALB/c Mice Can Learn Touchscreen Visual Discrimination and Reversal Tasks Faster than C57BL/6 Mice. *Front. Behav. Neurosci.* *11*.

Vallortigara, G., and Rogers, L.J. (2005). Survival with an asymmetrical brain: advantages and disadvantages of cerebral lateralization. *Behav. Brain Sci.* *28*, 575–589; discussion 589–633.

Vertechi, P., Lottem, E., Sarra, D., Godinho, B., Treves, I., Quendera, T., Oude Lohuis, M.N., and Mainen, Z.F. (2020). Inference-Based Decisions in a Hidden State Foraging Task: Differential Contributions of Prefrontal Cortical Areas. *Neuron* *106*, 166–176.e6.

Wilson, R.C., Takahashi, Y.K., Schoenbaum, G., and Niv, Y. (2014). Orbitofrontal cortex as a cognitive map of task space. *Neuron* *81*, 267–279.

Wyart, V., and Koechlin, E. (2016). Choice variability and suboptimality in uncertain environments. *Curr. Opin. Behav. Sci.* *11*, 109–115.

Yu, A., and Cohen, J.D. (2009). Sequential effects: Superstition or rational behavior? *Adv. Neural Inf. Process. Syst.* *21* - Proc. 2008 Conf. *21*, 1873–1880.

Yu, A.J., and Dayan, P. (2005). Uncertainty, Neuromodulation, and Attention. *Neuron* *46*, 681–692.

# CHAPITRE 3. Investigation neuromodulatrice de l'implication du cingulaire murin dans l'adaptation aux changements abrupts de l'environnement

---

## 1. INTRODUCTION

Les êtres vivants évoluent dans un environnement volatil et les changements qui interviennent peuvent être abrupts, récurrents et imprévisibles. Afin de comprendre quelles sont les réponses comportementales à ces changements ainsi que les processus computationnels qui les sous-tendent, nous avons développé une nouvelle tâche chez la souris, la tâche de *switch*. Dans le cadre de cette nouvelle tâche comportementale implémentée dans un système de chambre opérante automatisée permettant aux souris de réaliser la tâche sur des temps longs (un mois durant lequel deux changements de règles interviennent), nous avons montré précédemment au chapitre 2 de la section « Travail expérimental » :

- 1) La résolution de la tâche et l'adaptation aux changements abrupts de règles ne sont ni expliquées par un modèle d'apprentissage par renforcement simple, ni par un modèle utilisant des *task-sets*. Nos résultats montrent que l'adaptation des souris aux changements brusques repose plus probablement sur une stratégie de métapprentissage par renforcement impliquant une variation du taux d'apprentissage lorsque les changements de règles sont détectés.
- 2) À la suite d'un changement brutal des règles apprises, une phase de persévération est observable pour l'ensemble des figures. Durant cette phase, les souris continuent de répondre selon les règles précédentes et semblent apparemment insensibles aux nouveaux résultats associés aux différents stimuli. Néanmoins, cette phase semble être critique.
- 3) En effet, à la fin de cette phase de persévération, un changement de comportement est observé : les souris entrent dans une phase de latéralisation leur permettant d'apprendre plus rapidement une ou deux des associations en fonction du côté sur lequel elles latéralisent, bloquant temporairement l'apprentissage de l'autre/des autres

association(s). Cette entrée en latéralisation semble signifier que la souris a été en mesure de détecter le changement de règle, probablement au cours de la phase de persévération, et que la latéralisation peut consister en une stratégie comportementale d'exploration dirigée et un processus d'ajustement des variables d'apprentissage.

Ces analyses comportementales et hypothèses computationnelles ayant été établies, l'investigation des bases neurobiologiques de celles-ci consistait en la seconde partie de l'étude. Dans cette perspective, nous avons décidé de cibler chez la souris le cortex préfrontal dorsomédian (dmPFC) appelé aussi cortex cingulaire antérieur dorsal (dAAC) ou cortex cingulaire médian (MCC). Cette région est en effet un point d'intégration cérébral majeur permettant à la fois :

- de supporter les processus d'apprentissage (Holroyd and Coles, 2008; Holroyd and Yeung, 2011, 2012)
- de superviser les changements abrupts qui interviennent dans l'environnement (en étant sensible aux résultats inattendus (Hayden et al., 2011a; O'Reilly et al., 2013), en prenant en compte l'historique des récompenses (Kennerley et al., 2006), en estimant la volatilité de l'environnement (Behrens et al., 2007), en établissant un modèle interne de son environnement (Tervo et al., 2014))
- de contrôler le changement des stratégies comportementales (en ajustant dynamiquement les paramètres d'apprentissage (Khamassi et al., 2013), en contrôlant les transitions comportementales d'exploitation à exploration par exemple (Domenech et al., 2020; Durstewitz et al., 2010))

Afin d'investiguer le rôle de l'équivalent du dACC chez la souris dans le cadre de la tâche de *switch*, nous avons utilisé une technique de neuromodulation pharmacologique (Whissell et al., 2016) afin d'inhiber les neurones pyramidaux de cette région à la suite des changements de règles auxquels sont confrontés les animaux et d'en évaluer les conséquences comportementales.

Note de l'auteur :

*Ce troisième chapitre de travail expérimental consiste en la présentation d'un travail très préliminaire dont le bon déroulement a été entravé par le confinement des mois de mars à mai 2020 et par la perte de données liée à une panne informatique. Par ailleurs, toutes les histologies permettant de vérifier l'expression des récepteurs pharmacologiques n'ont pas encore été réalisées (seulement 2 sur 4). Les données sont à prendre ici avec beaucoup de recul et seront bien évidemment à compléter par la suite avec la réalisation des contrôles appropriés ainsi que l'augmentation du nombre de souris impliquées dans chaque condition.*

## 2. METHODES

### 2.1. Animaux

Un total de quatre souris mâles C57BL/6J de génotype sauvage a été impliqué dans ce protocole expérimental. À la suite de leur sevrage à vingt-et-un jours, elles ont grandi en groupe de quatre à six souris dans des cages ventilées individuellement, sous un rythme de lumière de douze heures (lumière de 8 heures du matin à 20 heures le soir) à une température de  $21^{\circ}\text{C} \pm 2^{\circ}\text{C}$  et  $50\% \pm 10\%$  d'humidité. Ces souris avaient un accès illimité à l'eau et l'alimentation, et avaient à leur disposition des carrés de coton pour construire leur nid. Les souris étaient incluses dans le protocole à partir de l'âge de quatre mois. Après l'étape de chirurgie (voir 2.2) et à partir du début du protocole d'expérimentation comportemental (voir 2.4), les souris étaient isolées dans des chambres opérantes comportementales. Durant la période des expériences, elles n'avaient plus accès à de la nourriture à volonté puisque leur prise alimentaire était conditionnée à leur réussite dans la tâche. A chaque bonne réponse, elles obtenaient une tablette alimentaire (20 mg AIN-76 rodent tablets, Test Diet). Les souris n'étaient en revanche pas restreintes d'un point de vue hydrique. Les animaux ont été manipulés par la même expérimentatrice tout au long du protocole. Les souris ont été pesées régulièrement : toute perte de masse en-deçà de 80 % était compensée par ajout de tablettes. Leur état général de santé était régulièrement évalué également : posture et faciès de l'animal, état de la fourrure, turgescence de la peau, activité générale évaluée par la quantité d'essais réalisés au cours des dernières vingt-quatre heures. Tout signe de détresse aurait entraîné l'exclusion du protocole de l'animal et les soins de celui-ci (Spangenberg and Keeling, 2016). L'élevage ainsi que les expériences ont été menés sur la plateforme PHENO-ICMice de l'Institut du Cerveau. Toutes les expériences ont été menées en accord avec les lois de bioéthiques françaises et européennes.

### 2.2. Chirurgie

#### 2.2.1. Justification de l'aire ciblée

Dans la perspective de notre étude et de la littérature, nous avons voulu évaluer les conséquences de l'inhibition du dACC sur le comportement à la suite des changements abrupts et non indicés des contingences expérimentales. Travaillant avec des souris, il a donc été nécessaire de cibler l'aire homologue du dACC chez les rongeurs. Suivant des arguments cytoarchitecturaux (Vogt and Paxinos, 2014), de connectivité (Heilbronner et al., 2016), et évolutifs (Laubach et al., 2018), le choix de cibler le cingulaire 1 et 2 en amont de Bregma a

été fait. Suivant le Paxinos 5<sup>ème</sup> édition (Paxinos and Franklin, 2019), c'est donc l'aire 24 qui a été ciblée.

### 2.2.2. Procédure chirurgicale

Vingt minutes avant la chirurgie les souris ont reçu une injection intrapéritonéale de buprenorphine (Buprecare ®, MedVet, 0,3 mg / mL diluée à 0,1 mg / mL) en analgésie pré-opératoire en respectant la dose de 0,1 mg / kg. L'anesthésie a été réalisée sous isoflurane (Iso-Vet ®, Laboratoire Osalia) à 4 % (flux 325 mL/min d'air) pour la période d'endormissement en chambre d'induction puis la concentration a été descendue progressivement à 2 % en cours de chirurgie (flux 180 mL/min d'air). Une fois les souris endormies, elles ont été placées sur l'appareil stéréotaxique grâce à des barres d'oreilles et un repose-dents. Leurs yeux ont été tout de suite protégés avec du gel oculaire (Lubrithal, Dechra). Tout au long de la chirurgie, les souris étaient sur un tapis chauffant réglé à 37°C grâce à une sonde rectale afin d'éviter toute hypothermie.

Une fois les poils retirés, la peau désinfectée à la bétadine (Vétédine, Med'Vet) et l'incision au scalpel réalisée après dépôt de lidocaïne (10 mg/kg - Xylovet® or Laocaine®) une craniotomie a été réalisée au foret de manière bilatérale suivant les zones d'injection souhaitée (+ 0,98 mm antérieur à Bregma, ± 0,38 mm latéralement à la ligne médiane). La seringue (Hamilton®, 1700-series, 10 µL) a alors été descendue dans chaque hémisphère au rythme de 100 µm toutes les 20 s jusqu'à – 1,27 mm par rapport à la surface du cerveau. Avant et après injection du virus à une vitesse de 0,05 µL / min (**voir 2.3 Virus**) un délai de cinq minutes a été respecté. 0,1 µL de virus a été injecté dans chacun des hémisphères. La seringue était ensuite remontée au rythme de 100 µm toutes les vingt secondes. La peau a été suturée avec du fil de suture (Ethicon, soie 3-0) et la plaie désinfectée (Vétédine, Med'Vet diluée à 10 %).

Les souris ont finalement reçu une injection sous-cutanée bilatérale de chlorure de sodium à 0,9 % à 37°C afin de les réhydrater (1 mL / 100 g, double injection) et ont été placées en chambre de réveil thermostatée à 37°C. Une fois complètement réveillées et alertes, elles ont été placées en cage ventilée individuelle avec eau, nourriture à volonté, coton de nidification et une petite maison en plastique. En cas de signe apparent de douleur ou de mal-être, à J1 et J2 post-chirurgie, les souris ont également reçu une injection intrapéritonéale de buprenorphine (Buprecare ®, MedVet, 0,3 mg / mL diluée à 0,1 mg / mL) en respectant la dose de 0,1 mg / kg. Un minimum de 7 jours de repos post-chirurgie a été respecté avant de

débuter les expériences comportementales et l'état des animaux a été vérifié régulièrement (masse, aspect général).

### **2.3. Virus**

Afin d'inhiber la région d'intérêt, il était nécessaire de faire exprimer un récepteur à action inhibitrice. Pour ce faire, nous avons transfecté les neurones pyramidaux du cingulaire 1 et 2 par le virus pAAV-CaMKIIa-hM4D(Gi)-mCherry (Adeno Associated Virus, sérotype 5, 50477-AAV5 Addgene,  $4.4 \times 10^{12}$  vg/mL). Le récepteur muscarinique inhibiteur hM4D (récepteur métabotropique couplé à petite protéine Gi) étant placé sous le contrôle du promoteur CAMKII, il était ainsi uniquement exprimé dans les neurones pyramidaux de la région infectée. Les 4 souris ont été injectées avec ce virus. Aucune souris injectée avec le virus contrôle vide (pAAV-CaMKIIa-EGFP, 50469-AAV5, Addgene) n'a pu encore être réalisée.

### **2.4. Comportement**

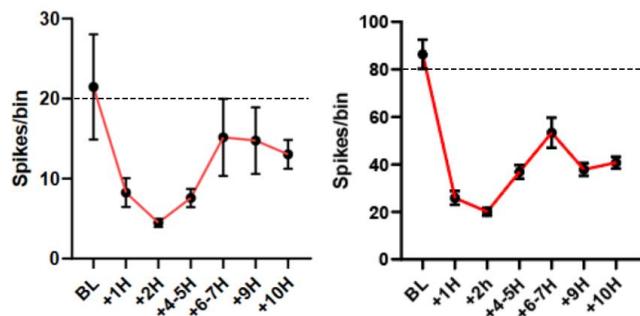
Les souris ont suivi le même protocole comportemental (habituation, entraînement et tâche de *switch*) qu'au chapitre 2 de la section « Travail expérimental » à l'exception du second switch qui n'était plus un *switch partiel* mais un *switch total*. Ce choix de séquence (2 *switches totaux* enchaînés) a été fait afin d'évaluer dans un premier temps l'effet de l'inhibition du cingulaire sur les comportements adaptatifs en cas de changement brutal et complet de l'environnement, les deux *switches* enchaînés pour un même animal servant de contrôle mutuel (voir 2.5. Neuromodulation inhibitrice).

## **2.5. Neuromodulation inhibitrice**

### **2.5.1. Choix du composé pharmacologique**

Du fait de la nouvelle ligne scientifique de l'équipe et de la réalisation de tests sur différents composés pharmacologiques employés dans le cadre de neuromodulation expérimentale (Schreiweis, Mondragon et al., *in prep*), le composé 21, agoniste des récepteurs hMD4 (Jendryka et al., 2019) a été utilisé sous sa forme hydrosoluble (DREADD agonist 21, *Compound 21*, dihydrochloride (water soluble), HB6124, HelloBio) et préparé par dissolution dans du chlorure de sodium à 0,9 % à la concentration de 0,5 mg / ml pour une utilisation à 5 mg / kg. Les travaux de Schreiweis et Mondragon chez la souris semblent en effet montrer qu'à cette concentration le composé 21 inhibe l'activité des neurones qui expriment le

récepteur muscarinique inhibiteur hM4D dès 1 heure après l'injection et pendant au moins 5 à 6 heures, voire peut-être plus (**Figure 1**). Ce composé permet donc d'obtenir un effet neuromodulateur sur une longue durée en une seule injection, ce qui peut être particulièrement intéressant dans le cadre de la tâche de *switch* puisque l'objectif est d'inhiber l'activité du cingulaire juste après le *switch* et de manière continue sur toute la période d'activité des animaux.



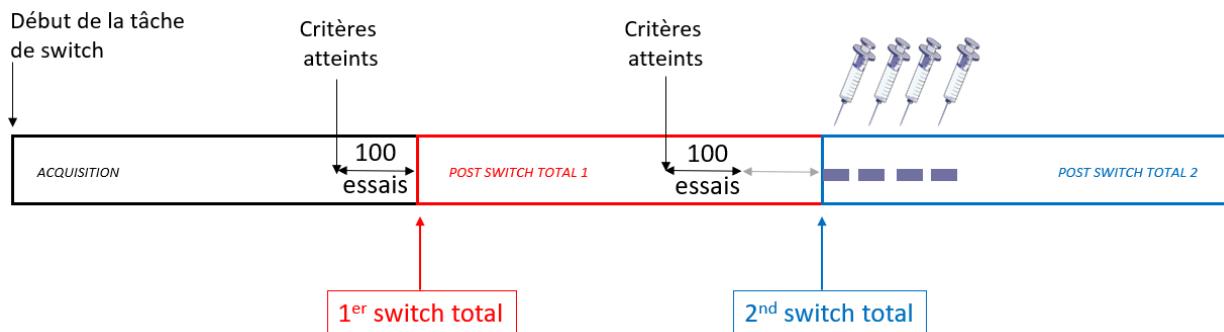
**Figure 1 : Effet du composé 21 sur l'activité neuronale.** Evolution de l'activité neuronale populationnelle de deux souris exprimant le récepteur muscarinique inhibiteur hM4D au niveau du cingulaire à la suite d'une injection de composé 21 (5 mg / kg) en comparaison avec l'activité basale sans injection (BL). Chaque graphe représente les données pour une souris. Chaque point représente la moyenne des enregistrements des différentes électrodes. Les barres d'erreurs représentent la déviation standard. L'activité est exprimée en nombre de décharges par unité de temps (un bin = 0,5 secondes). La neuromodulation inhibitrice grâce au composé 21 semble réduire de manière drastique l'activité des neurones jusqu'à 5 heures post-injection et peut-être plus en considérant les résultats de la deuxième souris. Figures incluses ici avec la permission du Dr Christiane Schreiweis et du Dr Lizbeth Mondragon (Schreiweis, Mondragon et al., *in prep*).

### 2.5.2. Temporalité des injections

Différentes périodes auraient pu être choisies au cours de la tâche, mais nous avons décidé dans un premier temps de nous concentrer sur la phase de persévération qui suit le changement de règles. Cette phase précède juste la phase d'adaptation comportementale et consiste en une période riche en informations pour les animaux puisque les souris sont confrontées aux nouveaux résultats associés à leurs actions après le changement de règles.

Dans le cadre de cette étude préliminaire, les souris ont reçu les injections du composé 21 lors du second *switch total*, le premier *switch total* servant donc de référence et de contrôle. Tout comme le protocole expérimental du chapitre 2, lorsque les souris atteignaient 80 % de performance globale sur les 40 derniers essais, ainsi que 80 % de performance pour les 40 dernières présentations de chaque stimulus, elles disposaient encore de 100 essais avec la règle en cours. Au bout de ces 100 essais, un message prévenait l'expérimentatrice de la nécessité d'injecter les animaux avec le composé 21. La première injection réalisée était concomitante au changement de règle. Les injections intrapéritonéales étaient réalisées entre

19 h 30 et 20 h le soir afin que le composé soit présent au cours des périodes d'activité de la souris durant la nuit (**Figure 2**).



**Figure 2 : Protocole de neuromodulation pharmacologique au cours de la tâche de switch.** Une injection de composé 21 (5 mg / kg) a été réalisée entre 19h30 et 20h chaque soir pendant quatre soirs après le second switch total.

## 2.6. Perfusion et post-perfusion

Une fois le protocole comportemental terminé, les souris ont été perfusées. À la suite d'une injection létale de pentobarbital (Euthasol, Med' Vet, 200mg/kg) permettant dans un premier temps la mise en coma profond et la perte de sensibilité, les souris ont été perfusées en intracardiaque grâce à une pompe (Gilson 1-channel peristaltic pump & tubing). 50 mL de NaCl à 0,9% et 4°C (élimination du sang) puis 50 mL de paraformaldéhyde à 4 % et 4°C (fixation des tissus) ont été utilisés. Une fois la décoloration des tissus et la rigidité du corps observées convenablement, les cerveaux ont été extraits et post-fixés dans du paraformaldéhyde à 4 % pendant 24 heures en chambre froide à 4°C. Une fois la post-fixation terminée, les cerveaux ont été déshydratés pendant 48 h dans une solution de saccharose à 30 % en chambre froide à 4°C. Une fois les cerveaux complètement déshydratés, ils ont été congelés dans du gel cryoprotecteur Tissue-Tek O.C.T dans l'isopentane à -45°C puis conservés à -80°C jusqu'au protocole d'immunohistochimie.

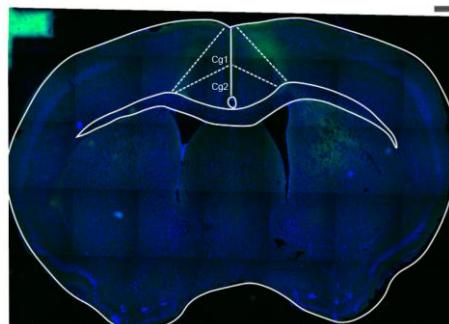
## 2.7. Immunohistochimie

Les cerveaux ont été coupés en sections coronales de 40 µm d'épaisseur au cryostat (Microm HM 560, Thermo Scientific) à -20°C. Les coupes ont été conservées dans du PBS-azide 0,1 % à 4°C jusqu'à l'immunohistochimie. Après trois lavages successifs de 10 minutes au PBS, 3 perméabilisations successives de 10 minutes dans du PBS - Triton 0.2% - Tween 0.1% (PBST), le blocage a été réalisé dans du PBST complémenté avec du sérum d'âne à 5% (PBSTA) pendant 2 heures à température ambiante. Les coupes ont été incubées avec

l'anticorps primaire anti-RFP (Anti-RFP lapin, Rockland antibodies and assays, dilution à 1:1000 dans du PBSTA) pendant une nuit à 4°C sous agitation. Suite à trois lavages successifs en PBST, les coupes ont été incubées pendant 2 heures à température ambiante, dans l'obscurité et sous agitation avec l'anticorps secondaire anti-anticorps de lapin (Anti-lapin A488, produit chez l'âne, Life Technologies, dilution à 1:400 dans du PBSTA). Puis, à la suite de trois lavages de 10 minutes (2 en PBST, puis 1 en PBS), les coupes ont été incubées à l'obscurité 2 minutes 30 avec du DAPI (10 µg/mL), rincées 3 fois en PBS et enfin montées sur lames (Thermo Scientific, Menzel Glaeser, Superfrost Plus). Après séchage, 200 µL de solution de montage (VECTASHIELD® Antifade Mounting Media) ont été déposés sur chaque lame permettant de sceller les lamelles. Les lames ont été conservées à 4°C, à l'abri de la lumière et imagées dans un délai maximum de 48 heures (voir 2.8 Microscopie).

## 2.8. Microscopie

Les images ont été acquises au microscope Axio Observer Apotome inversé Zeiss M4 (Ceilis), grossissement x10, en utilisant les filtres DAPI (light source intensity 20%) et eGFP (light source intensity 50%). Le logiciel employé était Zen (Zeiss microscopie).



**Figure 3 : Imagerie au microscope apotome d'une section coronale après immunohistochimie d'un cerveau d'une des souris ayant subi une transfection virale afin de faire exprimer le récepteur muscarinique inhibiteur hM4D dans les neurones pyramidaux du cingulaire murin. Localisation de la coupe : vers +0.98 antérieur à Bregma environ. Bleu : DAPI (noyaux cellulaires). Vert : marquage contre mCherry (permet de localiser les neurones ayant exprimé le récepteur hM4D). Dans cet exemple, le marquage semble essentiellement unilatéral et plus concentré vers le cingulaire 1.**

## 2.9. Analyses et statistiques

Les analyses des données comportementales et les statistiques ont été réalisées grâce à des scripts développés par mes soins sous Matlab (Matlab 2017b, MathWorks). Les analyses réalisées ont été exactement les mêmes que pour la première version d'article développées au chapitre 2 de la section « Travail expérimental ».

### 3. RESULTATS COMPORTEMENTAUX PRELIMINAIRES

#### 3.1. Comportement général dans la tâche de *switch*

Les souris ont présenté un comportement général analogue à celui des souris de l'étude précédente (tâche de *switch* sans neuromodulation pharmacologique). Elles ont réalisé en moyenne  $8\ 219 \pm 1\ 443$  essais en  $38 \pm 7$  jours après avoir été entraînées pendant  $155 \pm 70$  essais sur  $2 \pm 1$  jours. Les quatre souris ont également réalisé plus d'essais les nuits que les jours : un total  $10\ 814 \pm 2\ 840$  essais durant les nuits contre un total de  $1\ 155 \pm 557$  essais durant les jours. D'un point de vue qualitatif, on retrouve également les mêmes tendances lorsque l'on compare les essais de jour par rapport aux essais de nuit. Durant les journées, périodes moins actives des souris, on constate une plus grande proportion d'essais non répondus ( $6,7 \pm 0,5$  % contre  $2,1 \pm 0,6$  %), plus de récompenses non récupérées ( $21,7 \pm 0,2$  % contre  $3,2 \pm 0,2$  %), plus d'essais correctifs pour cause de latéralisation spatiale ( $11,5 \pm 2,7$  % contre  $7,4 \pm 0,8$  %). Tout comme l'étude précédente, les essais de jour ont été filtrés et non conservés pour les analyses principales.

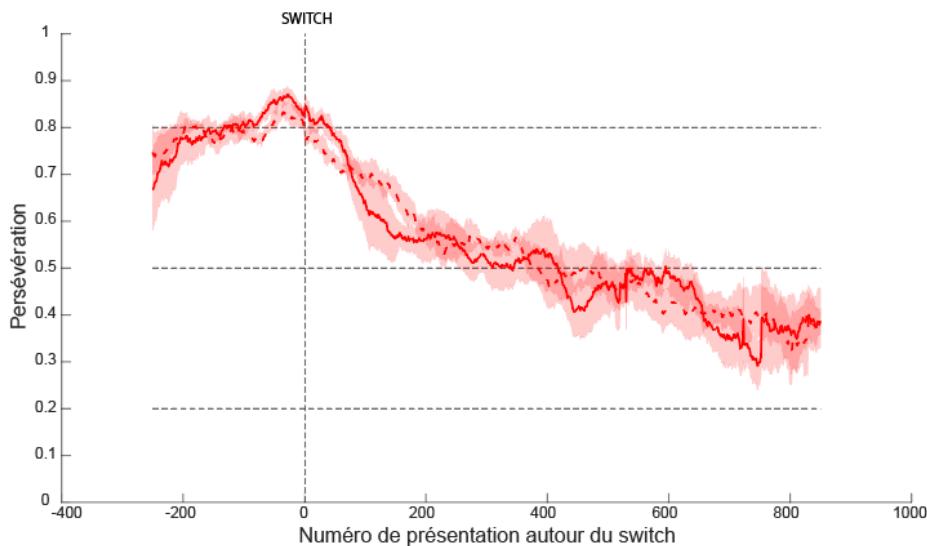
À la suite des injections de composé 21 durant les quatre soirs d'injection, aucune baisse d'activité n'a été constatée sur les nuits suivantes en comparaison au premier *switch total* pour lequel aucune injection n'a été réalisée ( $769 \pm 227$  essais durant les quatre nuits sans injection contre  $801 \pm 244$  pour les quatre nuits avec injections). Le composé 21 ne semble donc pas avoir eu d'effet sédatif sur les animaux.

Pour la suite des analyses et des graphes présentés, il est absolument essentiel de garder en tête que les souris ayant réalisé en moyenne 800 essais durant cette période et que la tâche ayant été conçue pour qu'il y ait autant de réponses données à gauche qu'à droite, cela signifie que ces 800 essais représentent environ 400 présentations du stimulus seul de son côté et 200 présentations de chacune des deux autres figures.

#### 3.2. Persévération autour du *switch*

Lors de l'étude précédente, l'analyse de la persévération, c'est-à-dire la propension de l'animal à répondre suivant les mêmes règles qu'avant le *switch* a révélé que l'adaptation des souris aux changements brusques reposait notamment sur une stratégie de méta-apprentissage par renforcement. Toute la question est d'investiguer si dans la suite de cette étude, l'inhibition du cortex cingulaire a un effet sur cette stratégie suivant le changement de contingences.

Le comportement de persévération a donc été évalué autour du premier *switch total* sans neuromodulation pharmacologique et autour du second *switch total avec* neuromodulation pharmacologique inhibitrice (**Figure 4**). Si l'on considère l'ensemble des stimuli avec leur performance moyennée, les deux courbes semblent globalement se superposer. L'observation seule de ces courbes ainsi moyennées ne permet donc pas de conclure à un effet potentiel de l'inhibition du cortex cingulaire.

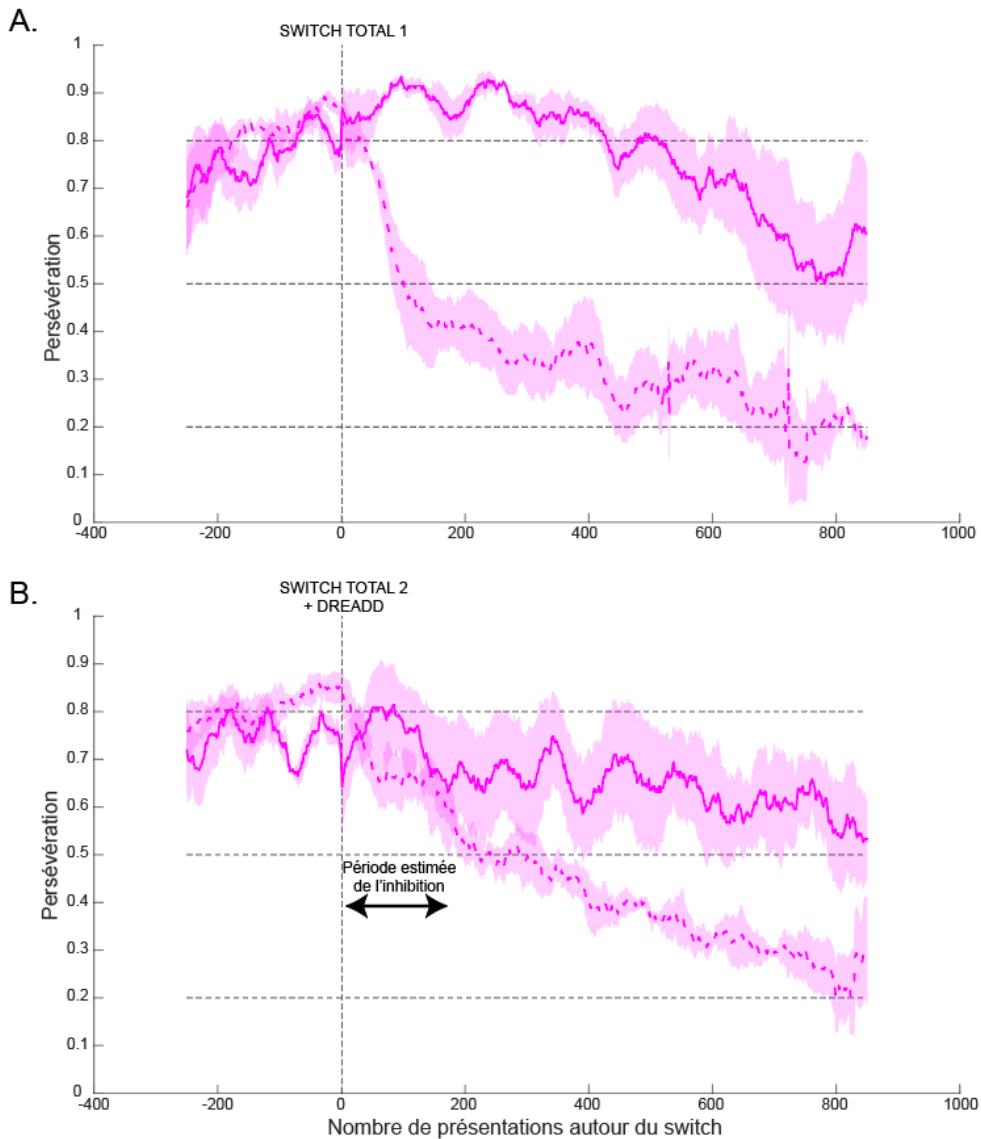


**Figure 4 : Evolution de la persévération autour des switches totaux.** Courbe rouge pleine : persévération autour du *switch total* sans neuromodulation. Courbe rouge en pointillés : persévération autour du *switch total avec* neuromodulation inhibitrice.

Comme montré durant l'étude précédente en revanche, lorsque l'on réalise l'analyse de la persévération autour des *switches* en dissociant les deux stimuli associés du même côté de réponse (**Figure 5A et B courbes roses en pointillés**), du stimulus seul de son côté de réponse (**Figure 5A et B courbes roses pleines**), des tendances semblent émerger.

Pour le premier *switch total*, les mêmes tendances que pour l'étude précédente sont observables (**Figure 5A**). Un plateau de persévération très court pour les stimuli associés du même côté de réponse, une chute abrupte de la persévération pour atteindre le niveau de chance autour de 175 présentations après le *switch* puis un ralentissement de la vitesse de persévération jusqu'à un réapprentissage des associations autour de 800 présentations après le *switch*. Pour la figure seule de son côté de réponse, on observe à l'inverse un très long plateau de persévération, donc un blocage de l'apprentissage de ce stimulus pendant environ 400 présentations, puis une diminution lente de la persévération. Après 800 présentations de cette figure, les animaux persévérent encore à hauteur de 50 % pour ce stimulus visuel. Ces résultats sont totalement cohérents avec ceux observés précédemment sur un échantillon de

dix-sept souris et semblent donc indiquer : (1) que le taux d'apprentissage est variable, suggérant ainsi bien l'existence d'un phénomène de méta-apprentissage, (2) que la variation du taux d'apprentissage dépend notamment de la manière dont les stimuli sont associés entre eux pour leur côté de réponse.



**Figure 5 : Evolution de la persévération autour des switches totaux.** **A.** Autour du *switch total* sans neuromodulation inhibitrice du cingulaire. **B.** Autour du *switch total* avec neuromodulation inhibitrice du cingulaire. Courbe rose en pointillés : persévération pour les figures associées au même côté de réponse. Courbe rose pleine : persévération pour la figure seule de son côté de réponse.

Pour le second *switch total* avec neuromodulation inhibitrice du cingulaire, on observe en revanche des dynamiques bien différentes (**Figure 5B**). Pour les stimuli associés du même côté de réponse, il ne semble pas y avoir de plateau de persévération, et la diminution rapide de la persévération jusqu'à 50 % n'est plus visible : il semblerait dans ce cas que la vitesse de réapprentissage soit constante. Pour le stimulus seul de son côté de réponse, à l'inverse, le

plateau de persévération ne semble quasiment pas observable et la diminution de la persévération ne semble pas être véritablement distincte de celle des stimuli associés du même côté de réponse durant la période d'inhibition estimée du cingulaire. En revanche, après cette période, la vitesse de diminution de la persévération semble moins rapide que celle pour les stimuli associés du même côté. En effet, après 800 présentations du stimulus seul de son côté de réponse, la persévération est encore de 50 % pour celui-ci, là où elle n'est plus que de 20 % pour les deux autres stimuli associés du même côté de réponse.

Ces résultats sembleraient donc indiquer que l'inhibition du cingulaire juste après le *switch* abolit le blocage de l'apprentissage pour le stimulus seul de son côté de réponse ainsi que la variation dynamique du taux d'apprentissage pour les stimuli associés du même côté de réponse. Néanmoins, après cette période, le taux d'apprentissage semble différer de nouveau entre les catégories de stimuli dépendamment de leur côté de réponse : les souris semblent de nouveau résoudre la tâche dans l'espace des actions avec des taux d'apprentissage différents.

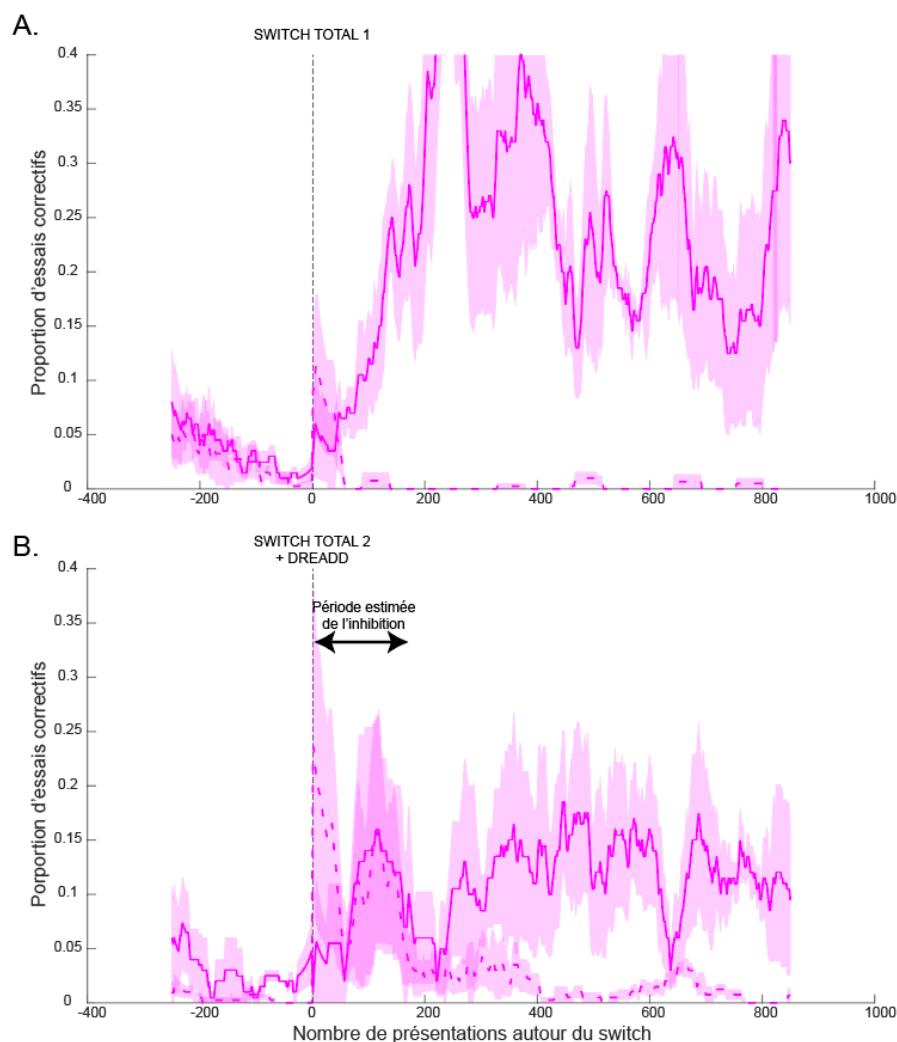
### 3.4. Estimation de la latéralisation autour du *switch*

Si les stratégies d'adaptation de la souris sont liées à l'espace des actions, il est donc possible d'envisager l'observation de motifs de réponses particuliers, comme de la latéralisation. Dans le cadre de la tâche de *switch*, la latéralisation peut s'estimer grâce à l'étude de la proportion d'essais correctifs. Les essais correctifs sont en effet déclenchés lorsqu'une séquence de latéralisation est détectée au cours de la tâche. Si la souris répond cinq fois du même côté avec une performance de moins de 50 %, lors des essais suivants, ne seront présentées que la ou les figures dont le côté de réponse est opposé au côté de latéralisation observé jusqu'à présent. Cette ou ces figures sont présentées tant que la souris n'a pas répondu correctement, c'est-à-dire tant qu'elle n'a pas changé son côté de réponse. Les essais correctifs sont donc bien des indicateurs du comportement de latéralisation de l'animal. L'analyse de ces essais correctifs a été réalisée autour des *switches* en dissociant encore une fois les deux stimuli associés du même côté de réponse (**Figure 6A et B courbes roses en pointillés**), du stimulus seul de son côté de réponse (**Figure 6A et B courbes roses pleines**).

Après le premier *switch total* sans neuromodulation inhibitrice, les essais correctifs sont déclenchés presque totalement pour la figure visuelle seule de son côté : cela signifie donc que les souris latéralisent bien du côté associé aux deux autres figures visuelles (**Figure 6A, courbe rose pleine**). En revanche, après le second *switch total*, durant la période estimée de la neuromodulation inhibitrice, on constate une proportion aussi élevée d'essais correctifs pour le stimulus seul de son côté de réponse que pour les deux autres stimuli associés au

même côté de réponse : les souris ne répondent donc pas de manière privilégiée et continue sur un côté particulier comme lors du premier *switch total* (**Figure 6B**). En revanche, à la fin de cette période de neuromodulation inhibitrice du cingulaire, les essais correctifs redeviennent majoritaires pour la figure seule de son côté, et cet effet est potentiellement lié à une reprise de comportement de latéralisation des souris du côté associé à deux stimuli.

L'ensemble de ces résultats pourraient donc suggérer que l'inhibition du cortex cingulaire murin contribue à une réduction du comportement de latéralisation préférentiel à un côté de réponse de la souris et à la résolution de la tâche dans l'espace des actions. Cette réduction s'observe conjointement au blocage du phénomène de méta-apprentissage.



**Figure 6 : Evolution de la proportion d'essais correctifs autour des *switches totaux* comme évaluation approximative de la latéralisation.** **A.** Autour du *switch total* sans neuromodulation inhibitrice du cingulaire. **B.** Autour du *switch total* avec neuromodulation inhibitrice du cingulaire. Courbe rose en pointillés : proportion d'essais correctifs pour les figures associées au même côté de réponse. Courbe rose pleine : proportion d'essais correctifs pour la figure seule de son côté de réponse.

#### 4. DISCUSSION

Au cours de cette étude préliminaire, nous avons montré que la neuromodulation inhibitrice du cingulaire chez la souris altère la stratégie d'adaptation à la suite d'un changement complet des contingences expérimentales. L'inhibition de cette région semble notamment abolir les variations du taux d'apprentissage ainsi que l'heuristique d'exploration fondée sur l'action. Il est à noter que durant la période d'inhibition du cingulaire, l'apprentissage n'est pas aboli mais celui-ci est désormais homogène entre tous les stimuli. A l'arrêt de la phase d'inhibition, l'asymétrie d'apprentissage semble en revanche réapparaître, suggérant de nouveau l'existence d'un taux d'apprentissage variable en fonction de la manière dont les stimuli sont associés entre eux. Ces résultats préliminaires peuvent donc suggérer que le dACC murin est impliqué dans le phénomène de méta-apprentissage par renforcement et d'ajustement des paramètres d'apprentissage.

Notamment, le fait que l'inhibition du dACC supprime la diminution rapide de la persévération pour les stimuli associés du même côté peut laisser penser que le dACC réajuste abruptement le taux d'apprentissage lorsque les changements de règles sont détectés. Ceci fait écho à certaines études qui semblent en effet suggérer que l'ajustement dynamique des paramètres d'apprentissage par le dACC permet de s'adapter de manière optimale à un changement brutal de l'environnement par une réinitialisation abrupte de certains métaparamètres (Khamassi et al., 2013; McGuire et al., 2014). Néanmoins, dans le cadre de cette tâche, cet effet n'est visible seulement pour les stimuli associés du même côté alors que l'apprentissage du stimulus seul de son côté de réponse reste bloqué. Cette observation est peut-être liée à des capacités cognitives de la souris plus limitées en termes de fonctions exécutives, notamment en termes de ressources attentionnelles, (Rmus et al., 2021), ne lui permettant probablement pas de considérer le problème dans son ensemble, nécessitant une résolution plus séquentielle de la tâche et une allocation dirigée de ses ressources cognitives.

Afin de creuser cet aspect, il aurait été pertinent et intéressant de prolonger l'inhibition du cingulaire murin au-delà de quatre soirs afin de voir si la courbe de persévération pour le stimulus seul de son côté évoluait de manière analogue et confondue avec celle pour les deux stimuli associés du même côté de réponse, suggérant ainsi une exploration des nouvelles associations ainsi qu'un réapprentissage identique pour l'ensemble des stimuli. Le fait que l'inhibition du dACC supprime l'heuristique d'exploration basée sur l'action pourrait suggérer que le dACC permet également de maintenir une représentation de la tâche dans l'espace des actions. Néanmoins, il aurait également été pertinent d'analyser comment une

inhibition prolongée du dACC affecte la courbe de réapprentissage : la persévération atteint-elle de nouveau 20 % au bout de 800 présentations de chacun des stimuli, ou stagne-t-elle autour de 50 % ? Si l'inhibition prolongée du cingulaire murin maintenait en effet les souris au niveau de la chance, ceci laisserait penser que le dACC permet en réalité de gérer les résultats associés aux actions réalisées par la souris et de prioriser le réapprentissage rapide de certaines associations en fonction des ressources cognitives dont elle dispose. De tels résultats pourraient faire écho aux travaux de Tervo et collaborateurs qui montrent que l'inhibition du cortex cingulaire chez le rat induit un comportement des animaux plus stochastique et bien moins dépendant de l'historique des résultats des derniers choix au cours de la tâche (Tervo et al., 2014).

Ces premiers points ainsi investigués, il serait par la suite très intéressant de raffiner la méthodologie de neuromodulation. En choisissant une autre fenêtre temporelle d'inhibition du cingulaire par exemple. L'inhibition de cette aire pourrait notamment être évaluée à la fin des plateaux de persévération. Ceci permettrait de vérifier si le cortex cingulaire est encore impliqué à ce moment ou si la phase critique d'adaptation comportementale médiée par cette région se joue bien durant de la phase de persévération. Néanmoins, une telle ambition nécessite de développer un algorithme de détection de ces périodes (persévération / latéralisation) en même temps que la tâche est effectuée par les souris, ce qui peut poser des contraintes techniques et algorithmiques non négligeables.

Le choix d'une autre technique de neuromodulation expérimentale pourrait également apporter des compléments de réponse aux questions posées. L'utilisation de l'optogénétique par exemple permettait de raffiner les périodes d'inhibition de la zone d'intérêt. Il serait par exemple envisageable d'inhiber l'activité du cingulaire au moment de l'observation des résultats associés aux actions réalisées par les souris, en inhibant soit au moment d'un résultat positif, soit au moment d'un résultat négatif, soit lors des deux types de résultats. L'inhibition du cortex cingulaire à ces moments perturberait donc uniquement l'intégration de ces résultats et il serait ainsi possible d'en évaluer les conséquences sur le comportement. Il a notamment été montré que l'intégration des résultats obtenus dans le passé est essentiel aux prises de décision futures (Hermoso-Mendizabal et al., 2020).

Néanmoins, cette technique reste encore assez sensible à implémenter dans notre système comportemental où les souris sont laissées en autonomie vingt-quatre heures sur vingt-quatre, sept jours sur sept. Avec les systèmes filaires en effet, les animaux ont sans cesse un fil relié à leur implant. Laisser les animaux sans surveillance pendant la nuit avec de telles conditions expérimentales peut être assez sensible et requiert encore quelques développements

techniques. Il faudrait sinon se diriger vers l'utilisation de systèmes sans fil afin de respecter au mieux le comportement, la physiologie des animaux tout en limitant leur stress (Montgomery et al., 2015).

Une fois le rôle du cingulaire défini dans les réponses comportementales ainsi décrites dans l'étude de *switch*, il pourrait être également très intéressant d'évaluer expérimentalement l'implication de neuromodulateurs sur le comportement décrit (Katz and Calin-Jageman, 2008). Si les résultats préliminaires mis en lumière ici venaient à se confirmer, l'implication de la noradrénaline pourrait être évaluée puisqu'il semblerait que ce neuromodulateur puisse induire des transitions comportementales, en s'affranchissant notamment de l'utilisation d'un modèle interne. Dans le cadre de notre tâche, il pourrait être pertinent de voir en quoi la neuromodulation noradrénergique permettrait d'induire directement un comportement d'exploration non latéralisé pour les souris en post-switch, sans phénomène de blocage d'apprentissage et de variation différentielle des taux d'apprentissage en fonction de la catégorie de stimuli.

Cette étude préliminaire apporte donc des résultats très encourageants et de multiples perspectives mais il est nécessaire de garder en tête quelques points importants à investiguer lors des prochaines expérimentations.

Notamment, l'expression des récepteurs au composé 21 n'a pas été vérifiée pour tous les animaux. Ainsi, il est pour l'instant impossible de savoir si les récepteurs ont bien été exprimés dans les zones d'intérêt. L'effet visible ici est peut-être lié à l'expression des récepteurs dans une autre zone que celle ciblée : la seule histologie réalisée montre notamment que le récepteur semble exprimé au niveau du cingulaire 1 et débordant sur les aires motrices (**Figure 3**). Il n'est donc pas possible de conclure quant à la stricte implication causale du cingulaire seul dans les résultats observés.

Par ailleurs, quelques contrôles seront à réaliser par la suite. Aucune souris n'a en effet suivi le protocole en effectuant deux *switches totaux* enchaînés sans neuromodulation pharmacologique. Or on sait notamment que lorsque des animaux enchaînent de multiples renversements de contingences les uns à la suite des autres, ils sont plus rapides à atteindre les critères (Bublitz et al., 2017; Castañé et al., 2010; Jarvers et al., 2016). Il existe donc peut-être un effet des séquences de renversement qu'on ne peut donc pas dissocier ici du potentiel effet de la neuromodulation pharmacologique. Néanmoins, cette hypothèse ne peut réellement expliquer la coïncidence temporelle de la réapparition de l'asymétrie d'apprentissage et de l'heuristique d'exploration par latéralisation avec la dissipation des effets du composé 21 (**Figure 5B et 6B**).

De même, aucune souris dont le cingulaire a été transfecté par un virus contrôle (vecteur vide sans le gène codant pour le récepteur inhibiteur muscarinique) et effectuant deux *switches totaux* enchaînés avec injection du composé 21 lors du second *switch* n'a pu encore être impliquée dans le protocole. De ce fait, les résultats observés peuvent potentiellement être attribués à un effet non spécifique du composé 21. Il ne serait effectivement pas prudent de considérer ce composé comme biologiquement inerte. Même si certaines études ne semblent pas montrer d'effet de ce composé sur le comportement locomoteur, exploratoire ou anxieux des animaux (Jendryka et al., 2019; Tran et al., 2020), il est à noter que les doses maximales utilisées sont de 1 mg / kg. Or dans une étude chez le rat, cette même dose de 1 mg / kg (soit une concentration cinq fois moins élevée que celle utilisée ici chez la souris), provoque une augmentation significative de l'activité des neurones dopaminergiques de la substance noire. Or la dopamine est un neuromodulateur essentiel dans les processus d'apprentissage et largement impliqué dans les processus de neuromodulation associés à l'adaptation comportementale (Cinotti et al., 2019; Koralek and Costa, 2019). Il est donc indispensable de réaliser ce contrôle par la suite afin de valider (ou non) l'effet spécifique de l'inhibition du cingulaire sur le comportement même s'il semblerait qu'en moyenne, la neuromodulation inhibitrice n'ait pas d'effet sur l'apprentissage global (**Figure 4**). Il serait par ailleurs également approprié de tester d'autres doses de ce composé et d'estimer la concentration minimale nécessaire permettant d'induire des conséquences comportementales chez les souris exprimant le récepteur muscarinique inhibiteur sensible à ce composé.

D'un point de vue pratique, il pourrait être également judicieux d'utiliser d'autres voies d'administration du composé 21 plutôt que des injections intrapéritonéales. Ces injections peuvent en effet stresser les animaux (Balcombe et al., 2004) et donc avoir des effets non négligeables sur les comportements observés par la suite. Néanmoins, l'administration du composé par oral dans l'eau de boisson reste peu maîtrisable car il est impossible de quantifier avec précision la quantité ingérée par les animaux ni de contrôler la temporalité de l'ingestion. L'utilisation de pompes osmotiques (Rodriguez et al., 2020) ou encore de plateformes sans fil d'injection pharmacologique (Zhang et al., 2019) pourraient consister en des alternatives intéressantes.



# **CONCLUSIONS, PERSPECTIVES ET DISCUSSION GENERALE**

---

## **INTRODUCTION**

A l'issue de mes quatre années de thèse, j'ai été en mesure de répondre aux différents objectifs définis au début de celle-ci.

Afin de comprendre les stratégies d'adaptation de la souris en environnement volatil, j'ai réussi à développer une toute nouvelle tâche comportementale permettant de désambiguïser les mécanismes computationnels d'adaptation parmi différentes hypothèses (apprentissage par renforcement simple, méta-apprentissage, task-set...). J'ai pu mettre en évidence des résultats originaux suggérant que l'adaptation aux changements abrupts de l'environnement chez la souris se fait certainement par le biais de méta-apprentissage et de stratégies d'exploration dirigée par latéralisation. Ces résultats ont des répercussions importantes quant à la considération du modèle souris pour étudier les processus comportementaux d'adaptation ainsi que leurs bases neurobiologiques.

Dans cette perspective, j'ai pu commencer à mettre en place un protocole de neuromodulation pharmacogénétique chez la souris lorsque celle-ci réalisait la tâche nouvellement développée. Les résultats préliminaires encourageants semblent montrer que le cortex cingulaire antérieur dorsal pourrait être impliqué dans l'ajustement des paramètres d'apprentissage à la suite de changements abrupts de contingences. Ces résultats seront à compléter par des expériences ultérieures.

L'ensemble de ces expériences a été rendu possible grâce à l'utilisation de chambres opérantes automatisées permettant l'acquisition de milliers de données. Dans la continuité des travaux initiés avant mon arrivée dans le laboratoire, j'ai développé un nouveau système de chambre opérante automatisée, open-source, utilisant des cartes électroniques programmables Arduino. J'ai pu valider ce système chez la souris et montrer que celui-ci permet de collecter des milliers de données sur des durées très longues d'expérimentation tout en assurant un bien-être optimal pour les animaux et en réduisant le temps de travail pour les expérimentateurs. J'ai également participé à l'adaptation hardware et software de ce système expérimental pour deux autres doctorantes de l'équipe. La BEATBox est aujourd'hui utilisée de manière quotidienne par les membres de l'équipe et permet l'implémentation de tâches complètement différentes de celles que j'ai pu valider, tout en enregistrant l'activité cérébrale

des souris par photométrie. La production à grande échelle de la BEATBox est aujourd’hui considérée afin de répondre aux besoins de l’équipe mais également de n’importe quel autre groupe de recherche dans le monde qui souhaiterait utiliser ce système.

Mon travail de thèse, de par ses fondements méthodologiques (développement et utilisation de chambres opérantes automatisées à haut rendement de données expérimentales), théoriques (développement d’une nouvelle tâche comportementale murine pour investiguer des stratégies comportementales d’adaptation et les mécanismes computationnels et neurobiologiques mis en jeu) et pratiques (mise en œuvre expérimentale et analytique), a ainsi tâché de répondre aux enjeux auxquels fait face la recherche aujourd’hui pour comprendre le cerveau et son fonctionnement (Griffiths, 2015; Niv, 2020). Mes résultats apportent des conclusions nouvelles à un ensemble de connaissances fascinant, mais encore controversé.

Dans cette perspective, il est critique aujourd’hui de comprendre les disparités observées entre les études et d’y apporter des solutions. Je souhaite désormais développer différents aspects de cette problématique. Ayant déjà discuté indépendamment des résultats, limites et perspectives de mon sujet au cours de l’état de l’art ainsi que lors de chacune de mes trois études expérimentales, je souhaite réaliser ici une discussion sur des thématiques transverses à mes travaux. Par ailleurs, ma thèse s’est jusqu’à présent ancrée dans une perspective de recherche essentiellement liée aux champs disciplinaires de la psychologie expérimentale et des neurosciences computationnelles. J’aimerais désormais discuter de mes résultats en adoptant une perspective plus liée à l’écologie comportementale et à l’éthologie cognitive car ces disciplines permettent de porter un regard critique sur les questions de l’apprentissage ou de l’adaptation, en relation avec des contraintes écologiques et des dynamiques évolutives.

## **1. « DES SOURIS ET DES HOMMES »<sup>4</sup> OU LA QUESTION DE LA RECHERCHE TRANSLATIONNELLE**

### **1.1. LE CHOIX DE LA SOURIS COMME MODELE D’ETUDE**

L’ensemble des travaux réalisés au cours de ces quatre années a été mené chez le modèle animal le plus couramment employé en recherche, à savoir la souris (Ministère de l’Enseignement Supérieur et de la Recherche, 2021). Il s’agit d’un animal de petite taille, se

---

<sup>4</sup> Titre inspiré de l’œuvre de John Steinbeck, *Of mice and men (Des souris et des hommes)*, 1937

reproduisant facilement, facile à élever, génétiquement modifiable, qui partage des similarités au niveau génétique avec l'homme (Chinwalla et al., 2002) ainsi que de nombreuses similitudes comportementales (van Meer and Raber, 2005). La souris constitue donc un excellent modèle animal pour explorer diverses questions de recherche, comprendre de multiples processus physiologiques, comportementaux, cognitifs et pathologiques afin d'en transposer les résultats à l'être humain dans le but de le caractériser biologiquement et de le soigner. Néanmoins, même si de nombreux mécanismes semblent conservés entre souris et humains, comme la capacité d'apprendre par essais – erreurs grâce à un processus de neuromodulation dopaminergique (Wise, 2004), il est de l'ordre du truisme d'écrire qu'une souris ne sera jamais un humain (Burkhardt and Zlotnik, 2013).

## **1.2. SOURIS VERSUS HUMAINS : DES STRATEGIES DIFFERENTES POUR S'ADAPTER EN MILIEU VOLATIL**

La tâche de *switch* nouvellement développée dans le cadre de ma thèse montre que les souris s'adaptent à des changements brusques de contingences grâce à une variation du taux d'apprentissage et que cette stratégie de métaprogrammation dépend également d'une heuristique d'exploration fondée sur l'action, les souris présentant une tendance à latéraliser de manière significative après les renversements de contingences.

Dans le cadre d'une expérimentation pilote menée au cours de ma thèse chez 11 sujets volontaires (7 hommes, 5 femmes, âgés en moyenne de 26 ans), afin de tester la possibilité d'implémenter la tâche de *switch* chez les humains, des résultats très différents ont émergé. Ces 11 sujets ont réalisé la même tâche déterministe que les souris (3 associations stimulus – action à apprendre, renversements de contingences complets ou partiels) en recevant le minimum de consignes afin qu'ils ne soient pas biaisés sur la structure de la tâche. Les stimuli visuels étaient présentés sur un écran d'ordinateur et les sujets devaient appuyer sur la touche Q ou la touche M afin d'associer un bouton de réponse à une figure visuelle. Les bonnes réponses étaient indiquées par un rond vert, les mauvaises par un rond rouge.

Les sujets humains ont été bien plus rapides que les souris pour réaliser cette tâche. Là où les souris ont nécessité en moyenne 10 000 essais et plus d'un mois d'expérience pour réaliser deux *switchs*, les sujets humains n'ont eu besoin que d'une heure et vingt minutes pour effectuer 1 300 essais et 45 *switchs*. Par ailleurs, après les *switchs partiels*, chez l'humain, j'ai pu observer une chute significative de la persévération pour le stimulus dont le côté de réponse ne change pas, résultat prédit par un modèle utilisant des *task-sets* et non

observable chez la souris. Les sujets humains n'ont également pas présenté de comportement de latéralisation comme les souris. Enfin, une fois le protocole achevé, en questionnant les sujets sur leur compréhension de la tâche, ceux-ci ont été capables d'en expliquer la structure (un bouton de réponse toujours associé à une figure, l'autre à deux figures, des renversements successifs de contingences, partiels ou complets) et ont indiqué avoir utilisé leur compréhension de cette structure afin de résoudre la tâche. Par exemple, à la suite d'un *switch*, si les deux stimuli associés à un même bouton de réponse étaient présentés l'un à la suite de l'autre, les sujets savaient d'ores et déjà que le troisième stimulus serait associé à l'autre bouton. Les humains ont donc utilisé les informations contrefactuelles contenues dans cette tâche. Ces données sont évidemment à prendre avec beaucoup de recul et à compléter puisqu'il s'agit d'une étude pilote informelle mais ces résultats restent néanmoins cohérents avec les données actuelles de la littérature qui semblent indiquer l'utilisation de structures mentales comme les *task-sets* pour s'adapter efficacement à un nouvel environnement à la suite d'un changement brutal de celui-ci (Domenech et al., 2020).

L'ensemble de ces résultats montre donc que les souris ont des capacités d'apprentissage plus limitées que les humains dans ce type de design expérimental. De plus, leur stratégie d'adaptation aux changements brutaux des contingences expérimentales est moins efficace que celle des humains du fait de l'absence de représentation mentale de la tâche et de l'absence d'utilisation d'informations contrefactuelles. Il est néanmoins à noter que cette tâche est en réalité très facile à réaliser chez l'homme, probablement moins chez la souris, ce qui peut expliquer les différences observées.

Néanmoins, un point essentiel est à remarquer. L'utilisation de *task-set* identifié chez les humains ainsi que les propos rapportés quant à la manière dont les sujets ont résolu la tâche (en se servant notamment de l'asymétrie d'association des boutons de réponse aux stimuli) pourrait indiquer que les humains auraient raisonné dans l'espace des actions pour apprendre dans cette tâche et s'adapter efficacement aux *switchs*. Chez la souris, nous avons mis en évidence une heuristique d'exploration fondée sur de la latéralisation ainsi qu'une variation du taux d'apprentissage liée à l'espace des actions et non des états. Ainsi, chez l'homme comme la souris, la résolution de la tâche en apprenant dans l'espace des actions, réduisant ainsi l'espace du problème, pourrait constituer une simplification efficace et une stratégie en soi pour apprendre et s'adapter. Des études récentes semblent d'ailleurs suggérer que l'espace des actions est effectivement une dimension à part entière dans le cadre de l'apprentissage par renforcement (Rmus et al., 2021). L'ensemble de ce travail doit être complété par des travaux de modélisation afin de soutenir les hypothèses avancées et de

proposer des modèles génératifs dont la simulation de données serait capable de reproduire les données expérimentales (Wilson and Collins, 2019). Ces recherches sont actuellement menées par Sami BEAUMONT dans le cadre de sa thèse.

Dans leur ensemble, les résultats ainsi obtenus sur le modèle murin C57BL6/J dans ce paradigme expérimental particulier ne peuvent se transposer à l'humain, soulevant donc des questions quant à l'utilisation du modèle souris pour étudier ce type de mécanismes dans la perspective des questions qui se posent pour l'Homme. Notamment, si l'on cherche à étudier en détail les substrats neuronaux impliqués dans les mécanismes computationnels d'adaptation chez l'Homme, la souris ne semble pas le meilleur modèle dans ce cas précis. En revanche, si la question est de comprendre l'évolution des capacités de prise de décision au travers des espèces, une telle étude fait sens dans la mesure où celle-ci est complétée par une approche trans-espèces.

La question ici n'est donc clairement pas de remettre en cause l'utilisation des modèles animaux en recherche car des années d'expériences et de preuves scientifiques en ont démontré tout l'intérêt. Tout l'enjeu consiste en revanche à se questionner sur les limites des protocoles expérimentaux employés et de leurs objectifs afin de proposer des solutions garantissant une meilleure validité, reproductibilité, fiabilité et surtout une meilleure translationnalité des études si cet aspect est recherché notamment (Garner et al., 2017). Il est donc essentiel de dépasser la vue réductionniste selon laquelle « la souris n'est qu'une éprouvette poilue » (Beans, 2018) et fondamental d'aborder les questions de recherche tant d'un point de vue écologique qu'évolutif car la considération de ces aspects peut ouvrir des perspectives d'amélioration technique et expérimentale pour répondre aux questions posées, de la manière la plus détaillée possible.

## **2. « LE RAT DE VILLE ET LE RAT DE LABORATOIRE »<sup>5</sup> OU LA QUESTION DE LA DISPARITE DES ANIMAUX ETUDES**

### **2.1. ANIMAL « NATUREL » OU ANIMAL « ARTIFICIEL » ?**

Le premier point de questionnement concerne le choix des animaux utilisés dans le cadre des études expérimentales. La majorité des animaux employés au laboratoire (rats, souris...) sont des animaux issus de lignées consanguines de sorte à limiter au maximum les

<sup>5</sup> Titre inspiré de la fable de Jean de La Fontaine « Le rat de ville et le rat des champs », *Fables*, Livre I

variations génétiques inter-individuelles (Chebib et al., 2021). Les animaux étant donc en théorie des clones les uns des autres, et étant élevés dans des conditions identiques, ceci permet de contrôler au maximum toutes les variables possibles (génétiques, environnementales...) et de tester ainsi l'effet spécifique des manipulations réalisées par l'expérimentateur (Calisi and Bentley, 2009). Néanmoins, tous les animaux de laboratoire ne sont pas équivalents et ceci est à prendre en considération.

Dans un premier temps, il est à noter que les animaux de laboratoires ne constituent pas une population homogène : il n'existe pas une souris de laboratoire mais des souris de laboratoire. Différentes lignées existent et ont été sélectionnées par les chercheurs en fonction de leurs propriétés et atypismes (les souris *nude* par exemple sont dépourvues de thymus et présentent de ce fait un système immunitaire altéré (Hoffman, 2017)). Le choix de la lignée qui sera impliquée dans le protocole expérimental peut donc avoir des conséquences importantes. Notamment, la réalisation d'expériences comportementales identiques chez des souris de lignées différentes n'aboutit pas aux mêmes conclusions (Luksys et al., 2009; Sultana et al.; Turner et al., 2017). Il est possible dans le cadre de ma thèse, par exemple, d'imaginer que le travail avec des souris BALBc au lieu de souris C57BL6/J ait pu donner des résultats différents comme une adaptation plus rapide des souris BABLbc à la suite des *switchs*.

Dans un second temps, un point essentiel concerne le choix du genre de l'animal étudié : les femelles restent encore sous-représentées dans de nombreuses études, notamment en neurosciences (Fritz et al., 2017). S'affranchir de l'étude de la moitié des individus d'une espèce constitue donc une réduction et une simplification expérimentales inappropriées car rien ne garantit que les mécanismes observés chez des mâles sont identiques à ceux des femelles (Harris et al., 2020). Cette question est d'ailleurs tout à fait pertinente dans le cadre de ma thèse puisqu'il semble que les souris mâles C57BL6/J utilisent une heuristique d'exploration fondée sur de la latéralisation pour s'adapter à un changement abrupt de contingences expérimentales. Or les femelles semblent présenter une plus forte propension à la latéralisation que les mâles (Chen et al., 2021; Harris et al., 2020) ce qui laisse donc en suspens une question importante quant aux stratégies d'adaptation des femelles dans la tâche de *switch* développée.

Comme discuté précédemment, la majorité des expériences de recherche sont menées avec des animaux de laboratoire. Toute la question est de savoir si les expériences menées avec ces animaux « artificiels », au sens où ils sont sélectionnés et façonnés par l'humain, reflètent des processus biologiques communs à ceux des animaux que l'on peut trouver dans

la nature. Les réponses tendent plutôt vers le non (Calisi and Bentley, 2009) : une souris ou un rat de laboratoire ne ressemble pas à une souris ou un rat des villes. Une étude chez le rat a notamment mis en évidence que le complexe basolatéral de l'amygdale, le noyau de la *stria terminalis*, et la partie caudale du bulbe olfactif accessoire sont plus développés chez les rats des villes que chez les rats de laboratoire, ce qui contribue ainsi potentiellement à leur plus haut niveau de vigilance (KOIZUMI et al., 2018). Dans le cadre de ma thèse par exemple, et de l'étude des stratégies d'adaptation de la souris, on pourrait notamment imaginer qu'une souris urbaine ayant cohabité pendant des siècles auprès de l'homme, présente des stratégies bien différentes et peut-être encore plus efficaces que celles identifiées chez la souris de laboratoire puisque les souris sauvages évoluent dans un milieu extrêmement complexe et volatil, contraignant donc ces animaux à user de ressources cognitives plus élaborées (Vrbanec et al., 2021).

Par ailleurs, les animaux de laboratoire présentent également des états microbiologiques bien spécifiques et répondent à des statuts sanitaires bien particuliers. La plupart des animaux sont hétéroxéniques, c'est-à-dire qu'ils hébergent une flore microbienne non pathogène afin de les protéger de toute maladie. Néanmoins, du fait des contrôles extrêmement précis de leur environnement (alimentation identique tout au long de leur vie, autoclavage de tout matériel mis à leur contact...), leur flore microbienne est donc très probablement peu diversifiée par rapport à celle des animaux sauvages. Or il est assez bien établi aujourd'hui que le microbiote intestinal a un rôle essentiel dans de multiples fonctions physiologiques mais aussi dans le comportement (Morais et al., 2021). La réalisation d'expériences comportementales identiques chez des souris de statuts sanitaires différents n'aboutit pas aux mêmes résultats (Heijtz et al., 2011).

Etudier des animaux sauvages offre donc des perspectives de recherche tout à fait fascinantes mais malheureusement complexes à mettre en œuvre dans la nature. De même capturer des animaux sauvages pour les étudier au laboratoire ne représente ni une solution éthique (Swart, 2004), ni une solution scientifiquement acceptable car la captivité d'animaux sauvages altère indéniablement leur état physiologique, mental et comportemental (Calisi and Bentley, 2009).

Des alternatives scientifiques doivent ainsi être envisagées car restreindre les études expérimentales à un cas particulier de modèle animal issu d'un environnement spécifique occulte probablement l'investigation de mécanismes évolutifs, génétiques et environnementaux. Comment mener alors des travaux de recherche afin d'étudier ces différents aspects ?

## **2.2. TIRER PROFIT DES SYSTEMES EXPERIMENTAUX AUTOMATISES AFIN DE REALISER DES ETUDES COMPORTEMENTALES A GRANDE ECHELLE**

Afin de répondre à ces questions, la première option envisageable et la plus simple serait de réaliser au laboratoire les expériences avec des animaux de différentes lignées ou différents statuts sanitaires dans le cadre de la même étude afin de comparer les résultats obtenus et d'évaluer s'ils divergent ou non, permettant ainsi d'assurer la reproductibilité des résultats ou au contraire d'investiguer les causes de ces divergences. Si, comme dans le cadre de ma thèse, des questions évolutives, quant aux mécanismes mis en jeu, sont soulevées, réaliser la même étude chez différentes espèces pourrait s'avérer tout à fait pertinent. Dans le cadre de la nouvelle tâche de *switch*, envisager une étude trans-lignées de souris, et trans-espèces avec d'autres équipes travaillant sur des modèles animaux différents serait une perspective d'étude particulièrement intéressante puisqu'elle permettrait d'apporter des réponses aux questions qui se posent quant à l'aspect évolutif des stratégies d'adaptation à travers les espèces en lien avec l'évolution de la complexité cérébrale.

Le développement de systèmes expérimentaux automatisés étant en pleine expansion aujourd'hui, il serait facilement envisageable d'implémenter ce genre d'expérimentations sans induire un coût élevé en termes de temps expérimental. En effet, ces systèmes expérimentaux automatisés permettent de collecter rapidement des quantités massives de données tout en limitant l'intervention de l'expérimentateur qui peut alors se consacrer à d'autres aspects de son travail de recherche comme l'analyse des données. Cet avantage a été largement mis en lumière avec l'utilisation de la BEATBOX notamment. De plus, une des particularités de la BEATBOX repose sur son haut degré de modularité hardware, facilitant ainsi son adaptation à différents animaux. Enfin, ce type de système est également plus respectueux de l'animal puisqu'il est possible pour lui de réaliser les tâches expérimentales à son rythme, en limitant la présence de l'expérimentateur et le stress associé : ceci ne peut donc être que bénéfique sur le résultat des études menées et sur le bien-être des animaux (Gouveia and Hurst, 2013, 2017).

Par ailleurs, au cours de ma thèse, j'ai été en mesure de développer rapidement, facilement et à bas coût ce système expérimental très simple d'utilisation et de construction, permettant pourtant la réalisation de tâches comportementales très complexes et la collecte efficace de milliers de données par animal. Pourquoi ne pas envisager l'adaptation et la diffusion de ce genre de système à l'extérieur du laboratoire : dans des zoos ? chez des particuliers pour les animaux domestiques ? dans la nature même ? Cette tendance semble en

effet se développer aujourd’hui. « Lab cognition going wild », est le projet de recherche audacieux d’Erica van de Waal qui souhaite caractériser les capacités cognitives de singes vervet qu’elle a déjà bien étudiés sous l’angle de l’écologie comportementale (Canteloup et al., 2020; Hemelrijk et al., 2020). Erica van de Waal implémente désormais des tâches de psychologie expérimentale à l’aide d’écrans tactiles dans des réserves d’Afrique du Sud afin de mieux comprendre les capacités cognitives de ces animaux sauvages en comparaison avec des résultats déjà bien établis au laboratoire (Inkawu Vervet Project).

Parce que les approches au laboratoire et dans la nature sont tout à fait complémentaires (Calisi and Bentley, 2009), il devient donc indispensable aujourd’hui d’essayer d’implémenter les travaux de recherche dans ces différents milieux grâce à une technologie expérimentale qui rend largement accessible ce genre d’approches. La comparaison des résultats parmi différentes espèces (« naturelles ou non ») permettrait ainsi d’envisager plus précisément la reproductibilité et le caractère translationnel des études et des résultats tout en considérant l’aspect évolutif des questions.

### **3. « SI LE LABORATOIRE NE VA PAS A LA NATURE, LA NATURE IRA AU LABORATOIRE »<sup>6</sup>**

#### **3.1. DEVELOPPER DES SYSTEMES EXPERIMENTAUX PLUS ECOLOGIQUES**

Les différentes propositions mentionnées précédemment n’étant pas toujours réalisables, il peut donc être intéressant de repenser l’hébergement et la manière dont les animaux de laboratoire réalisent les protocoles comportementaux notamment. Il est largement démontré aujourd’hui que les conditions d’hébergement et d’enrichissement ont des effets sur la neurogenèse et le comportement des animaux (Clemenson et al., 2015). S’affranchir des conditions d’hébergement classiques permettrait ainsi de se rapprocher au mieux de conditions de vie naturelle et de faire en sorte que les animaux ne voient pas toute leur vie les quatre mêmes parois en plexiglass de leur cage et les mêmes congénères qui partagent celle-ci.

Dans la perspective de faire évoluer leurs souris dans des environnements plus réalistes, deux chercheurs d’instituts scientifiques différents, Andrea Graham et Wayne Potts, ont

---

<sup>6</sup> Titre inspiré de l’œuvre de Paul Féval, le *Bossu*, 1858

développé de grands enclos (à l'intérieur ou à l'extérieur) dans lesquels plusieurs dizaines de souris (parfois mâles et femelles ensemble) peuvent évoluer librement. Tout en améliorant le bien-être des animaux, ces chercheurs ont pu mettre en évidence une différence de conclusions entre les études qu'ils ont menées dans leurs enclos et celles menées dans des conditions plus classiques (Beans, 2018). Ces études illustrent ainsi la nécessité de prendre en considération ces développements expérimentaux afin d'améliorer la validité et l'aspect translationnel des résultats (Graham, 2021).

Dans le cadre du comportement, ces environnements d'hébergement combinés à des systèmes automatisés de passation de tâches comportementales et d'acquisition de données constitueraient une approche expérimentale optimale. Ce type d'approche est notamment incarné par le système Autonomouse (Erskine et al., 2019) et le système Souris City (Torquet et al., 2018). Au sein de ces systèmes, les souris sont implantées avec des puces RFID qui permettent de les identifier et de les faire entrer dans des espaces dédiés à la réalisation de tâches comportementales. Il est ainsi possible d'acquérir des données comportementales complexes pour chacun des animaux. En combinaison avec des systèmes d'enregistrement d'activité neuronale ou de neuromodulation sans fil (Barbera et al., 2019; Fan et al., 2011; Zhang et al., 2019), il est parfaitement envisageable de réaliser des expérimentations complexes afin d'étudier le comportement et les bases neuronales sous-jacentes.

### **3.2. ENJEUX ET AVANTAGES DE JEUX DE DONNEES COMPLEXES ET BRUITES**

Bien évidemment, de par la diminution du contrôle de l'environnement de l'animal, les résultats obtenus peuvent être plus variables, plus bruités car ce type d'approche expérimentale en grand enclos avec de nombreuses et multiples interactions sociables favorise l'expression de l'individualité de chaque animal (Freund et al., 2013; Torquet et al., 2018). Néanmoins en collectant suffisamment de données, il est possible d'en filtrer une partie afin d'en retirer le bruit.

Dans le cadre de mon étude comportementale de *switch* par exemple, j'ai notamment dû implémenter des techniques de filtrage des données afin d'éliminer le maximum de bruit au sein de celles-ci en retirant les essais isolés ou réalisés au cours des périodes de nuit de la souris. Chaque souris présentant en réalité son propre rythme (certaines suivant un rythme strictement nyctéméral, d'autres présentant un décalage de phase au cours de la tâche...), il m'a été nécessaire de développer des techniques d'analyses objectives sur les temps inter-

essais ou encore l'activité cumulée des animaux afin de limiter au maximum le choix de critères subjectifs d'exclusion ou d'inclusion des données et de trouver une méthode qui, appliquée à chacune des souris, permet d'extraire les données pertinentes. Les méthodes de filtrage que j'ai mises en place se sont révélées efficaces et m'ont ainsi permis de conserver l'individualité propre de chacune des souris et de rester aussi proche que possible des données comportementales tout en réduisant la complexité du jeu de données initial en en retirant le bruit. Grâce à ces jeux de données ainsi obtenus, j'ai pu réaliser des analyses individuelles pour chacun des animaux, permettant ainsi de caractériser finement les séquences d'apprentissage de chacun, et d'identifier les biais de latéralité individuels. Ce type d'analyse m'a notamment permis d'identifier deux groupes d'animaux distincts utilisant des stratégies de réapprentissage différentes à la suite du *switch* partiel. Dépasser la vision collective et moyennée des résultats comportementaux permet donc d'ouvrir des perspectives différentes d'interprétation et d'analyse.

## **4. « LE CHAT QUI APPRIT A UNE MOUETTE A VOLER »<sup>7</sup> OU LA QUESTION DU CHOIX DES PARADIGMES EXPERIMENTAUX**

### **4.1. LES PARADIGMES EXPERIMENTAUX FACE A L'ECOLOGIE DE L'ESPECE ETUDIEE**

Ce titre, bien qu'humoristique permet de faire écho aux limites des études de psychologie comparative. Comme le soulignent Schubiger et collaborateurs, « *La conception des tests reste le problème majeur de la psychologie comparative contemporaine et il est essentiel que les chercheurs valident et redéfinissent si nécessaire les tests cognitifs, afin de s'assurer que les tâches reflètent fidèlement les capacités cognitives* » des espèces (Schubiger et al., 2020).

Êtres humains et autres animaux ont évolué depuis des millénaires dans des niches écologiques différentes et ont donc été confrontés à des contraintes et des problèmes largement différents. Il est ainsi nécessaire de prendre en compte les contraintes environnementales (et même sociétales dans le cas de l'humain) qui ont pu façonner certains aspects cognitifs et comportementaux afin de comprendre les résultats des expériences.

---

<sup>7</sup> Titre inspiré de l'œuvre de Luis Sepùlveda, *História de una gaviota y del gato que le enseñó a volar* (*Histoire d'une mouette et du chat qui lui apprit à voler*), 1996

Chaque espèce présente des spécificités qui lui sont propres et le choix du modèle ne peut être décorrélé de l'expérience menée et des résultats : un rat ne peut pas être considéré comme une grosse souris et l'interprétation des résultats doit prendre en considération cette divergence évolutive (Ellenbroek and Youn, 2016).

Ben Vermaercke et ses collaborateurs par exemple, ont mené chez le rat et l'être humain une étude translationnelle au cours de laquelle les rats se sont avérés plus performants que les humains, résultat peu habituel en comparaison aux autres études translationnelles. Il est à noter que les sujets humains dans le cadre de cette tâche ont reçu le strict minimum d'informations pour la réaliser et se retrouvent donc au même niveau que les rats pour comprendre la configuration générale de la procédure expérimentale. Cette étude est donc bien différente des études classiques chez l'Homme durant lesquelles les sujets reçoivent des informations très détaillées qui peuvent donc les influencer dans la réalisation de la tâche, biaisant ainsi les résultats comportementaux. Au cours de deux tâches, les sujets murins et humains devaient apprendre à distinguer des motifs cibles, puis ils étaient testés sur leur capacité à appliquer les règles apprises à de nouveaux motifs. Dans la première tâche, les motifs ne variaient que sur une seule dimension - l'orientation ou la position des figures sur l'écran - et les rats et les humains obtenaient des performances similaires. Dans la seconde en revanche, les motifs variaient sur les deux dimensions, et les rats étaient alors plus performants que les humains (Vermaercke et al., 2014). Or une des théories de l'apprentissage des catégories distingue deux systèmes concurrents (Ashby et al., 1998) : (1) un système explicite, fondé sur l'utilisation de règles, (2) un système plus implicite qui intervient lorsque des catégories ne peuvent pas être distinguées par une solution simple fondée sur des règles, et qui nécessite ainsi l'intégration de plusieurs paramètres. Une des explications des auteurs quant aux différences observées est liée au fait que les humains ont une propension majeure à chercher et extraire des règles dans l'environnement puisque c'est ce à quoi nous sommes conditionnés depuis l'enfance par nos parents, à l'école, dans la société ou encore au travail (Beard, 2015).

Une autre étude a notamment montré chez les primates non-humains que les chimpanzés préfèrent de manière significative des options risquées contrairement aux bonobos qui préfèrent des options fixes (Heilbronner et al., 2008). Or dans la nature, les chimpanzés exploitent des ressources de nourriture dont la temporalité et la spatialité sont variables alors que les bonobos se nourrissent préférentiellement de plantes présentes en quantité abondante et continue dans le temps. Cette étude montre donc que malgré la proximité des espèces étudiées, la pression écologique sur celles-ci peut largement influencer

leurs processus de prise de décision : l'écologie de chaque espèce est donc une force sélective qui façonne des répertoires comportementaux uniques.

De même, lorsque certains aspects des processus décisionnels sont étudiés chez l'être humain à différents âges, les résultats sont largement différents. Les différentes formes d'exploration ne sont pas implémentées de la même manière chez un enfant et chez un adulte (Schulz and Gershman, 2019; Wilson et al., 2021). La proportion d'exploration aléatoire diminue fortement à mesure que les enfants vieillissent (Meder et al., 2020) et ceci pourrait être en lien avec l'expérience qu'ils accumulent dans leur milieu de vie et le processus d'apprentissage des règles qui régissent cet environnement.

L'ensemble de ces études suggèrent donc que l'interprétation des résultats d'une tâche donnée doit nécessairement prendre en compte la nature de l'espèce étudiée, sa physiologie propre et les contraintes écologiques auxquelles elle est censée s'adapter au quotidien. Dans le cadre de ma thèse par exemple, il peut être légitime de se poser la question quant à la réalisation d'une tâche de discrimination visuelle pour des animaux dont d'autres capacités sensorielles sont plus développées. Il a été montré par exemple que des rats sont plus performants dans une tâche d'acquisition de *learning-set* si celle-ci met en jeu l'olfaction et non la vue (Slotnick and Katz, 1974). Les souris présentent donc peut-être des stratégies d'adaptation différentes en fonction des modalités sensorielles mises en jeu et l'utilisation de modalités tactiles, auditives ou olfactives auraient pu révéler des résultats comportements différents de ceux obtenus au cours de l'étude menée au cours de ma thèse.

## 4.2. LES BIG DATA AU SERVICE DES PARADIGMES EXPERIMENTAUX

Au-delà même de la question des tests de psychologie comparative, la question de la validité des tâches de psychologie expérimentale implantée au laboratoire pour l'être humain peut également se poser. Ces tâches couplées à des modèles computationnels ont permis des progrès considérables dans la compréhension de l'esprit humain mais il pourrait être intéressant de sortir du laboratoire pour étudier le comportement et la cognition dans des environnements plus complexes, plus vastes et plus écologiquement pertinents.

Aujourd'hui en effet, les enregistrements omniprésents du comportement humain offrent la possibilité d'étudier la cognition humaine à une échelle et à un niveau de validité qui ne pourront jamais être atteints en laboratoire. Des services comme Twitter peuvent donner accès au flux de conscience de millions de personnes, Facebook peut fournir des informations sur leurs connexions et leurs interactions (Griffiths, 2015) tandis que des applications de

livraison de plats cuisinés comme Deliveroo a permis d'investiguer les stratégies de *fourrageage* et le dilemme *exploration – exploitation* chez des milliers de sujets (Schulz et al., 2019). En parallèle, le développement de systèmes d'enregistrement de l'activité cérébrale sans fils (Topalovic et al., 2020) et miniaturisés (Blum et al., 2020) constitue des perspectives exaltantes pour envisager d'étudier l'activité cérébrale en dehors du laboratoire.

## **CONCLUSION**

Aujourd'hui ce ne sont ni les moyens techniques ni les moyens analytiques qui manquent pour répondre à des questions aussi complexes que l'*apprentissage*, la balance *exploration – exploitation* et leur importance dans le cadre des stratégies d'*adaptation* à un environnement *volatil*. Tout l'enjeu consiste donc à repenser les paradigmes expérimentaux en perspective de considérations évolutives, écologiques ainsi que computationnelles. En quelque sorte, « Think outside the (BEAT)Box » !

# BIBLIOGRAPHIE

---

- Abzhanov, A. (2010). Darwin's Galapagos finches in modern biology. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* *365*, 1001–1007.
- Addicott, M.A., Pearson, J.M., Sweitzer, M.M., Barack, D.L., and Platt, M.L. (2017). A Primer on Foraging and the Explore/Exploit Trade-Off for Psychiatry Research. *Neuropsychopharmacology* *42*, 1931–1939.
- Aguillon, V., Angelaki, D., Bayer, H., Bonacchi, N., Carandini, M., Cazettes, F., Churchland, A., Chapuis, G., Dan, Y., Dewitt, E., et al. (2020). A standardized and reproducible method to measure decision-making in mice.
- Amiez, C., Joseph, J.-P., and Procyk, E. (2006). Reward encoding in the monkey anterior cingulate cortex. *Cereb Cortex* *16*, 1040–1055.
- Amodeo, D.A., Hassan, O., Klein, L., Halberstadt, A.L., and Powell, S.B. (2020). Acute serotonin 2A receptor activation impairs behavioral flexibility in mice. *Behav Brain Res* *395*, 112861.
- Andreelli, F., and Mosbah, H. (2014). IRM fonctionnelle cérébrale : les principes. *Médecine des Maladies Métaboliques* *8*, 13–19.
- Ashby, F.G., Alfonso-Reese, L.A., Turken, A.U., and Waldron, E.M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychol Rev* *105*, 442–481.
- Aston-Jones, G., and Cohen, J.D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu Rev Neurosci* *28*, 403–450.
- Atasoy, D., Aponte, Y., Su, H.H., and Sternson, S.M. (2008). A FLEX Switch Targets Channelrhodopsin-2 to Multiple Cell Types for Imaging and Long-Range Circuit Mapping. *J. Neurosci.* *28*, 7025–7030.
- Audet, J.-N., and Lefebvre, L. (2017). What's flexible in behavioral flexibility? *Behav Ecol* *28*, 943–947.
- Audet, J.-N., Kayello, L., Duceatz, S., Perillo, S., Cauchard, L., Howard, J.T., O'Connell, L.A., Jarvis, E.D., and Lefebvre, L. (2018). Divergence in problem-solving skills is associated with differential expression of glutamate receptors in wild finches. *Science Advances* *4*, eaao6369.
- Auer, P., and Cesa-Bianchi, N. Finite-time Analysis of the Multiarmed Bandit Problem. *22*.
- Barentzen, S., Casado-Sainz, A., Lange, D., Shalgunov, V., Tejada, I.M., Xiong, M., L'Estrade, E.T., Edgar, F.G., Lee, H., Herth, M.M., et al. (2019). The Chemogenetic Receptor Ligand Clozapine N-Oxide Induces in vivo Neuroreceptor Occupancy and Reduces Striatal Glutamate Levels. *Front Neurosci* *13*.
- Balcombe, J.P., Barnard, N.D., and Sandusky, C. (2004). Laboratory routines cause animal stress. *Contemp Top Lab Anim Sci* *43*, 42–51.
- Barbera, G., Liang, B., Zhang, L., Li, Y., and Lin, D.-T. (2019). A wireless miniScope for deep brain imaging in freely moving mice. *J Neurosci Methods* *323*, 56–60.
- Bartolo, R., and Averbeck, B.B. (2020). Prefrontal Cortex Predicts State Switches during Reversal Learning. *Neuron* *106*, 1044–1054.e4.
- Bayne, T., Brainard, D., Byrne, R.W., Chittka, L., Clayton, N., Heyes, C., Mather, J., Ölveczky, B., Shadlen, M., Suddendorf, T., et al. (2019). What is cognition? *Curr. Biol.* *29*, R608–R615.
- Beans, C. (2018). News Feature: What happens when lab animals go wild. *Proc Natl Acad Sci U S A* *115*, 3196–3199.

- Beard, A. (2015). Rats Can Be Smarter Than People. *Harvard Business Review*.
- Behrens, T.E.J., Woolrich, M.W., Walton, M.E., and Rushworth, M.F.S. (2007). Learning the value of information in an uncertain world. *Nat. Neurosci.* *10*, 1214–1221.
- Belkaid, M., Bousseyrol, E., Durand-de Cuttoli, R., Dongelmans, M., Duranté, E.K., Ahmed Yahia, T., Didienne, S., Hanesse, B., Come, M., Mourot, A., et al. (2020). Mice adaptively generate choice variability in a deterministic task. *Communications Biology* *3*, 1–9.
- Benzina, N. (2019). Apport d'une approche translationnelle dans l'étude du lien entre flexibilité cognitive et compulsion : une étude comportementale trans-espèces. These de doctorat. Sorbonne université.
- Benzina, N., N'Diaye, K., Pelissolo, A., Mallet, L., and Burguière, E. (2021). A cross-species assessment of behavioral flexibility in compulsive disorders. *Communications Biology* *4*, 1–12.
- Bernstein, J.G., Garrity, P.A., and Boyden, E.S. (2012). Optogenetics and thermogenetics: technologies for controlling the activity of targeted cells within intact neural circuits. *Curr Opin Neurobiol* *22*, 61–71.
- Blanchard, R.J., McKittrick, C.R., and Blanchard, D.C. (2001). Animal models of social stress: effects on behavior and brain neurochemical systems. *Physiol. Behav.* *73*, 261–271.
- Blum, S., Emkes, R., Minow, F., Anlauff, J., Finke, A., and Debener, S. (2020). Flex-printed forehead EEG sensors (fEEGGrid) for long-term EEG acquisition. *Journal of Neural Engineering* *17*.
- Bouret, S., and Sara, S.J. (2005). Network reset: a simplified overarching theory of locus coeruleus noradrenaline function. *Trends Neurosci* *28*, 574–582.
- Boyden, E.S., Zhang, F., Bamberg, E., Nagel, G., and Deisseroth, K. (2005). Millisecond-timescale, genetically targeted optical control of neural activity. *Nat Neurosci* *8*, 1263–1268.
- Braun, D.A., Mehring, C., and Wolpert, D.M. (2010). Structure learning in action. *Behav Brain Res* *206*, 157–165.
- Bretschtajn, L., and Gedroyc, W. (2018). Brain-focussed ultrasound: what's the “FUS” all about? A review of current and emerging neurological applications. *Br J Radiol* *91*, 20170481.
- Brigman, J.L., Graybeal, C., and Holmes, A. (2010). Predictably irrational: assaying cognitive inflexibility in mouse models of schizophrenia. *Front Neurosci* *4*.
- Brodmann, K. (1909). Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues (Leipzig : Barth).
- Brown, V., and Tait, D. (2010). Behavioral Flexibility: Attentional Shifting, Rule Switching and Response Reversal. pp. 209–213.
- Brunelli, M., Castellucci, V., and Kandel, E.R. (1976). Synaptic facilitation and behavioral sensitization in Aplysia: possible role of serotonin and cyclic AMP. *Science* *194*, 1178–1181.
- Bublitz, A., Weinhold, S.R., Strobel, S., Dehnhardt, G., and Hanke, F.D. (2017). Reconsideration of Serial Visual Reversal Learning in Octopus (*Octopus vulgaris*) from a Methodological Perspective. *Front. Physiol.* *8*.
- Budygin, E.A., Bass, C.E., Grinevich, V.P., Deal, A.L., Bonin, K.D., and Weiner, J.L. (2020). Opposite Consequences of Tonic and Phasic Increases in Accumbal Dopamine on Alcohol-Seeking Behavior. *IScience* *23*, 100877.
- Burguière, E., Monteiro, P., Feng, G., and Graybiel, A.M. (2013). Optogenetic stimulation of lateral orbitofronto-striatal pathway suppresses compulsive behaviors. *Science* *340*, 1243–1246.
- Burkhardt, A.M., and Zlotnik, A. (2013). Translating translational research: mouse models of human disease. *Cellular & Molecular Immunology* *10*, 373–374.

Calisi, R.M., and Bentley, G.E. (2009). Lab and field experiments: Are they the same animal? *Hormones and Behavior* *56*, 1–10.

Canteloup, C., Hoppitt, W., and van de Waal, E. (2020). Wild primates copy higher-ranked individuals in a social transmission experiment. *Nature Communications* *11*, 459.

Carlén, M. (2017). What constitutes the prefrontal cortex? *Science* *358*, 478–482.

Castañé, A., Theobald, D.E.H., and Robbins, T.W. (2010). Selective lesions of the dorsomedial striatum impair serial spatial reversal learning in rats. *Behav Brain Res* *210*, 74–83.

Cavanagh, J.F., Wiecki, T.V., Cohen, M.X., Figueroa, C.M., Samanta, J., Sherman, S.J., and Frank, M.J. (2011). Subthalamic nucleus stimulation reverses mediofrontal influence over decision threshold. *Nat Neurosci* *14*, 1462–1467.

Chari, A., Thornton, R.C., Tisdall, M.M., and Scott, R.C. (2020). Microelectrode recordings in human epilepsy: a case for clinical translation. *Brain Communications* *2*.

Charnov, E.L. (1976). Optimal foraging, the marginal value theorem. *Theoretical Population Biology* *9*, 129–136.

Charron, S., and Collette, S. (2012). Ready to Face the Future: Brain Mechanisms for Cognitive Flexibility and Exploration. *Recherches économiques de Louvain* Vol. *78*, 73–81.

Chau, B.K.H., Kolling, N., Hunt, L.T., Walton, M.E., and Rushworth, M.F.S. (2014). A neural mechanism underlying failure of optimal choice with multiple alternatives. *Nat Neurosci* *17*, 463–470.

de Chaumont, F., Coura, R.D.-S., Serreau, P., Cressant, A., Chabout, J., Granon, S., and Olivo-Marin, J.-C. (2012). Computerized video analysis of social interactions in mice. *Nat Methods* *9*, 410–417.

Chebib, J., Jackson, B.C., López-Cortegano, E., Tautz, D., and Keightley, P.D. (2021). Inbred lab mice are not isogenic: genetic variation within inbred strains used to infer the mutation rate per nucleotide site. *Heredity* *126*, 107–116.

Chen, C.S., Ebitz, R.B., Bindas, S.R., Redish, A.D., Hayden, B.Y., and Grissom, N.M. (2021). Divergent Strategies for Learning in Males and Females. *Curr Biol* *31*, 39-50.e4.

Chinwalla, A.T., Cook, L.L., Delehaunty, K.D., Fewell, G.A., Fulton, L.A., Fulton, R.S., Graves, T.A., Hillier, L.W., Mardis, E.R., McPherson, J.D., et al. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* *420*, 520–562.

Chow, P.K.Y., Lurz, P.W.W., and Lea, S.E.G. (2018). A battle of wits? Problem-solving abilities in invasive eastern grey squirrels and native Eurasian red squirrels. *Animal Behaviour* *137*, 11–20.

Chun, M.M., and Wolfe, J.M. (1996). Just say no: how are visual searches terminated when there is no target present? *Cogn Psychol* *30*, 39–78.

Cinotti, F., Fresno, V., Aklil, N., Coutureau, E., Girard, B., Marchand, A.R., and Khamassi, M. (2019). Dopamine blockade impairs the exploration-exploitation trade-off in rats. *Scientific Reports* *9*, 6770.

Clemenson, G., Deng, W., and Gage, F. (2015). Environmental enrichment and neurogenesis: From mice to humans. *Current Opinion in Behavioral Sciences* *1*.

Cohen, J.D., McClure, S.M., and Yu, A.J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* *362*, 933–942.

Collins, A., and Koechlin, E. (2012). Reasoning, learning, and creativity: frontal lobe function and human decision-making. *PLoS Biol.* *10*, e1001293.

- Collins, A.G.E., and Frank, M.J. (2013). Cognitive control over learning: creating, clustering, and generalizing task-set structure. *Psychol Rev* 120, 190–229.
- Cooper, G.F. (1990). The computational complexity of probabilistic inference using bayesian belief networks. *Artificial Intelligence* 42, 393–405.
- Corkin, S. (1984). Lasting Consequences of Bilateral Medial Temporal Lobectomy: Clinical Course and Experimental Findings in H.M. *Semin Neurol* 4, 249–259.
- Costa, V.D., Tran, V.L., Turchi, J., and Averbeck, B.B. (2015). Reversal Learning and Dopamine: A Bayesian Perspective. *J. Neurosci.* 35, 2407–2416.
- Cristóbal-Azkarate, J., and Arroyo-Rodríguez, V. (2007). Diet and activity pattern of howler monkeys (*Alouatta palliata*) in Los Tuxtlas, Mexico: effects of habitat fragmentation and implications for conservation. *American Journal of Primatology* 69, 1013–1029.
- Dajani, D.R., and Uddin, L.Q. (2015). Demystifying cognitive flexibility: Implications for clinical and developmental neuroscience. *Trends Neurosci.* 38, 571–578.
- D'Angelo, E., Solinas, S., Garrido, J., Casellato, C., Pedrocchi, A., Mapelli, J., Gandolfi, D., and Prestori, F. (2013). Realistic modeling of neurons and networks: towards brain simulation. *Funct Neurol* 28, 153–166.
- Darwin, C. (2012). *On the Origin of the Species and The Voyage of the Beagle* (Graphic Arts Books).
- Davis, K.M., Kangassalo, L., Spapé, M., and Ruotsalo, T. (2020). Brainsourcing: Crowdsourcing Recognition Tasks via Collaborative Brain-Computer Interfacing. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, (New York, NY, USA: Association for Computing Machinery), pp. 1–14.
- Daw, N.D., O'Doherty, J.P., Dayan, P., Seymour, B., and Dolan, R.J. (2006). Cortical substrates for exploratory decisions in humans. *Nature* 441, 876–879.
- Dayan, P., and Niv, Y. (2008). Reinforcement learning: The Good, The Bad and The Ugly. *Current Opinion in Neurobiology* 18, 185–196.
- Deisseroth, K. (2011). Optogenetics. *Nat Methods* 8, 26–29.
- Del Giudice, M., and Crespi, B.J. (2018). Basic functional trade-offs in cognition: An integrative framework. *Cognition* 179, 56–70.
- Deng, H., Xiao, X., and Wang, Z. (2016). Periaqueductal Gray Neuronal Activities Underlie Different Aspects of Defensive Behaviors. *J. Neurosci.* 36, 7580–7588.
- Dennett, D.C. and The Journal of Philosophy, Inc. (1971). Intentional Systems: *Journal of Philosophy* 68, 87–106.
- Devilbiss, D.M., and Waterhouse, B.D. (2011). Phasic and Tonic Patterns of Locus Coeruleus Output Differentially Modulate Sensory Network Function in the Awake Rat. *J Neurophysiol* 105, 69–87.
- Dhawan, S.S., Tait, D.S., and Brown, V.J. (2019). More rapid reversal learning following overtraining in the rat is evidence that behavioural and cognitive flexibility are dissociable. *Behav. Brain Res.* 363, 45–52.
- D'Hondt, F., and Sequeira, H. (2013). Chapitre 4. La magnétoencéphalographie dans l'étude de la cognition (Dunod).
- Domenech, P., and Koechlin, E. (2015). Executive control and decision-making in the prefrontal cortex. *Current Opinion in Behavioral Sciences* 1, 101–106.
- Domenech, P., Rheims, S., and Koechlin, E. (2020). Neural mechanisms resolving exploitation-exploration dilemmas in the medial prefrontal cortex. *Science* 369.

- Donoso, M., Collins, A.G.E., and Koechlin, E. (2014). Foundations of human reasoning in the prefrontal cortex. *Science* *344*, 1481–1486.
- Doya, K. (2002). Metalearning and neuromodulation. *Neural Networks* *15*, 495–506.
- Durstewitz, D., Vittoz, N.M., Floresco, S.B., and Seamans, J.K. (2010). Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. *Neuron* *66*, 438–448.
- Ebitz, R.B., and Hayden, B.Y. (2016). Dorsal anterior cingulate: a Rorschach test for cognitive neuroscience. *Nature Neuroscience* *19*, 1278–1279.
- Ellenbroek, B., and Youn, J. (2016). Rodent models in neuroscience research: is it a rat race? *Dis Model Mech* *9*, 1079–1087.
- Ellwood, I.T., Patel, T., Wadia, V., Lee, A.T., Liptak, A.T., Bender, K.J., and Sohal, V.S. (2017). Tonic or Phasic Stimulation of Dopaminergic Projections to Prefrontal Cortex Causes Mice to Maintain or Deviate from Previously Learned Behavioral Strategies. *J Neurosci* *37*, 8315–8329.
- Erskine, A., Bus, T., Herb, J.T., and Schaefer, A.T. (2019). AutonoMouse: High throughput operant conditioning reveals progressive impairment with graded olfactory bulb lesions. *PLoS ONE* *14*, e0211571.
- Euvrard, M., Benzina, N., Travert, Y., Tissier, P., Penderia, G., and Burguière, E. (2019). The Behavioural and Ecological Automated operanT Box (BEATBox): a flexible and low-cost system for high-throughput behavioural data acquisition. Meeting of the International Basal Ganglia Society (IBAGS). Biarritz, France.
- Fan, D., Rich, D., Holtzman, T., Ruther, P., Dalley, J., Lopez, A., Rossi, M., Barter, J., Salas-Meza, D., Herwik, S., et al. (2011). A Wireless Multi-Channel Recording System for Freely Behaving Mice and Rats. *PloS One* *6*, e22033.
- Farashahi, S., Rowe, K., Aslami, Z., Lee, D., and Soltani, A. (2017). Feature-based learning improves adaptability without compromising precision. *Nature Communications* *8*, 1768.
- Faraut, M. (2015). Apprendre à apprendre dans un environnement incertain, et dynamique des réseaux corticaux pour la flexibilité comportementale. These de doctorat. Lyon 1.
- Findling, C., Skvortsova, V., Dromnelle, R., Palminteri, S., and Wyart, V. (2019). Computational noise in reward-guided learning drives behavioral variability in volatile environments. *Nature Neuroscience* *22*, 2066–2077.
- Freund, J., Brandmaier, A.M., Lewejohann, L., Kirste, I., Kritzler, M., Krüger, A., Sachser, N., Lindenberger, U., and Kempermann, G. (2013). Emergence of individuality in genetically identical mice. *Science* *340*, 756–759.
- Fritz, A., Amrein, I., and Wolfer, D.P. (2017). Similar reliability and equivalent performance of female and male mice in the open field and water-maze place navigation task. *Am J Med Genet C Semin Med Genet* *175*, 380–391.
- Fuster, J. (2013). Cognitive Functions of the Prefrontal Cortex. In *Frontiers in Human Neuroscience*, pp. 11–22.
- Gallistel, C.R., Mark, T.A., King, A.P., and Latham, P.E. (2001). The rat approximates an ideal detector of changes in rates of reward: implications for the law of effect. *J Exp Psychol Anim Behav Process* *27*, 354–372.
- Gallistel, C.R., Fairhurst, S., and Balsam, P. (2004). The learning curve: implications of a quantitative analysis. *Proc Natl Acad Sci U S A* *101*, 13124–13131.
- Gallo, M. (2007). Reversible Inactivation of Brain Circuits in Learning and Memory Research. In *Neural Plasticity and Memory: From Genes to Brain Imaging*, F. Bermúdez-Rattoni, ed. (Boca Raton (FL): CRC Press/Taylor & Francis), p.

- Gao, P., and Ganguli, S. (2015). On simplicity and complexity in the brave new world of large-scale neuroscience. *Curr Opin Neurobiol* 32, 148–155.
- Garner, J., Gaskill, B., Weber, E., Ahloy Dallaire, J., and Pritchett-Corning, K. (2017). Introducing Therioepistemology: The study of how knowledge is gained from animal research. *Lab Animal* 46, 103–113.
- Genon, S., Reid, A., Langner, R., Amunts, K., and Eickhoff, S.B. (2018). How to Characterize the Function of a Brain Region. *Trends in Cognitive Sciences* 22, 350–364.
- Gershman, S.J. (2018). Deconstructing the human algorithms for exploration. *Cognition* 173, 34–42.
- Gershman, S.J. Uncertainty and exploration. 17.
- Gershman, S.J., and Niv, Y. (2010). Learning latent structure: Carving nature at its joints. *Curr Opin Neurobiol* 20, 251–256.
- Gershman, S.J., Blei, D.M., and Niv, Y. (2010). Context, learning, and extinction. *Psychol Rev* 117, 197–209.
- Gershman, S.J., Norman, K.A., and Niv, Y. (2015). Discovering latent causes in reinforcement learning. *Current Opinion in Behavioral Sciences* 5, 43–50.
- Gittins, J.C. (1979). Bandit Processes and Dynamic Allocation Indices. *Journal of the Royal Statistical Society: Series B (Methodological)* 41, 148–164.
- Gomez, J.L., Bonaventura, J., Lesniak, W., Mathews, W.B., Sysa-Shah, P., Rodriguez, L.A., Ellis, R.J., Richie, C.T., Harvey, B.K., Dannals, R.F., et al. (2017). Chemogenetics revealed: DREADD occupancy and activation via converted clozapine. *Science* 357, 503–507.
- Gomez-Marin, A., and Mainen, Z.F. (2016). Expanding perspectives on cognition in humans, animals, and machines. *Current Opinion in Neurobiology* 37, 85–91.
- Gottlieb, J., Oudeyer, P.-Y., Lopes, M., and Baranes, A. (2013). Information seeking, curiosity and attention: computational and neural mechanisms. *Trends Cogn Sci* 17, 585–593.
- Gouveia, K., and Hurst, J.L. (2013). Reducing mouse anxiety during handling: effect of experience with handling tunnels. *PLoS ONE* 8, e66401.
- Gouveia, K., and Hurst, J.L. (2017). Optimising reliability of mouse performance in behavioural testing: the major role of non-aversive handling. *Scientific Reports* 7, 44999.
- Graham, A.L. (2021). Naturalizing mouse models for immunology. *Nature Immunology* 22, 111–117.
- Grienberger, C., and Konnerth, A. (2012). Imaging Calcium in Neurons. *Neuron* 73, 862–885.
- Griffin, A.S., Guez, D., Lermite, F., and Patience, M. (2013). Tracking Changing Environments: Innovators Are Fast, but Not Flexible Learners. *PLOS ONE* 8, e84907.
- Griffiths, T. (2015). Manifesto for a new (computational) cognitive revolution. *Cognition*.
- Grossman, C.D., Bari, B.A., and Cohen, J.Y. (2020). Serotonin neurons modulate learning rate through uncertainty. *BioRxiv* 2020.10.24.353508.
- Hamilton, D.A., and Brigman, J.L. (2015). Behavioral flexibility in rats and mice: contributions of distinct frontocortical regions. *Genes Brain Behav.* 14, 4–21.
- Harris, C., Aguirre, C., Kolli, S., Das, K., Izquierdo, A., and Soltani, A. (2020). Unique features of stimulus-based probabilistic reversal learning. *BioRxiv* 2020.09.24.310771.
- Hayden, B.Y. (2017). The foraging perspective on economic choice. *BioRxiv* 190991.

- Hayden, B.Y., Pearson, J.M., and Platt, M.L. (2009). Fictive reward signals in the anterior cingulate cortex. *Science* *324*, 948–950.
- Hayden, B.Y., Heilbronner, S.R., Pearson, J.M., and Platt, M.L. (2011a). Surprise Signals in Anterior Cingulate Cortex: Neuronal Encoding of Unsigned Reward Prediction Errors Driving Adjustment in Behavior. *J Neurosci* *31*, 4178–4187.
- Hayden, B.Y., Pearson, J.M., and Platt, M.L. (2011b). Neuronal basis of sequential foraging decisions in a patchy environment. *Nat Neurosci* *14*, 933–939.
- Healy, S., and Hurly, T. (2013). What hummingbirds can tell us about cognition in the wild. *Comp. Cogn. Behav. Rev.* *8*, 13–28.
- Heijtz, R.D., Wang, S., Anuar, F., Qian, Y., Björkholm, B., Samuelsson, A., Hibberd, M.L., Forssberg, H., and Pettersson, S. (2011). Normal gut microbiota modulates brain development and behavior. *PNAS* *108*, 3047–3052.
- Heilbronner, S.R., and Hayden, B.Y. (2016). Dorsal Anterior Cingulate Cortex: A Bottom-Up View. *Annu Rev Neurosci* *39*, 149–170.
- Heilbronner, S.R., Rosati, A.G., Stevens, J.R., Hare, B., and Hauser, M.D. (2008). A fruit in the hand or two in the bush? Divergent risk preferences in chimpanzees and bonobos. *Biology Letters* *4*, 246–249.
- Heilbronner, S.R., Rodriguez-Romaguera, J., Quirk, G.J., Groenewegen, H.J., and Haber, S.N. (2016). Circuit-Based Corticostriatal Homologies Between Rat and Primate. *Biol. Psychiatry* *80*, 509–521.
- Hemelrijk, C.K., Wubs, M., Gort, G., Botting, J., and van de Waal, E. (2020). Dynamics of Intersexual Dominance and Adult Sex- Ratio in Wild Vervet Monkeys. *Front. Psychol.* *11*.
- Hermoso-Mendizabal, A., Hyafil, A., Rueda-Orozco, P.E., Jaramillo, S., Robbe, D., and de la Rocha, J. (2020). Response outcomes gate the impact of expectations on perceptual decisions. *Nature Communications* *11*, 1057.
- Heukelum, S. van, Mars, R.B., Guthrie, M., Buitelaar, J.K., Beckmann, C.F., Tiesinga, P.H.E., Vogt, B.A., Glennon, J.C., and Havenith, M.N. (2020). Where is Cingulate Cortex? A Cross-Species View. *Trends in Neurosciences* *43*, 285–299.
- Hills, T.T., Todd, P.M., Lazer, D., Redish, A.D., Couzin, I.D., and Cognitive Search Research Group (2015). Exploration versus exploitation in space, mind, and society. *Trends Cogn. Sci. (Regul. Ed.)* *19*, 46–54.
- Hoffman, R.M. (2017). Nude Mouse☆. In Reference Module in Life Sciences, (Elsevier), p.
- Holroyd, C., and Coles, M. (2002). Holroyd, C.B. & Coles, M.G.H. The neural basis of error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychol. Rev.* *109*, 679–709. *Psychological Review* *109*, 679–709.
- Holroyd, C., and Yeung, N. (2011). An Integrative Theory of Anterior Cingulate Cortex Function: Option Selection in Hierarchical Reinforcement Learning. *Neural Basis of Motivational and Cognitive Control* *16*, 333–349.
- Holroyd, C.B., and Coles, M.G.H. (2008). Dorsal anterior cingulate cortex integrates reinforcement history to guide voluntary behavior. *Cortex* *44*, 548–559.
- Holroyd, C.B., and Yeung, N. (2012). Motivation of extended behaviors by anterior cingulate cortex. *Trends Cogn. Sci. (Regul. Ed.)* *16*, 122–128.
- Hong, G., and Lieber, C.M. (2019). Novel electrode technologies for neural recordings. *Nat Rev Neurosci* *20*, 330–345.

Hurtubise, J.L., and Howland, J.G. (2017). Effects of stress on behavioral flexibility in rodents. *Neuroscience* 345, 176–192.

Hvoslef-Eide, M., Nilsson, S.R.O., Saksida, L.M., and Bussey, T.J. (2016). Cognitive Translation Using the Rodent Touchscreen Testing Approach. *Curr Top Behav Neurosci* 28, 423–447.

Izquierdo, A., Brigman, J.L., Radke, A.K., Rudebeck, P.H., and Holmes, A. (2017). The neural basis of reversal learning: An updated perspective. *Neuroscience* 345, 12–26.

Jarvers, C., Brosch, T., Brechmann, A., Woldeit, M.L., Schulz, A.L., Ohl, F.W., Lommerzheim, M., and Neumann, H. (2016). Reversal Learning in Humans and Gerbils: Dynamic Control Network Facilitates Learning. *Front Neurosci* 10.

Jendryka, M., Palchaudhuri, M., Ursu, D., van der Veen, B., Liss, B., Kätsel, D., Nissen, W., and Pekcec, A. (2019). Pharmacokinetic and pharmacodynamic actions of clozapine-N-oxide, clozapine, and compound 21 in DREADD-based chemogenetics in mice. *Scientific Reports* 9, 4522.

Juczewski, K., Koussa, J.A., Kesner, A.J., Lee, J.O., and Lovinger, D.M. (2020). Stress and behavioral correlates in the head-fixed method. *BioRxiv* 2020.02.24.963371.

Kaczmarek, L., and Levitan, I. (1987). Neuromodulation : the biochemical control of neuronal excitability.

Kane, G.A., Vazey, E.M., Wilson, R.C., Shenhav, A., Daw, N.D., Aston-Jones, G., and Cohen, J.D. (2017). Increased locus coeruleus tonic activity causes disengagement from a patch-foraging task. *Cogn Affect Behav Neurosci* 17, 1073–1083.

Karlsson, M.P., Tervo, D.G.R., and Karpova, A.Y. (2012). Network Resets in Medial Prefrontal Cortex Mark the Onset of Behavioral Uncertainty. *Science* 338, 135–139.

Katz, P., and Calin-Jageman, R. (2008). Neuromodulation. In *Encyclopedia of Neuroscience*, pp. 497–503.

Kennerley, S.W., Walton, M.E., Behrens, T.E.J., Buckley, M.J., and Rushworth, M.F.S. (2006). Optimal decision making and the anterior cingulate cortex. *Nat Neurosci* 9, 940–947.

Kennerley, S.W., Behrens, T.E.J., and Wallis, J.D. (2011). Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nat Neurosci* 14, 1581–1589.

Khamassi, M., Lallée, S., Enel, P., Procyk, E., and Dominey, P.F. (2011). Robot Cognitive Control with a Neurophysiologically Inspired Reinforcement Learning Model. *Front Neurorobot* 5.

Khamassi, M., Enel, P., Dominey, P.F., and Procyk, E. (2013). Medial prefrontal cortex and the adaptive regulation of reinforcement learning parameters. *Prog Brain Res* 202, 441–464.

Knight, M.H., Jaarsveld, A.S.V., and Mills, M.G.L. (1992). Allo-suckling in spotted hyenas (*Crocuta crocuta*): an example of behavioural flexibility in carnivores. *African Journal of Ecology* 30, 245–251.

Koechlin, E. (2016). Prefrontal executive function and adaptive behavior in complex environments. *Curr. Opin. Neurobiol.* 37, 1–6.

KOIZUMI, R., KIYOKAWA, Y., MIKAMI, K., ISHII, A., TANAKA, K.D., TANIKAWA, T., and TAKEUCHI, Y. (2018). Structural differences in the brain between wild and laboratory rats (*Rattus norvegicus*): Potential contribution to wariness. *J Vet Med Sci* 80, 1054–1060.

Kolling, N., and O'Reilly, J. (2018). State-change decisions and dorsomedial prefrontal cortex: the importance of time. *Current Opinion in Behavioral Sciences* 22.

Kolling, N., Behrens, T.E.J., Mars, R.B., and Rushworth, M.F.S. (2012). Neural Mechanisms of Foraging. *Science* 336, 95–98.

- Kolling, N., Behrens, T., Wittmann, M.K., and Rushworth, M. (2016a). Multiple signals in anterior cingulate cortex. *Curr Opin Neurobiol* 37, 36–43.
- Kolling, N., Wittmann, M.K., Behrens, T.E.J., Boorman, E.D., Mars, R.B., and Rushworth, M.F.S. (2016b). Value, search, persistence and model updating in anterior cingulate cortex. *Nat. Neurosci.* 19, 1280–1285.
- Koralek, A.C., and Costa, R.M. (2019). Sustained Dopaminergic Plateaus and Noradrenergic Depressions Bias Transitions into Exploitative Behavioral States. *BioRxiv* 822650.
- Krebs, J.R., Kacelnik, A., and Taylor, P. (1978). Test of optimal sampling by foraging great tits. *Nature* 275, 27–31.
- Krug, K., Salzman, C.D., and Waddell, S. (2015). Understanding the brain by controlling neural activity. *Philosophical Transactions of the Royal Society B: Biological Sciences* 370, 20140201.
- Kwisthout, J., Wareham, T., and van Rooij, I. (2011). Bayesian Intractability Is Not an Ailment That Approximation Can Cure. *Cognitive Science* 35, 779–784.
- Ladouce, S., Donaldson, D.I., Dudchenko, P.A., and Ietswaart, M. (2017). Understanding Minds in Real-World Environments: Toward a Mobile Cognition Approach. *Front Hum Neurosci* 10.
- Lamothe, H., Schreiweis, C., Mallet, L., and Burguière, E. (2020). The Sapap3-knockout mouse model manifests a spectrum of repetitive behaviours. *BioRxiv* 2020.01.22.915215.
- Laubach, M., Amarante, L.M., Swanson, K., and White, S.R. (2018). What, If Anything, Is Rodent Prefrontal Cortex? *ENeuro* 5.
- Laureiro Martinez, D., Brusoni, S., Canessa, N., and Zollo, M. (2014). Understanding the exploration-exploitation dilemma: An fMRI study of attention control and decision-making performance. *Strategic Management Journal* 36.
- Laureiro-Martínez, D., Brusoni, S., and Zollo, M. (2010). The neuroscientific foundations of the exploration-exploitation dilemma. *Journal of Neuroscience, Psychology, and Economics* 3, 95–115.
- Lea, S.E.G., Chow, P.K.Y., Leaver, L.A., and McLaren, I.P.L. (2020). Behavioral flexibility: A review, a model, and some exploratory tests. *Learn Behav* 48, 173–187.
- Logiaco, L., Quilodran, R., Procyk, E., and Arleo, A. (2015). Spatiotemporal Spike Coding of Behavioral Adaptation in the Dorsal Anterior Cingulate Cortex. *PLOS Biology* 13, e1002222.
- Lottem, E., Banerjee, D., Vertechi, P., Sarra, D., Lohuis, M.O., and Mainen, Z.F. (2018). Activation of serotonin neurons promotes active persistence in a probabilistic foraging task. *Nat Commun* 9, 1000.
- Luksys, G., Gerstner, W., and Sandi, C. (2009). Stress, genotype and norepinephrine in the prediction of mouse behavior using reinforcement learning. *Nat. Neurosci.* 12, 1180–1186.
- Mainen, Z.F., Häusser, M., and Pouget, A. (2016). A better way to crack the brain. *Nature* 539, 159–161.
- Mallet, L., Polosan, M., Jaafari, N., Baup, N., Welter, M.-L., Fontaine, D., du Montcel, S.T., Yelnik, J., Chéreau, I., Arbus, C., et al. (2008). Subthalamic nucleus stimulation in severe obsessive-compulsive disorder. *N Engl J Med* 359, 2121–2134.
- Manning, E.E., Bradfield, L.A., and Iordanova, M.D. (2021). Adaptive behaviour under conflict: Deconstructing extinction, reversal, and active avoidance learning. *Neuroscience & Biobehavioral Reviews* 120, 526–536.
- Mar, A.C., Horner, A.E., Nilsson, S.R.O., Alsiö, J., Kent, B.A., Kim, C.H., Holmes, A., Saksida, L.M., and Bussey, T.J. (2013). The touchscreen operant platform for assessing executive function in rats and mice. *Nat Protoc* 8, 1985–2005.

- Marder, E., and Thirumalai, V. (2002). Cellular, synaptic and network effects of neuromodulation. *Neural Netw* *15*, 479–493.
- Mathis, M.W., and Mathis, A. (2020). Deep learning tools for the measurement of animal behavior in neuroscience. *Current Opinion in Neurobiology* *60*, 1–11.
- Matsumoto, K., Suzuki, W., and Tanaka, K. (2003). Neuronal correlates of goal-based motor selection in the prefrontal cortex. *Science* *301*, 229–232.
- McGuire, J.T., Nassar, M.R., Gold, J.I., and Kable, J.W. (2014). Functionally dissociable influences on learning rate in a dynamic environment. *Neuron* *84*, 870–881.
- Meder, B., Wu, C., Schulz, E., and Ruggeri, A. (2020). Development of directed and random exploration in children.
- van Meer, P., and Raber, J. (2005). Mouse behavioural analysis in systems biology. *Biochem J* *389*, 593–610.
- Mehlhorn, K., Newell, B., Todd, P., Lee, M., Morgan, K., Braithwaite, V., Hausmann, D., Fiedler, K., and Gonzalez, C. (2015). Unpacking the Exploration–Exploitation Tradeoff: A Synthesis of Human and Animal Literatures. *Decision* *2*.
- Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences* *7*, 134–140.
- Montague, P.R., Hyman, S.E., and Cohen, J.D. (2004). Computational roles for dopamine in behavioural control. *Nature* *431*, 760–767.
- Montgomery, K.L., Yeh, A.J., Ho, J.S., Tsao, V., Mohan Iyer, S., Grosenick, L., Ferenczi, E.A., Tanabe, Y., Deisseroth, K., Delp, S.L., et al. (2015). Wirelessly powered, fully internal optogenetics for brain, spinal and peripheral circuits in mice. *Nature Methods* *12*, 969–974.
- Morais, L.H., Schreiber, H.L., and Mazmanian, S.K. (2021). The gut microbiota–brain axis in behaviour and brain disorders. *Nature Reviews Microbiology* *19*, 241–255.
- Mosser, C.-A., Haqee, Z., Nieto-Posadas, A., Murai, K., Stifani, S., Williams, S., and Brandon, M.P. (2020). The McGill-Mouse-Marmoset Platform: A Standardized Approach for High-throughput Imaging of Neuronal Dynamics During Behavior. *BioRxiv* *2020.02.06.937573*.
- Nadim, F., and Bucher, D. (2014). Neuromodulation of Neurons and Synapses. *Curr Opin Neurobiol* *0*, 48–56.
- Nakajima, M., and Schmitt, L.I. (2020). Understanding the circuit basis of cognitive functions using mouse models. *Neuroscience Research* *152*, 44–58.
- Narayanan, N.S., Cavanagh, J.F., Frank, M.J., and Laubach, M. (2013). Common medial frontal mechanisms of adaptive control in humans and rodents. *Nat Neurosci* *16*, 1888–1895.
- Nath, T., Mathis, A., Chen, A.C., Patel, A., Bethge, M., and Mathis, M.W. (2019). Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nature Protocols* *14*, 2152–2176.
- Naudé, J., Tolu, S., Dongelmans, M., Torquet, N., Valverde, S., Rodriguez, G., Pons, S., Maskos, U., Mourot, A., Marti, F., et al. (2016). Nicotinic receptors in the ventral tegmental area promote uncertainty-seeking. *Nat Neurosci* *19*, 471–478.
- Neisser, U. (2014). *Cognitive Psychology: Classic Edition* (Psychology Press).
- Nicolas, S., and Ferrand, L. (2008). Chapitre 4. Développement d'une psychologie internationale : le cognitivisme. *Ouvertures psychologiques* 132–144.
- Nimpf, S., and Keays, D.A. (2017). Is magnetogenetics the new optogenetics? *EMBO J* *36*, 1643–1646.
- Niv, Y. (2020). On the Primacy of Behavioral Research for Understanding the Brain. pp. 134–151.

- Northcutt, R.G. (2002). Understanding Vertebrate Brain Evolution1. Integrative and Comparative Biology 42, 743–756.
- O'Driscoll, K., and Leach, J.P. (1998). "No longer Gage": an iron bar through the head. BMJ 317, 1673–1674.
- O'Reilly, J.X., Schüffelgen, U., Cuell, S.F., Behrens, T.E.J., Mars, R.B., and Rushworth, M.F.S. (2013). Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. PNAS 110, E3660–E3669.
- Osugi, T., Deng Kun, and Scott, S. (2005). Balancing Exploration and Exploitation: A New Algorithm for Active Machine Learning. In Fifth IEEE International Conference on Data Mining (ICDM'05), (Houston, TX, USA: IEEE), pp. 330–337.
- Otchy, T.M., Wolff, S.B.E., Rhee, J.Y., Pehlevan, C., Kawai, R., Kempf, A., Gobes, S.M.H., and Ölveczky, B.P. (2015). Acute off-target effects of neural circuit manipulations. Nature 528, 358–363.
- Paxinos, G., and Franklin, K.B.J. (2019). Paxinos and Franklin's the Mouse Brain in Stereotaxic Coordinates (5th Ed.).
- Pearce, J.M., and Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychological Review 87, 532–552.
- Pegel'man, S.G. (1966). [Adaptive nesting behaviour of some species of wild and laboratory animals]. Zh. Obshch. Biol. 27, 331–335.
- di Pellegrino, G., Ciaramelli, E., and Làdavas, E. (2007). The regulation of cognitive control following rostral anterior cingulate cortex lesion in humans. J Cogn Neurosci 19, 275–286.
- Piantadosi, P.T., Lieberman, A.G., Pickens, C.L., Bergstrom, H.C., and Holmes, A. (2019). A novel multichoice touchscreen paradigm for assessing cognitive flexibility in mice. Learn Mem 26, 24–30.
- Poirier, F.E. (1969). Behavioral flexibility and intertroop variation among nilgiri langurs (*Presbytis johnii*) of south India. Folia Primatol. 11, 119–133.
- Preuss, T.M. (1995). Do rats have prefrontal cortex? The rose-woolley-akert program reconsidered. J Cogn Neurosci 7, 1–24.
- Price, C.J., Tonn, W.M., and Paszkowski, C.A. (1991). Intraspecific patterns of resource use by fathead minnows in a small boreal lake. Can. J. Zool. 69, 2109–2115.
- Pritchard, D.J., Hurly, T.A., Tello-Ramos, M.C., and Healy, S.D. (2016). Why study cognition in the wild (and how to test it)? J Exp Anal Behav 105, 41–55.
- Quilodran, R., Rothé, M., and Procyk, E. (2008). Behavioral shifts and action valuation in the anterior cingulate cortex. Neuron 57, 314–325.
- Ragozzino, M.E. (2002). The effects of dopamine D(1) receptor blockade in the prelimbic-infralimbic areas on behavioral flexibility. Learn Mem 9, 18–28.
- Ragozzino, M.E., Mohler, E.G., Prior, M., Palencia, C.A., and Rozman, S. (2009). Acetylcholine activity in selective striatal regions supports behavioral flexibility. Neurobiol Learn Mem 91, 13–22.
- Rangel, A., Camerer, C., and Montague, P.R. (2008). Neuroeconomics: The neurobiology of value-based decision-making. Nat Rev Neurosci 9, 545–556.
- Remmelink, E., Aartsma-Rus, A., Smit, A.B., Verhage, M., Loos, M., and van Putten, M. (2016). Cognitive flexibility deficits in a mouse model for the absence of full-length dystrophin. Genes Brain Behav. 15, 558–567.
- Rescorla, R.A. (1971). Variation in the effectiveness of reinforcement and nonreinforcement following prior inhibitory conditioning. Learning and Motivation 2, 113–123.

Richmond, J. (2000). The 3Rs - Past, Present and Future. Scandinavian Journal of Laboratory Animal Science 27, 84–92.

Rmus, M., McDougle, S., and Collins, A. (2021). The role of executive function in shaping reinforcement learning. Current Opinion in Behavioral Sciences 38, 66–73.

Rodriguez, G.A., Barrett, G.M., Duff, K.E., and Hussaini, S.A. (2020). Chemogenetic attenuation of neuronal activity in the entorhinal cortex reduces A $\beta$  and tau pathology in the hippocampus. PLOS Biology 18, e3000851.

Roth, B.L. (2016). DREADDs for Neuroscientists. Neuron 89, 683–694.

Rushworth, M., Hadland, K., Gaffan, D., and Passingham, R. (2003). The Effect of Cingulate Cortex Lesions on Task Switching and Working Memory. Journal of Cognitive Neuroscience 15, 338–353.

Sakai, K. (2008). Task set and prefrontal cortex. Annu. Rev. Neurosci. 31, 219–245.

Sallet, J., Mars, R.B., Noonan, M.P., Neubert, F.-X., Jbabdi, S., O'Reilly, J.X., Filippini, N., Thomas, A.G., and Rushworth, M.F. (2013). The Organization of Dorsal Frontal Cortex in Humans and Macaques. J. Neurosci. 33, 12255–12274.

Saunders, T.L. (2020). The History of Transgenesis. Methods Mol. Biol. 2066, 1–26.

Schaefer, A.T., and Claridge-Chang, A. (2012). The surveillance state of behavioral automation. Curr. Opin. Neurobiol. 22, 170–176.

Schoenbaum, G., Nugent, S.L., Saddoris, M.P., and Setlow, B. (2002). Orbitofrontal lesions in rats impair reversal but not acquisition of go, no-go odor discriminations. Neuroreport 13, 885–890.

Scholl, J., and Klein-Flügge, M. (2018). Understanding psychiatric disorder by capturing ecologically relevant features of learning and decision-making. Behavioural Brain Research 355, 56–75.

Schreiweis, C., Feltgen, Q., Euvrard, M., Daunizeau, J., and Burguière, E. (2019). Motivational control of decision making in wild-type mice and a mouse model for compulsive-like behaviours.

Schubiger, M.N., Fichtel, C., and Burkart, J.M. (2020). Validity of Cognitive Tests for Non-human Animals: Pitfalls and Prospects. Front. Psychol. 11.

Schultz, W., Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction and reward. Science 275, 1593–1599.

Schulz, E., and Gershman, S.J. (2019). The algorithmic architecture of exploration in the human brain. Current Opinion in Neurobiology 55, 7–14.

Schulz, E., Bhui, R., Love, B., Brier, B., Todd, M.T., and Gershman, S. (2019). Structured, uncertainty-driven exploration in real-world consumer choice. Proceedings of the National Academy of Sciences.

Schwarz, C., Hentschke, H., Butovas, S., Haiss, F., Stützgen, M.C., Gerdjikov, T.V., Bergner, C.G., and Waiblinger, C. (2010). The Head-fixed Behaving Rat—Procedures and Pitfalls. Somatosens Mot Res 27, 131–148.

Schweighofer, N., and Doya, K. (2003). Meta-learning in Reinforcement Learning. Neural Networks 16, 5–9.

Schweighofer, N., Shishida, K., Han, C.E., Okamoto, Y., Tanaka, S.C., Yamawaki, S., and Doya, K. (2006). Humans Can Adopt Optimal Discounting Strategy under Real-Time Constraints. PLOS Computational Biology 2, e152.

Seamans, J.K., Lapish, C.C., and Durstewitz, D. (2008). Comparing the prefrontal cortex of rats and primates: insights from electrophysiology. Neurotox Res 14, 249–262.

- Seo, H., and Lee, D. (2007). Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. *J Neurosci* 27, 8366–8377.
- Seu, E., Lang, A., Rivera, R.J., and Jentsch, J.D. (2009). Inhibition of the norepinephrine transporter improves behavioral flexibility in rats and monkeys. *Psychopharmacology (Berl)* 202, 505–519.
- Sharma, P., and Pienaar, I.S. (2018). Chapter 24 - The Use of DREADDs for Dissecting the Contribution of Cellular and Neural Circuit Mechanisms in Models of Neurodegenerative Disease. In *Molecular-Genetic and Statistical Techniques for Behavioral and Neural Research*, R.T. Gerlai, ed. (San Diego: Academic Press), pp. 565–596.
- Sheets-Johnstone, M. (2015). Phenomenology and the life sciences: Clarifications and complementarities. *Progress in Biophysics and Molecular Biology* 119, 493–501.
- Sidor, M.M., and McClung, C.A. (2014). Timing matters: using optogenetics to chronically manipulate neural circuitry and rhythms. *Front Behav Neurosci* 8.
- Silvetti, M., Vassena, E., Abrahamse, E., and Verguts, T. (2018). Dorsal anterior cingulate-brainstem ensemble as a reinforcement meta-learner. *PLOS Computational Biology* 14, e1006370.
- Slotnick, B., and Katz, H. (1974). Olfactory Learning-Set Formation in Rats. *Science (New York, N.Y.)* 185, 796–798.
- Soltani, A., and Izquierdo, A. (2019). Adaptive learning under expected and unexpected uncertainty. *Nature Reviews Neuroscience* 20, 635–644.
- Sorge, R.E., Martin, L.J., Isbester, K.A., Sotocinal, S.G., Rosen, S., Tuttle, A.H., Wieskopf, J.S., Acland, E.L., Dokova, A., Kadoura, B., et al. (2014). Olfactory exposure to males, including men, causes stress and related analgesia in rodents. *Nat. Methods* 11, 629–632.
- Spangenberg, E.M.F., and Keeling, L.J. (2016). Assessing the welfare of laboratory mice in their home environment using animal-based measures--a benchmarking tool. *Lab. Anim.* 50, 30–38.
- Spencer, H., and Desclos, A. (1909). *Qu'est-ce que la morale?* (Paris, France: Schleicher frères).
- Squire, L.R. (2009). The Legacy of Patient H.M. for Neuroscience. *Neuron* 61, 6–9.
- Staddon, J.E.R. (2016). *Adaptive Behavior and Learning* (Cambridge: Cambridge University Press).
- Steurer, M.M., Aust, U., and Huber, L. (2012). The Vienna comparative cognition technology (VCCT): an innovative operant conditioning system for various species and experimental procedures. *Behav Res Methods* 44, 909–918.
- Sultana, R., Ogundele, O.M., and Lee, C.C. Contrasting characteristic behaviours among common laboratory mouse strains. *Royal Society Open Science* 6, 190574.
- Sun, Y., O'Reilly, R.C., Bhattacharyya, R., Smith, J.W., Liu, X., and Wang, H. (2015). Latent structure in random sequences drives neural learning toward a rational bias. *PNAS* 112, 3788–3792.
- Sutton, R.S., and Barto, A.G. (2018). *Reinforcement Learning: An Introduction*, Second edition.
- Swart, Jac.A.A. (2004). The Wild Animal as a Research Animal. *Journal of Agricultural and Environmental Ethics* 17, 181–197.
- Tervo, D.G.R., Proskurin, M., Manakov, M., Kabra, M., Vollmer, A., Branson, K., and Karpova, A.Y. (2014). Behavioral Variability through Stochastic Choice and Its Gating by Anterior Cingulate Cortex. *Cell* 159, 21–32.
- Thompson, W.R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika* 25, 285–294.

Tolman, E.C. (1948). Cognitive maps in rats and men. *Psychological Review* *55*, 189–208.

Topalovic, U., Aghajan, Z.M., Villaroman, D., Hiller, S., Christov-Moore, L., Wishard, T.J., Stangl, M., Hasulak, N.R., Inman, C., Fields, T.A., et al. (2020). Wireless Programmable Recording and Stimulation of Deep Brain Activity in Freely Moving Humans. *BioRxiv* 2020.02.12.946434.

Torquet, N., Marti, F., Campart, C., Tolu, S., Nguyen, C., Oberto, V., Benallaoua, M., Naudé, J., Didienne, S., Debray, N., et al. (2018). Social interactions impact on the dopaminergic system and drive individuality. *Nature Communications* *9*, 3081.

Tran, F.H., Spears, S.L., Ahn, K.J., Eisch, A.J., and Yun, S. (2020). Does chronic systemic injection of the DREADD agonists clozapine-N-oxide or Compound 21 change behavior relevant to locomotion, exploration, anxiety, and depression in male non-DREADD-expressing mice? *Neuroscience Letters* *739*, 135432.

Turner, K.M., Simpson, C.G., and Burne, T.H.J. (2017). BALB/c Mice Can Learn Touchscreen Visual Discrimination and Reversal Tasks Faster than C57BL/6 Mice. *Front Behav Neurosci* *11*.

Uddin, L.Q. (2021). Cognitive and behavioural flexibility: neural mechanisms and clinical considerations. *Nature Reviews Neuroscience* *1*–13.

Uylings, H.B., and van Eden, C.G. (1990). Qualitative and quantitative comparison of the prefrontal cortex in rat and in primates, including humans. *Prog Brain Res* *85*, 31–62.

Uylings, H.B.M., Groenewegen, H.J., and Kolb, B. (2003). Do rats have a prefrontal cortex? *Behav Brain Res* *146*, 3–17.

Valero-Cabré, A., Pascual-Leone, A., and Coubard, O.A. (2011). La stimulation magnétique transcrânienne (SMT) dans la recherche fondamentale et clinique en neuroscience. *Rev Neurol (Paris)* *167*, 291–316.

Vassena, E., Holroyd, C.B., and Alexander, W.H. (2017). Computational Models of Anterior Cingulate Cortex: At the Crossroads between Prediction and Effort. *Front Neurosci* *11*, 316.

Verharen, J.P.H., Kentrop, J., Vanderschuren, L.J.M.J., and Adan, R.A.H. (2019). Reinforcement learning across the rat estrous cycle. *Psychoneuroendocrinology* *100*, 27–31.

Vermaercke, B., Cop, E., Willems, S., D'Hooge, R., and Op de Beeck, H.P. (2014). More complex brains are not always better: rats outperform humans in implicit category-based generalization by implementing a similarity-based strategy. *Psychon Bull Rev* *21*, 1080–1086.

Vertechi, P., Lottem, E., Sarra, D., Godinho, B., Treves, I., Quendera, T., Oude Lohuis, M.N., and Mainen, Z.F. (2020). Inference-Based Decisions in a Hidden State Foraging Task: Differential Contributions of Prefrontal Cortical Areas. *Neuron* *106*, 166–176.e6.

Vogt, B.A., and Paxinos, G. (2014). Cytoarchitecture of mouse and rat cingulate cortex with human homologies. *Brain Struct Funct* *219*, 185–192.

Volle, E., and Levy, R. (2014). [Role of the prefrontal cortex in human behavioral adaptation]. *Med Sci (Paris)* *30*, 179–185.

Vrbanec, L., Matijević, V., and Guenther, A. (2021). Enhanced problem-solving ability as an adaptation to urban environments in house mice. *Proceedings of the Royal Society B: Biological Sciences* *288*, 20202504.

Warbrick, T., Reske, M., and Shah, N.J. (2014). Transferring Cognitive Tasks Between Brain Imaging Modalities: Implications for Task Design and Results Interpretation in fMRI Studies. *J Vis Exp*.

Watakabe, A., Ohtsuka, M., Kinoshita, M., Takaji, M., Isa, K., Mizukami, H., Ozawa, K., Isa, T., and Yamamori, T. (2015). Comparative analyses of adeno-associated viral vector serotypes 1, 2, 5, 8 and 9 in marmoset, mouse and macaque cerebral cortex. *Neurosci Res* *93*, 144–157.

Weber, F., and Dan, Y. (2016). Circuit-based interrogation of sleep control. *Nature* *538*, 51–59.

Welch, J.M., Lu, J., Rodriguez, R.M., Trotta, N.C., Peca, J., Ding, J.-D., Feliciano, C., Chen, M., Adams, J.P., Luo, J., et al. (2007). Cortico-striatal synaptic defects and OCD-like behaviours in Sapap3-mutant mice. *Nature* *448*, 894–900.

Whissell, P.D., Tohyama, S., and Martin, L.J. (2016). The Use of DREADDs to Deconstruct Behavior. *Front. Genet.* *7*.

Wilson, E.O. (1968). *The Dance Language and Orientation of Bees*. Karl von Frisch. Translated from the German edition (Berlin, 1965) by Leigh E. Chadwick. Belknap Press (Harvard University Press), Cambridge, Mass., 1967. xiv + 566 pp., illus. \$15. *Science* *159*, 864–865.

Wilson, C.R.E., and Gaffan, D. (2008). Prefrontal–Inferotemporal Interaction Is Not Always Necessary for Reversal Learning. *J. Neurosci.* *28*, 5529–5538.

Wilson, R.C., and Collins, A.G. (2019). Ten simple rules for the computational modeling of behavioral data. *ELife* *8*, e49547.

Wilson, R., Bonawitz, E., Costa, V., and Ebitz, R. (2021). Balancing exploration and exploitation with information and randomization. *Current Opinion in Behavioral Sciences* *38*, 49–56.

Wilson, R.C., Takahashi, Y.K., Schoenbaum, G., and Niv, Y. (2014a). Orbitofrontal cortex as a cognitive map of task space. *Neuron* *81*, 267–279.

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

Wisdom, T.N., Song, X., and Goldstone, R.L. (2013). Social Learning Strategies in Networked Groups. *Cogn Sci* *37*, 1383–1425.

Wise, R. (2004). Dopamine, learning and motivation. *Nature Reviews. Neuroscience* *5*, 483–494.

Wong, B.B.M., and Candolin, U. (2015). Behavioral responses to changing environments. *Behavioral Ecology* *26*, 665–673.

Wyart, V., and Koechlin, E. (2016). Choice variability and suboptimality in uncertain environments. *Current Opinion in Behavioral Sciences* *11*, 109–115.

Xue, G., Chen, C., Lu, Z.-L., and Dong, Q. (2010). Brain Imaging Techniques and Their Applications in Decision-Making Research. *Xin Li Xue Bao* *42*, 120–137.

Yizhar, O., Fenno, L.E., Davidson, T.J., Mogri, M., and Deisseroth, K. (2011). Optogenetics in Neural Systems. *Neuron* *71*, 9–34.

Yu, A., and Cohen, J.D. (2009). Sequential effects: Superstition or rational behavior? *Advances in Neural Information Processing Systems 21 - Proceedings of the 2008 Conference 21*, 1873–1880.

Yu, A.J., and Dayan, P. (2005). Uncertainty, Neuromodulation, and Attention. *Neuron* *46*, 681–692.

Zhang, Y., Castro, D.C., Han, Y., Wu, Y., Guo, H., Weng, Z., Xue, Y., Ausra, J., Wang, X., Li, R., et al. (2019). Battery-free, lightweight, injectable microsystem for in vivo wireless pharmacology and optogenetics. *PNAS* *116*, 21427–21437.

Ministère de l'Enseignement supérieur et de la recherche, « Utilisation des animaux à des fins scientifiques, Enquête statistique sur l'utilisation des animaux à des fins scientifiques », 1ère publication : 9.10.2014 - Mise à jour : 25.01.2021, <https://www.enseignementsup-recherche.gouv.fr/cid70613/enquete-statistique-sur-l-utilisation-des-animaux-a-des-fins-scientifiques.html>

Inkawu Vervet Project, <https://inkawuvervetproject.weebly.com/>

20 minutes, M.F, Paris: 3.000 poubelles anti-rats et anti-corneilles vont être (enfin) déployées en 2019, Publié, 20/02/2019 à 18h01, <https://www.20minutes.fr/paris/2455831-20190220-paris-3000-poubelles-anti-rats-anti-corneilles-vont-etre-enfin-deployees-2019>

## **ANNEXES**

---



## **1. AUTRES PUBLICATIONS**

Benzina, N., Morgièvre, M., **Euvrard, M.**, Flores Alves Dos Santos, J., Pelissolo, A., and Mallet, L. (2020). Personalised 360o video exposure therapy for the treatment of obsessive-compulsive disorder: A single case study. *Fr. J. Psychiatry* 1, 31–38.

## **2. CONGRES**

### **XIIIth International Basal Ganglia Society Meeting - IBAGS - (Biarritz) – Mai 2021**

- 1) The Behavioural and Ecological Automatized operanT Box (BEATBox): a flexible and low-cost system for high-throughput behavioural data acquisition (POSTER)  
**M. Euvrard**, N. Benzina, Y. Travert, P. Tissier, G. Penderia, E. Burguière.
- 2) Motivational control of decision making in wild-type mice and a mouse model for compulsive-like behaviours (POSTER)  
C. Schreiweis, Q. Feltgen, **M. Euvrard**, J. Daunizeau\*, E. Burguière\*

### **Neuro OpenScience Workshop – (Paris) – Février 2019**

- 1) The Behavioural and Ecological Automatized operanT Box (BEATBox): a flexible and low-cost system for high-throughput behavioural data acquisition (POSTER)  
**M. Euvrard**, N. Benzina, Y. Travert, P. Tissier, G. Penderia, E. Burguière
- 2) Co-animateuse de l'atelier Arduino (<http://open-neuro.org/program/>)

## **3. MEDIATION SCIENTIFIQUE**

**2018 – 2021** : Membre de l'équipe de conception et de correction du concours Brain Bee France, organisé par l'Institut du Cerveau

**2018 - 2020** : Encadrante d'un projet de découverte de la recherche pour collégiens et lycéens dans le cadre de l'initiative « Chercheurs en herbe » sur le sujet : « *Le rôle des interneurones striataux à parvalbumine dans les troubles des comportements répétitifs* » (1 mercredi après-midi par semaine pendant 10 semaines)

**2020** : Ambassadrice DECLICS (Cercle FSER) à la rencontre de lycéens du lycée Honoré de Balzac (75017 PARIS) afin de présenter le milieu de la recherche

**2019** : Animatrice d'un atelier microscopie lors de l'opération INSERM « Destination Labo »

## **4. ACTIVITES D'ENSEIGNEMENT**

**2019 – 2021** : Vacataire à l'Ecole Normale Supérieure Paris-Saclay : cours de Neurosciences en M2FESUP BGB et en M1 Biologie Santé, cours d'histoire des Sciences en L3 Biologie Santé

**2017 – 2019** : Monitrice (64h / an) à l'Ecole Normale Supérieure Paris-Saclay : cours, TP, et TD, en Neurosciences et Biologie cellulaire en M2FESUP BGB, L3 et M1 Biologie Santé

## 5. FORMATIONS

### **2019 :**

Formation réglementaire de base en chirurgie chez le rongeur de laboratoire (25h30), Vetsalius

### **2018 :**

Formation « Expérimentation animale, éthique et bientraitance, Formation des concepteurs de projets scientifiques utilisant des animaux » (10 jours), Sorbonne Université, INSERM

Formation à l'éthique scientifique dans le cadre de la journée scientifique des doctorants de 1<sup>ère</sup> année organisée par l'Ecole doctorale Cerveau, Cognition, Comportement

Formation « Communiquer sa recherche » (2 h), Master class organisée par la direction des affaires scientifiques et médicales de l'Institut du Cerveau

Formation « Développer son leadership conscient et légitimer son autorité » (21 h), Master class organisée par la direction des affaires scientifiques et médicales de l'Institut du Cerveau et le collectif XX initiative

# POSTFACE

---

## UNE ODE A LA PERSEVERANCE.

### Les deux grenouilles et le seau de lait *(librement mis en vers d'après un conte que me racontait ma Mémé)*

Deux grenouilles par la saison sèche assoiffées,  
Contraintes de quitter leur mare desséchée,  
    Parvinrent à une ferme coquette,  
    Oasis méritée après leur longue quête.  
    La traite du soir était terminée,  
    Et de l'étable s'exhalait,  
    La douce et chaude odeur du lait.  
Nos deux grenouilles en gourmandes invétérées,  
    Plongent dans un seau pour s'y délecter.  
    Mais voilà que le festin terminé,  
Commères les grenouilles au fond sont coincées.  
    Nageant et sautant, sans cesse elles glissent,  
    Car les parois du seau sont hautes et lisses.  
    Le premier batracien alors se désespère :  
    « Mais que faisons-nous dans cette galère,  
        Nous n'arriverons point à en sortir,  
        Je préfère me résigner au pire. »  
    La première grenouille, en disant ces mots  
    Se laisse couler, mourir, tout au fond du seau.  
    Mais la seconde, elle, refuse d'abandonner.  
    Et de toutes ses forces, de toute son âme,  
    De tout son cœur, elle remue ses petites palmes.  
    Elle nage et saute, et saute et nage,  
    Tant et si bien que le blanc breuvage,  
        Finit par prendre en consistance,  
        Et la crème se transforme en beurre dense.  
Notre courageux amphibien peut alors prendre appui,  
    Sauter hors du sceau, regagner la vie.

M.E.

