Caractérisation des déficits d'apprentissage par renforcement et de motivation dans les troubles bipolaires : approche méta-analytique, comportementale et computationnelle

Arnaud Pouchon

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : ISCE - Ingénierie pour la Santé la Cognition et l'Environnement
Spécialité : PCN - Sciences cognitives, psychologie et neurocognition
Unité de recherche : Grenoble Institut des Neurosciences

Caractérisation des déficits d'apprentissage par renforcement et de motivation dans les troubles bipolaires : approche méta-analytique, comportementale et computationnelle

Characterization of reinforcement learning deficits and motivation in bipolar disorder: a meta-analytic, behavioral and computational approach

Présentée par :

Arnaud POUCHON

Direction de thèse :

Julien BASTIN DIRECTEUR DE RECHERCHE, INSERM DELEGATION AUVERGNE-RHONE-ALPES	Directeur de thèse
Mircea POLOSAN PROFESSEUR DES UNIV. - PRATICIEN HOSP., UNIVERSITE GRENOBLE ALPES	Co-directeur de thèse
Clément DONDE PROFESSEUR DES UNIV. - PRATICIEN HOSP., UNIVERSITE GRENOBLE ALPES	Co-directeur de thèse

Rapporteurs :

ANNE SAUVAGET
PROFESSEURE DES UNIVERSITES - PRATICIENNE HOSPITALIERE, UNIVERSITE DE NANTES
ARTHUR KALADJIAN
PROFESSEUR DES UNIVERSITES - PRATICIEN HOSPITALIER, UNIVERSITE DE REIMS - CHAMPAGNE ARDENNE

Thèse soutenue publiquement le **19 décembre 2023**, devant le jury composé de :

MIRCEA POLOSAN PROFESSEUR DES UNIVERSITES - PRATICIEN HOSPITALIER, UNIVERSITE GRENOBLE ALPES	Co-directeur de thèse
ANNE SAUVAGET PROFESSEURE DES UNIVERSITES - PRATICIENNE HOSPITALIERE, UNIVERSITE DE NANTES	Rapporteure
ARTHUR KALADJIAN PROFESSEUR DES UNIVERSITES - PRATICIEN HOSPITALIER, UNIVERSITE DE REIMS - CHAMPAGNE ARDENNE	Rapporteur
MONICA BACIU PROFESSEURE DES UNIVERSITES, UNIVERSITE GRENOBLE ALPES	Présidente
EMMANUEL POULET PROFESSEUR DES UNIVERSITES - PRATICIEN HOSPITALIER, UNIVERSITE LYON 1 - CLAUDE BERNARD	Examinateur

Invités :

JULIEN BASTIN
DIRECTEUR DE RECHERCHE, INSERM DELEGATION AUVERGNE-RHONE-ALPES
CLEMENT DONDE COQUELET
PROFESSEUR DES UNIVERSITES - PRATICIEN HOSPITALIER, UNIVERSITE GRENOBLE ALPES

Remerciements

Je tiens tout d'abord à remercier les membres du jury, les rapporteurs (le **Pr. Arthur KALADJIAN** et la **Pr. Anne SAUVAGET**) et les examinateurs (la **Pr. Monica BACIU** et le **Pr. Emmanuel POULET**) pour avoir accepté d'évaluer ce travail. Merci de votre confiance pour CerCog@UGA et la section STEP de l'AFPBN.

Je remercie également mes directeurs/encadrant de thèse. Merci au **Pr. Mircea POLOSAN**, mon maître, merci pour ta confiance et tes encouragements tout au long de mon parcours. Merci au **Dr. Julien BASTIN** pour ton soutien et ta bienveillance sans limite, j'espère que nous continuerons longtemps à travailler ensemble. Merci au **Pr. Clément DONDE**, ancien co-CCA, pour ton soutien, j'espère également qu'on aura de belles années de travail ensemble devant nous. Merci à vous d'avoir cru en moi (probablement plus que moi-même) et de m'avoir poussé à aller jusqu'au bout dans les moments de doute.

Je tiens à remercier vivement le **Pr. Fabien VINCKIER**, merci pour tes conseils tout au long de ce travail, j'espère que nous serons amenés à davantage collaborer dans le futur. Merci à ton équipe de l'ICM pour le coup de main. Je remercie également les membres de mon CSI (le **Pr. Bruno ETAIN** et le **Dr. Mehdi KHAMASSI**) pour leurs conseils.

Je remercie l'équipe 21 du GIN, notamment mes co-doctorants pour votre aide et vos conseils quand j'en avais besoin, merci à **Audrey**, **Elodie**, **Clarissa** et **Maeva**. Merci à **Romane** pour ton aide sans faille, tu as toute ma gratitude. Merci aux stagiaires qui m'ont aidé. Merci à **Yelena**, **Mylène** et **Léa**. Et une mention spéciale à **Michi**.

Un grand merci à l'équipe du Centre expert des troubles bipolaires pour votre soutien, en commençant par **Maud**. Merci à **Solange**, **Pauline**, **Benjamin**, **Ariane**, **Quentin**. Merci à toute l'équipe de l'UTNS, merci **Marie**, **Evelyne**, **Deborah**, **Christine**, **Mégane**, **Leslie**, et **Annabelle**. Merci à tous les participants pour leur temps.

Je tiens à remercier mes collègues du service de psychiatrie de l'adulte du CHU Grenoble-Alpes, merci à **Gaëlle**, **Jérôme**, **Claire**, **Anne-Sophie**, **Marc**. Une mention spéciale à **Antoine**, simplement merci d'être là et d'être toi, tu as grandement contribué à ce travail. Merci aux internes de l'UTNS (**Léonard**, **Marine** et **Maxime**) pour votre travail.

Enfin, et surtout, un immense merci à mes proches. Rien de tout cela n'aurait été possible sans votre soutien inconditionnel. Bien que ce ne soit pas une justification, j'espère que ce manuscrit vous aidera à comprendre mes absences (physiques ou parfois cognitives)...

Merci à mes amis, pour votre présence depuis toutes ces années. Merci à **Boris**, **Alex**, **Jerem**, **Matthieu**, **Kévin**, **Flo**, **PG**. J'espère que nous aurons encore de beaux moments ensemble devant nous. Mention spéciale à **Tristan** pour ton amitié depuis toutes ces années.

Merci à ma famille, à **ma tante** et **mes cousins** et **cousine**. A mes **grands-parents**. Merci d'avoir et d'être présents et soutenants malgré mes absences. Merci à ma belle-famille, merci à **Jérôme** et **Nathalie**, pour votre bienveillance, votre confiance et votre accueil dans la famille. Merci à **Justine** et **Marius** pour les moments passés ensemble.

Je tiens particulièrement à remercier mes frères, **Axel** et **Charles**. Vous êtes des modèles pour moi, chacun à votre façon. J'attends nos futurs moments passés ensemble avec impatience. Merci **Julie**, **Robin**, et **Côme** pour le bonheur que vous m'apportez.

Bien évidemment je tiens particulièrement à remercier **mes parents**. Merci pour tout... Vous savez à quel point je suis pudique, ces mots ne sauraient refléter tout l'amour et le respect que je vous porte. Tout ceci n'aurait pas été possible sans votre amour, votre soutien, votre confiance depuis toutes ces années.

Et pour finir je tiens à remercier **Margaux**, ma future épouse, pour absolument tout... Merci d'être présente, d'être toi, de me comprendre. Merci pour tous les merveilleux moments que nous avons passés ensemble. Je n'aurais jamais pu faire cela sans toi, merci pour ton amour, ton soutien, tes encouragements, ta confiance, ta patience et ta tolérance. A tous nos futurs projets.

Liste des études

Reinforcement learning in bipolar disorder: a systematic review and meta-analysis of behavioral studies

Arnaud Pouchon, Clément Dondé, Julien Bastin[1], Mircea Polosan[1]
En préparation

Reward and punishment learning deficits among bipolar disorder subtypes

Arnaud Pouchon, Fabien Vinckier, Clément Dondé, Maëlle CM Gueguen, Mircea Polosan[1], Julien Bastin[1]

- **Article publié** : **Pouchon A**, Vinckier F, Dondé C, Gueguen MC, Polosan M, Bastin J. Reward and punishment learning deficits among bipolar disorder subtypes. J Affect Disord. 2023 Nov 1;340:694-702. doi: 10.1016/j.jad.2023.08.075. Epub 2023 Aug 15. PMID: 37591352.
- **Communication affichée (poster)** : **Arnaud Pouchon**, Julien Bastin, Mircea Polosan. Sensitivity to reward and punishment: searching for a bipolar disorder subtype marker. Journée de la Recherche Médicale de Grenoble (JRM), 07 juin 2019, Grenoble.
- **Prix des chefs de clinique – assistants : Arnaud POUCHON (orateur)**, Fabien VINCKIER, Clément DONDE, Maelle GUEGUEN, Julien BASTIN, Mircea POLOSAN. La modélisation computationnelle pour capturer les différences d'apprentissage par renforcement selon le sous-type de trouble bipolaire. Congrès de l'Encéphale 2022, 25 janvier 2022, Paris.

Asymmetric influence of agency on mood during reward and punishment learning

Arnaud Pouchon, Fabien Vinckier, Marc Benhamou, Clément Dondé, Julien Bastin[1], Mircea Polosan[1]
En préparation

[1]Les auteurs ont contribué de façon égale à ce travail

Table de matières

« Non, n'essaie pas, fais-le ! Ou ne le fais pas. Mais il n'y a pas d'essai. »

– Maître Yoda

Il y a fort longtemps, dans une galaxie lointaine, la Force permettait de se passer de l'apprentissage par essai et erreur pour parvenir à ses objectifs. Aujourd'hui la science a remplacé la Force…

INTRODUCTION GENERALE

I

L'apprentissage par renforcement en psychologie expérimentale

Afin de mieux comprendre la problématique de ce travail de thèse, notamment pourquoi nous nous sommes intéressés à l'apprentissage par renforcement dans le trouble bipolaire, il m'apparaissait nécessaire d'exposer dans une première partie théorique les notions importantes de psychologie expérimentale sur l'apprentissage par renforcement. Nous verrons en effet dans cette partie ce qu'est le conditionnement, notamment instrumental, avant d'introduire la notion de renforcement et de récompense, puis les formalisations mathématiques de l'apprentissage par renforcement. Nous aborderons ensuite le trouble bipolaire. Un résumé sera proposé en fin de chaque partie.

1. Le conditionnement en psychologie expérimentale

a. L'histoire du conditionnement

Les premières expériences sur le conditionnement ont été menées en Russie à la fin du XIXe siècle par **Ivan Petrovic Pavlov** (1849-1936). Pavlov étudiait le fonctionnement

gastrique des chiens en collectant, mesurant, et analysant les réactions salivaires et alimentaires de ses chiens dans différentes conditions. Pavlov avait remarqué que les chiens commençaient à saliver abondamment avant même d'être nourris, voire même à partir du moment où ils entraient dans la pièce pour être nourris. Cette salivation s'est donc produite en réponse à un stimulus contextuel (leur présence dans la salle d'examen) et non plus physiologique (la présence de nourriture dans leur bouche). Grâce à ces observations, Pavlov développa une théorie en 1927, qu'il appela alors la « salivation psychique » et plus tard « **conditionnement classique** » ou « **conditionnement pavlovien** ». La consommation d'aliments déclenchant systématiquement la salivation avant même le début de l'expérience, il s'agissait donc d'une « **réponse inconditionnée** » (RI) (la salivation) à un « **stimulus inconditionné** » (SI) (la nourriture). A l'inverse, le simple fait d'être dans la pièce ne faisait pas saliver le chien, constituant un « stimulus neutre » (SN). Selon Pavlov, répéter l'expérience à ses chiens leur a permis d'apprendre à associer différents stimuli neutres (SN) qui étaient systématiquement présentés avant de manger (l'odeur et la vue des aliments, par exemple) à cette ingestion de nourriture (SI) (Pavlov, 1927). Ces stimuli auparavant neutres, parce qu'ils ne provoquaient pas de salivation par eux-mêmes, sont devenus des « **stimuli conditionnés** » (SC), du fait qu'ils provoquaient la salivation après répétition de l'expérience. La salivation observée en réponse à ces stimuli conditionnés est alors devenue une « **réponse conditionnée** » (RC) (**Figure I.1 : Décours temporel du protocole expérimental de Pavlov. (1)** Préalablement au conditionnement, un stimulus inconditionné (SI) déclenche systématiquement une réponse inconditionnée (RI). **(2)** Un stimulus neutre (SN) seul n'induit aucune réponse. **(3)** Pendant le conditionnement, la répétition de l'association SN puis SI induit tout le temps une RI. **(4)** Après conditionnement, un SN devenu stimulus conditionné (SC) permet donc à lui seul le déclenchement d'une réponse conditionnée (RC). *Figure issue de (Subedi, 2022).*).

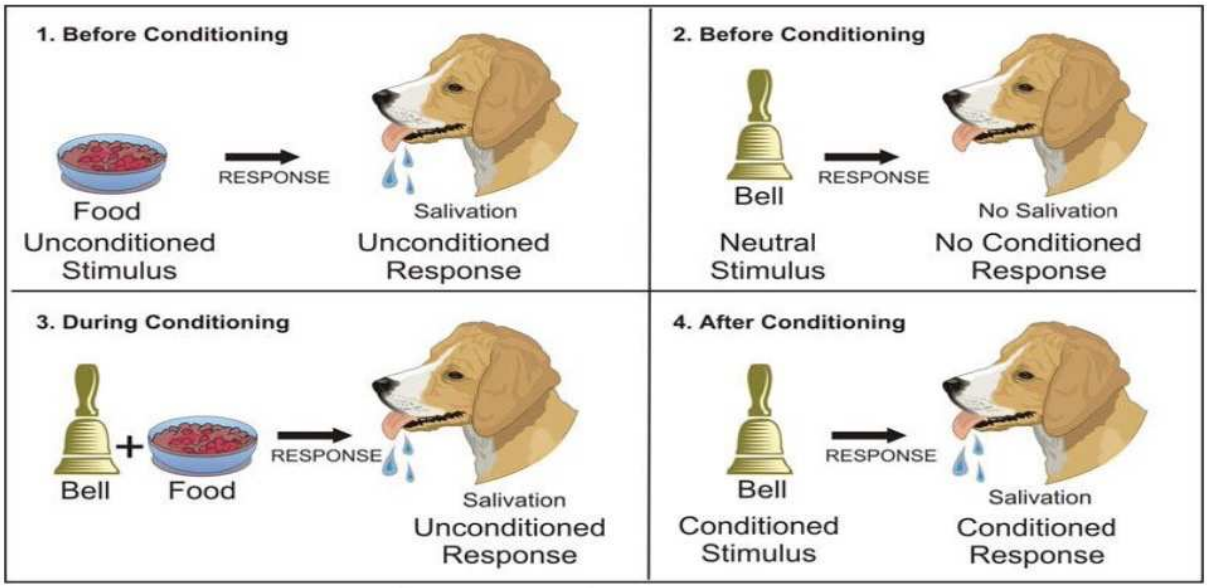

Figure I.1 : Décours temporel du protocole expérimental de Pavlov. (1) Préalablement au conditionnement, un stimulus inconditionné (SI) déclenche systématiquement une réponse inconditionnée (RI). **(2)** Un stimulus neutre (SN) seul n'induit aucune réponse. **(3)** Pendant le conditionnement, la répétition de l'association SN puis SI induit tout le temps une RI. **(4)** Après conditionnement, un SN devenu stimulus conditionné (SC) permet donc à lui seul le déclenchement d'une réponse conditionnée (RC). *Figure issue de (Subedi, 2022).*

Les études sur le conditionnement se sont poursuivies avec les travaux d'**Edward L. Thorndike** (1874-1949). La conception expérimentale de Thorndike consistait à mesurer le temps qu'il fallait à un chat pris au piège dans une boite pour tirer une ficelle et ainsi ouvrir la porte, s'échapper, et manger la nourriture visible à l'extérieure de la boite. Une différence fondamentale entre les conceptions expérimentales de Pavlov et de Thorndike est que dans le premier cas, la récompense était donnée indépendamment du comportement de l'animal, tandis que dans le second, **la récompense dépendait de la réponse comportementale**. Thorndike observa qu'alors que le chat découvrait le comportement cible par hasard, **après plusieurs essais, celui-ci mettait de moins en moins de temps pour sortir de la boite car il réalisait le comportement cible (tirer sur la corde) plus rapidement** (Figure I.2). Cet apprentissage par essai erreur est permis grâce à ce que Thorndike décrira comme une association « action-effet » ou « **instrumentalisation** » : une action menant à l'obtention d'une récompense verra sa probabilité de réalisation renforcée. Pour ce type de conditionnement, appelé « **conditionnement opérant** » ou « **conditionnement instrumental** », Thorndike décrivit

et publia ainsi la « **loi de l'effet** » en 1911, stipulant qu'une action a plus de chances de se répéter si elle apporte une satisfaction à l'individu (récompense), et aura tendance à être abandonnée si elle provoque une insatisfaction (punition) (Thorndike, 1911).

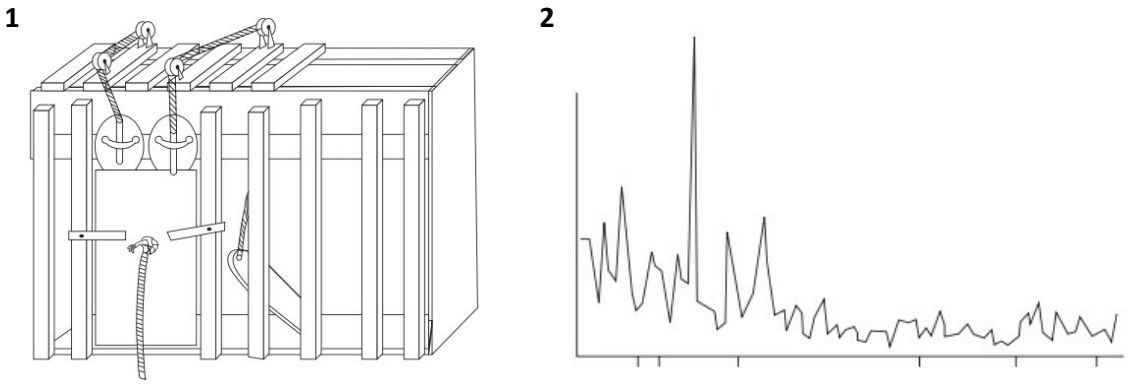

Figure I.2 : Expérience de Thorndike. (1) Une des boîtes à puzzle utilisées par Thorndike pour étudier l'acquisition de nouveaux comportements chez les chats. **(2)** Temps de fuite en fonction des essais pour un chat dans une expérience de boîte à puzzle. *Figures issues de (Blaisdell, 2008).*

Les études sur le conditionnement instrumental se sont par la suite poursuivies par les travaux de **Burrhus F. Skinner** (1904-1990), qui simplifia la recherche sur le conditionnement instrumental chez l'animal libre de mouvement grâce à sa « boite de Skinner » testée sur des rats (**Figure I.3**), permettant d'étudier les associations entre stimulus-action-renforcement de manière précise, et dont les variantes sont toujours utilisées à ce jour pour étudier le comportement animal. La contribution la plus importante de Skinner à l'étude du conditionnement opérant a été le concept de « **contingence du renforcement** » en 1938, pour définir l'environnement qui suscite le comportement. Trois facteurs entrent en jeu : 1) le **contexte** dans lequel le comportement se produit, 2) le **comportement** lui-même et 3) le **renforcement** qui suit le comportement. La contingence de renforcement correspond à **la probabilité de la relation de cause à effet entre le comportement et le renforcement associé** (Skinner, 1938). Plus la contingence comportement-renforcement est forte, plus le comportement sera renforcé rapidement.

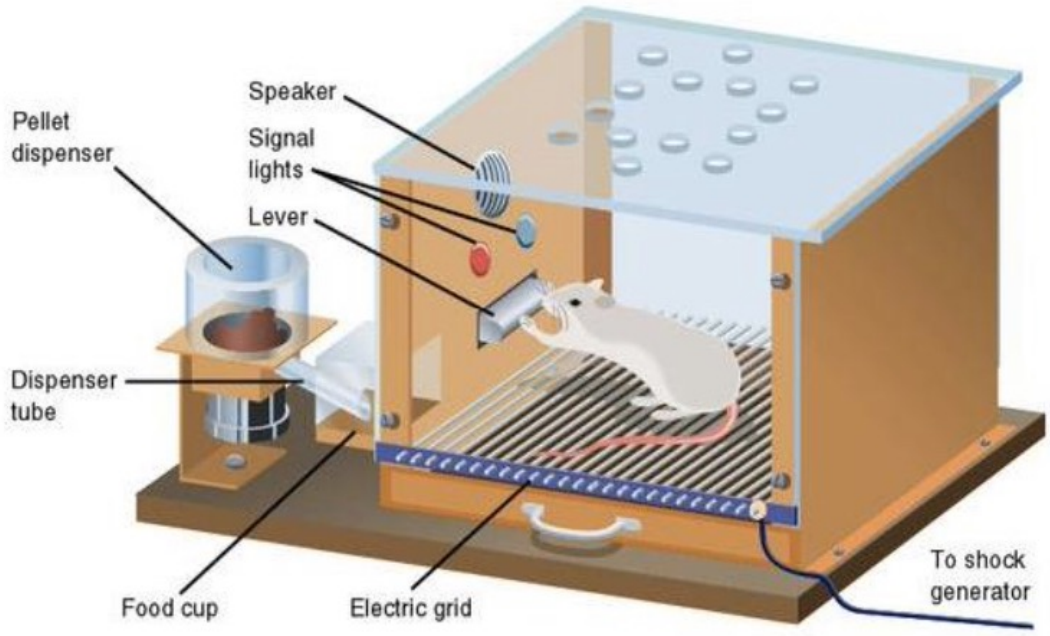

Figure I.3 : Expérience de Skinner. Un rat est placé dans une boite pour étudier le conditionnement opérant grâce à des récompenses (nourriture) et des punitions (plancher électrifié). *Figure issue de (Subedi, 2022).*

Parallèlement à cela, les comportementalistes Clark Hull et Kenneth Spence ont tenté de formaliser **mathématiquement** les lois régissant l'apprentissage par le conditionnement instrumental (Hull, 1943). Tout comme Thorndike qui observait que le chat sortait de plus en plus rapidement à force de répétition, et Skinner qui pointait du doigt l'influence de la contingence du renforcement sur la vitesse d'exécution du comportement cible menant au renforcement, ils ont mis en évidence **l'importance du temps pour réaliser l'action attendue pour étudier cet apprentissage**. Des expériences avec des rats qui devaient identifier un bras contenant de la nourriture dans un labyrinthe en T leur ont permis de montrer que **les performances des animaux suivent une courbe approximativement exponentielle au cours de l'apprentissage**. A partir de ces observations, plusieurs types de conditionnement ont été décrits.

b. Conditionnement classique et conditionnement instrumental

Le conditionnement classique et le conditionnement instrumental se distinguent par le fait que le conditionnement classique n'exige pas que l'agent prenne des mesures pour obtenir le renforçateur, alors que **dans le conditionnement instrumental, le renforçateur dépend du comportement de l'agent** (Dickinson, 1980; Staddon, 2016). Ainsi, ce n'est que dans le conditionnement instrumental que **les sujets influencent leur environnement pour déterminer l'apparition de renforçateurs**. De plus, la réponse dans le conditionnement classique est innée et dépend du type de renforçateur (la nourriture induit la salivation chez le chien). Dans le conditionnement instrumental, les réponses ne sont pas intrinsèquement « naturellement » associées aux renforçateurs. En d'autres termes, dans le conditionnement classique, le renforçateur (qui sera défini dans la partie suivante) (SI) détermine la forme de la réponse conditionnée (RC), alors que **dans le conditionnement instrumental, seules la force et la fréquence (probabilité) de la réponse, et non la forme, dépendent du renforçateur**.

Les recherches menées au cours des deux dernières décennies ont montré de manière convaincante que les deux types de structures de régulation existent dans l'apprentissage instrumental (Redgrave et al., 2010; Yin & Knowlton, 2006). En fait, le comportement instrumental peut être divisé en deux sous-classes : **orienté vers un but** et **habituel**, sur la base de cette distinction même. Il a été suggéré que le comportement d'apprentissage précoce est principalement déterminé par l'association entre la réponse et le résultat (*outcome*) (en anglais *Reponse - Outcome* ; R-O). Cette étape est également appelée « **orientée vers les buts** » **car son expression est motivée par les attentes des objectifs** (résultats). Au stade ultérieur, appelé « **habituel** », le même comportement est **principalement motivé par les associations entre** stimulus et réponse (S–R). A ce stade, les actions deviennent des habitudes qui s'expriment automatiquement dans un certain contexte, quel que soit le résultat attendu (**Figure I.4**). Le terme de « renforcement » semble donc prendre différents sens selon son origine.

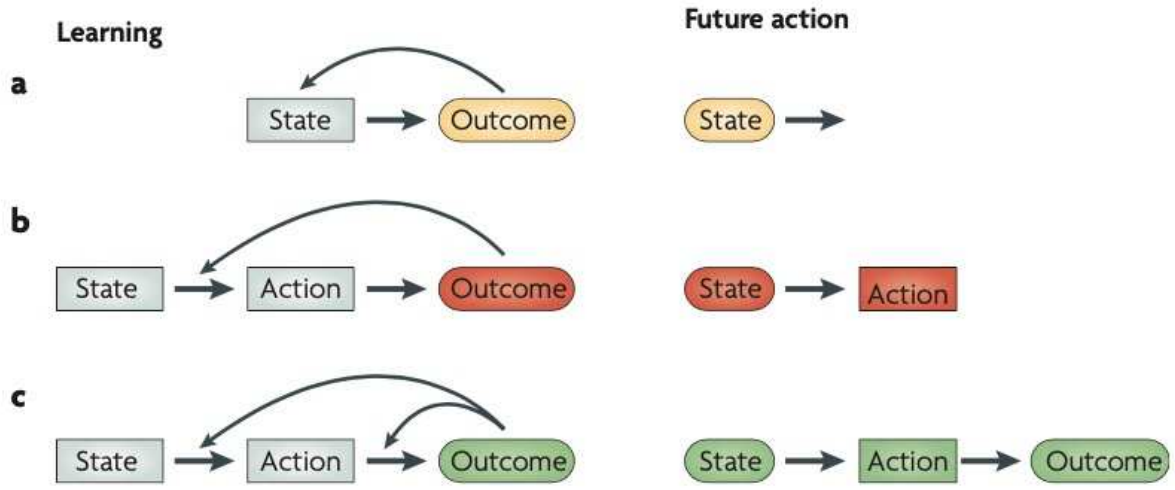

Figure I.4 : Les différents mécanismes d'apprentissage. (a) L'apprentissage pavlovien entraîne des réponses pavloviennes : l'association prédictive répétée d'un état ou d'un indice arbitraire avec un résultat motivant (une récompense ou une aversion) entraîne l'émission d'une réponse conditionnée et typiquement innée lorsque l'état est rencontré à l'avenir, anticipant ainsi le résultat de manière appropriée. **(b) L'apprentissage instrumental des habitudes** entraîne la formation d'habitudes : si une action est exécutée alors que le sujet se trouve dans un certain état et que cette action donne lieu à une récompense, l'action est renforcée de telle sorte que la rencontre de cet état à l'avenir rendra l'exécution de cette action plus probable. Si la valeur du résultat était aversive, l'action est inhibée à l'avenir. **(c) L'apprentissage instrumental orienté vers un but** aboutit à des actions orientées vers un but : si une action effectuée à partir d'un certain état aboutit à une récompense, une représentation explicite de la séquence est mémorisée, ce qui permet de guider les actions lorsque l'état est rencontré à l'avenir. *Figure issue de (Seymour et al., 2007).*

2. Le renforcement

a. Définition du renforcement

Le terme de **renforcement** a été introduit pour la première fois par Pavlov pour décrire le renforcement des associations entre les stimuli inconditionnés et conditionnés. Pour Pavlov, la nourriture (SI) était un facteur de renforcement, et la combinaison de tels stimuli avec des stimuli neutres (comme rentrer dans la pièce) formait le processus de renforcement et visait à renforcer (et à construire) des liens avec les SI (Pavlov, 1927). **Pour Thorndike, le terme faisait référence au renforcement (et à l'établissement) des associations Stimulus-Réponse-Résultat (*outcome*) (en anglais *Stimulus-Response-Outcome* ; S-R-O)** (Thorndike, 1911).

Le terme renforçateur est souvent associé au terme récompense, **la distinction entre les deux étant subtile. Cependant, ces deux termes peuvent (et doivent) être formellement distingués**. Une récompense est tout objet ou événement qui 1) produit des comportements d'approche, 2) augmente la probabilité et l'intensité d'actes comportementaux menant à de tels objets, et évoque des sentiments subjectifs de plaisir ou de déplaisir (Berridge & Robinson, 2003).

b. Récompense vs. renforcement

Les deux premières propriétés de la récompense sont liées au conditionnement. Les comportements d'approche sont des réponses généralement suscitées lors du conditionnement classique (comme la salivation) après que le SC est devenu un prédicteur du SI. Les changements de vitesse et d'intensité du comportement sont les variables expérimentales les plus importantes étudiées dans les paradigmes de conditionnement instrumental (Glickman & Schiff, 1967; Landauer, 1969). **La troisième propriété de la récompense évoque le plaisir subjectif, et n'a pas besoin d'induire de conditionnement car des stimuli non hédoniques peuvent également susciter des réponses instrumentales**. À ce jour, la ségrégation fonctionnelle entre ces traits a été démontrée à l'aide de diverses expérimentations, tant au niveau comportemental que neurobiologique (Cannon & Palmiter, 2003; S. Robinson et al., 2005). **Chez l'homme, une séparation a par exemple été observée entre les effets de renforcement et les effets hédoniques de la récompense dans la toxicomanie**. Il a ainsi été montré que ces patients continuaient à essayer d'obtenir la substance de l'abus même lorsqu'elle cesse d'être agréable (Everitt et al., 2001; White, 1996). **Ainsi, bien que les récompenses soient des renforçateurs positifs, les renforçateurs n'ont pas besoin d'être des récompenses**. De plus, la classe des renforçateurs comprend également les renforçateurs négatifs. **Le renforcement négatif est un renforcement résultant de la cessation d'un état de dégoût en cours.** Les renforçateurs négatifs sont aussi appelés « punisseurs », mais la différence entre les deux termes est presque identique à celle des renforçateurs positifs et des récompenses (Seymour et al., 2007).

Il existe donc **quatre types de renforcements possibles** dans le cadre de l'apprentissage par renforcement ou apprentissage par conditionnement instrumental :

- Renforcement appétitif positif ou récompense positive, qui **augmente** la probabilité de reproduction d'une réponse grâce à l'**ajout d'un stimulus appétitif** contingent à celle- ci.
- Renforcement appétitif négatif ou récompense négative, qui **augmente** la probabilité de reproduction d'une réponse grâce au **retrait d'un stimulus aversif** contingent à celle-ci.
- Renforcement aversif positif ou punition positive, qui **diminue** la probabilité de reproduction d'une réponse grâce à l'**ajout d'un stimulus appétitif** contingent à celle- ci.
- Renforcement aversif négatif ou punition négative, qui **diminue** la probabilité de reproduction d'une réponse grâce au **retrait d'un stimulus aversif** contingent à celle- ci.

Il peut cependant exister certaines **variations à la sensibilité au renforcement** (et donc à la récompense) au sein des individus, de manière physiologique ou pathologique, conditionnant les comportements d'approche ou d'évitement (Warr et al., 2021).

c. Sensibilité au renforcement et à la récompense

Selon les travaux de Kent Berridge et Terry Robinson, les mécanismes impliqués dans le traitement de ces récompenses se séparent en 2 phases : la **phase d'anticipation** (*Wanting*) et la **phase de consommation** (*Liking*) de la récompense**. Il est maintenant admis que les mécanismes déterminant l'intensité de l'anticipation de la récompense sont différents de ceux qui déterminent l'intensité associée à l'expérience du plaisir ou de consommation de cette récompense** (Berridge & Robinson, 2016). Lors de leur expérience initiale en 1989, Kent Berridge et ses collaborateurs avaient émis l'hypothèse que la suppression de dopamine au niveau cérébral chez les rats diminuerait l'expérience du plaisir associé aux stimuli agréables, partant du fait que la dopamine était à l'origine du

Liking (Berridge et al., 1989; Berridge & Robinson, 2016). Au contraire, cette expérience a montré que malgré la suppression quasi totale de dopamine au niveau cérébral, l'expérience du *Liking* chez les rats était restée complètement normale (Berridge et al., 1989), mais que **ces altérations dopaminergiques avaient inhibé toute motivation**. Les rats étaient alors retrouvés dans un état profond **d'apathie** et ne présentaient plus de volonté vers la consommation de la nourriture assimilée comme une récompense (Berridge et al., 1989; T. E. Robinson & Berridge, 1993). Durant les années suivantes, des études similaires ont été menées chez l'Homme, montrant que la suppression dopaminergique au niveau cérébral n'avait pas supprimé l'intensité associée à l'expérience du plaisir lors de la consommation de drogue telle que la cocaïne ou les amphétamines, malgré la diminution de l'appétence pour la consommation de drogue (Brauer & De Wit, 1997; Leyton et al., 2005). Ces mêmes observations ont été rapportées lors de la suppression dopaminergique chez les sujets sains ou ceux atteints de la maladie de Parkinson, constatant qu'il n'y avait pas de diminution de l'intensité du plaisir lors de la consommation de nourriture savoureuse (Hardman et al., 2012; Sienkiewicz-Jarosz et al., 2013). **Ainsi, il est suggéré que la dopamine serait plus étroitement liée à la médiation du *Wanting* plutôt que de celle du plaisir (*Liking*)** (Salamone & Correa, 2012).

Au niveau neurobiologique, le « désir » (ou la phase d'anticipation) est souvent déclenché par des stimuli liés à une récompense ou par une image précise de cette dernière (Berridge, 2012). Cette phase serait principalement médiée par le **système méso-cortico-limbique**, impliquant les projections dopaminergiques du mésencéphale vers des structures du cerveau antérieur comme le **noyau accumbens** ou certaines parties du **striatum** (Eldar & Niv, 2015). A l'inverse, l'expérience du plaisir ou phase de consommation de la récompense impliquerait un circuit de « **hotspots hédoniques** » (Berridge & Kringelbach, 2015). Ces hotspots se retrouvent au niveau du **cortex préfrontal**, du **cortex orbitofrontal** et certaines régions de **l'insula** (Kringelbach et al., 2003; Kringelbach & Berridge, 2010; Small et al., 2001). Il a donc été suggéré que ces deux phases du traitement de la récompense étaient traitées distinctement au niveau cérébral (Knutson et al., 2001) (**Figure I.5**).

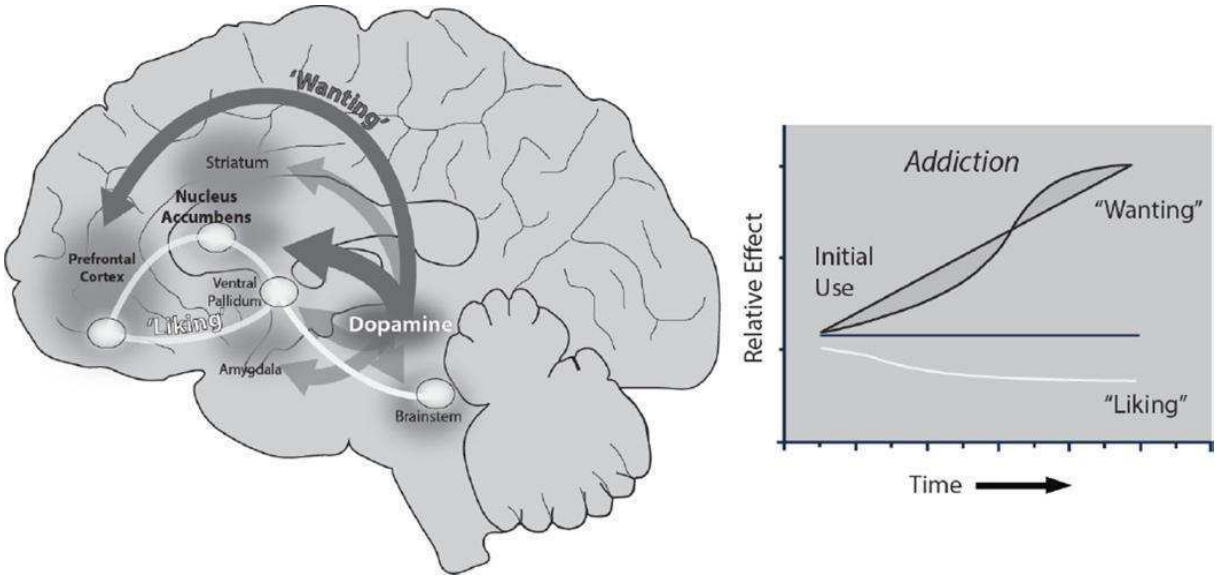

Figure I.5 : *Liking* et *Wanting* au niveau cérébral illustré dans l'addiction. Le *Wanting* est médié par un système comprenant des projections dopaminergiques (à gauche, en gris foncé), tandis que *Liking* est médié par un système dit de « hotspots hédoniques » (en blanc) (Berridge & Kringelbach, 2015). La théorie de la dépendance incitative-sensibilisation (à droite) montre comment le « désir » peut croitre au fil du temps indépendamment du « plaisir » lorsqu'un individu entretient des conduites toxicomaniques, en raison de la sensibilisation des systèmes mésolimbiques cérébraux (T. E. Robinson & Berridge, 1993). *Figure issue de (Berridge & Robinson, 2016).*

d. Le rôle du renforcement

La **théorie du conditionnement** vise à expliquer les relations qui sous-tendent l'apprentissage et la manière dont elles déterminent le comportement (Mowrer & Klein, 2000). Les traitements théoriques de ces conditionnements portent sur plusieurs questions. Notamment, Thorndike pensait que le chat avait appris à associer une séquence spécifique d'actions pour sortir de la boite. Ainsi, le rôle de l'objectif à atteindre (l'ouverture de la porte de la boite) est de mémoriser **l'association stimulus-réponse (S-R)** (Thorndike, 1933). Une autre option est qu'il aurait appris que cette séquence d'actions permettait l'ouverture de la porte, mettant en évidence une **association réponse-résultat (R-O)**. Skinner a réussi à concilier ces deux hypothèses grâce à la prise en compte du rôle d'un stimulus de contrôle dans ce qu'il a appelé une **contingence à trois termes** : le renforcement favorise une réponse en la présence de ce stimulus de contrôle ou « discriminant » (**association S-R-O**) (Skinner, 1938).

C'est le rôle du renforcement qui fait principalement la différence entre ces deux hypothèses (S-R et R-O). Pour l'hypothèse S-R, il joue un rôle dans l'apprentissage. Mais une fois celui-ci terminé, le comportement est globalement **indépendant du résultat** (*outcome*). Pour l'hypothèse R-O, il est proposé une **représentation directe de l'*outcome* au sein de l'association contrôlant le comportement**. L'ouverture de la porte d'une boite ne deviendrait plus désirable pour un chat enfermé dedans si un chien attendait devant la boite. Si seule la théorie S-R s'appliquait, le chat effectuerait la séquence d'actions pour ouvrir la porte, malgré la présence du chien. Mais d'après la théorie R-O, le chat aurait plutôt tendance à réprimer le comportement (de Wit & Dickinson, 2009; Dickinson, 1985; Seymour et al., 2007). Mais d'où découle l'apprentissage en lien avec ce renforçateur ? Nous allons voir cela dans la partie suivante.

3. L'apprentissage en lien avec le renforcement

a. Les conditions nécessaires à l'apprentissage

Nous avons vu que plusieurs facteurs nécessaires pour établir des associations entre les stimuli, les actions et les résultats ont été identifiés. Les conditions unanimement citées comme nécessaires au conditionnement sont la **contiguïté temporelle**, le **caractère aléatoire (contingence)** et **l'erreur de prédiction**.

L'importance de la contiguïté temporelle se reflète dans le fait que **la relation temporelle entre SC et SI**, ou réponse et résultat, détermine **la vitesse à laquelle ils sont associés**. Les stimuli sont appris de manière cohérente (Gibbon et al., 1977). La contingence est définie comme **la différence entre la probabilité de l'occurrence du SI en présence du SC et celle de l'occurrence du SI en l'absence du SC**. La contingence est donc une mesure simple de la co-occurrence statistique de deux stimuli, c'est à dire une mesure de la prédictivité du SC par rapport au SI (Rescorla, 1967). La troisième condition nécessaire à l'établissement d'une association est l'erreur de prédiction.

b. L'erreur de prédiction

Historiquement, l'importance des erreurs de prédiction a été montrée grâce à deux paradigmes clés du conditionnement, le **blocage** et **l'inhibition conditionnée**. Dans le **paradigme du blocage de Kamin**, un animal est exposé à un premier stimulus conditionné (SC1) prédisant l'occurrence d'un renforcement. Après avoir appris l'association appariant le SC1 et le SI, un second stimulus (SC2) est présenté en même temps que le SC1. Le décours temporel est donc le suivant : apparition simultanée de SC1 et de SC2, suivie de la délivrance du renforcement. Ainsi, **le SC1 et le SC2 sont tous deux prédictifs du SI**. Néanmoins, l'animal ne présente que peu voire aucune association entre le SC2 et le SI au cours des tests. **L'association SC2-SI a été bloquée par la première (SC1-SI)** puisque le SC1 permettait déjà de prédire complètement l'occurrence du SI, le SC2 n'apportant rien de nouveau. La délivrance du renforcement suite à l'apparition du stimulus SC2 n'induit pas d'erreur de prédiction positive (renforcement inattendu) car celui-ci était déjà entièrement prédit par l'apparition du stimulus SC1. **Les erreurs de prédiction n'apparaissent donc que lorsque qu'il y a une part d'inattendu associée au renforcement : renforcement plus ou moins important que ce qui était prévu**. Ces paradigmes de conditionnement ont ainsi montré que la **contiguïté temporelle** et la **contingence** ne suffisaient pas à induire un apprentissage associatif. Afin d'expliquer ces phénomènes, Rescorla et Wagner proposèrent un **modèle mathématique** simple incluant le concept d'erreur de prédiction comme facteur nécessaire au conditionnement (Rescorla, 1972).

4. Modélisation computationnelle de l'apprentissage par renforcement

a. Le modèle de Rescorla et Wagner

Le modèle Rescorla-Wagner (RW) tente d'expliquer **le changement de force associative (V) entre un SC et un SI ultérieur à la suite d'un essai de conditionnement**. De manière plus formelle, le modèle RW propose que le conditionnement se produit non pas parce que deux événements se produisent en même temps, mais parce que leur co-occurrence était inattendue d'après la force associative actuelle.

Au cours d'un essai d'apprentissage où deux stimuli *A* et *X* sont suivis par un SI, le modèle RW prédit que **les règles de changement de force associative** de *A* et *X* sont :

$$\Delta V_A = [\alpha_A \beta]\ (\lambda - V_{AX})$$

Et :

$$\Delta V_X = [\alpha_X \beta]\ (\lambda - V_{AX})$$

Où :

$$V_{AX} = V_A + V_X$$

Dans ces équations, λ est l'effet maximum que SI peut produire, représentant la limite de l'apprentissage. α et β ont des valeurs comprises entre 0 et 1, et sont des paramètres d'apprentissage qui dépendent respectivement de SC et SI. **α fait référence au taux d'apprentissage**, et **β représente la propension que les évènements inattendus ont à modifier les choix (également appelés « température »)**. Ces paramètres ont des valeurs fixes basées sur les propriétés physiques de SC et SI. Dans chaque essai, la force associative globale V_{AX} est comparée à λ, et la différence est considérée comme une erreur à corriger, grâce à un changement de la force associative (ΔV). Il s'agit donc d'un **modèle de correction des erreurs**.

Le modèle RW a fourni un cadre pour décrire l'effet de blocage de Kamin. Cela signifie que les appariements précédents du stimulus A et SI invalident les appariements ultérieurs de A et X avec le SI en produisant une association entre X et le SI. Le conditionnement préalable de A implique que VA soit proche de λ, donc lors d'un essai AX, puisque VX est nulle, VAX est proche de λ donc l'erreur correspond à $(\lambda - VAX)$ qui est proche de zéro. Ainsi, ΔVX est elle aussi proche de zéro donc VX change peu.

Ce modèle a depuis été largement étendu pour mieux décrire (1) l'apprentissage de manière plus générale, et (2) comment certaines régions du cerveau implémentent ces processus d'apprentissage. De nombreux modèles mathématiques du conditionnement ont été développés par la suite en prenant en compte l'influence d'autres phénomènes non expliqués par le modèle RW. Ces modèles sont

largement étudiés dans le domaine de **l'apprentissage automatique** (« Machine learning »), permettant d'aller plus loin dans la compréhension du conditionnement.

b. Définition de la modélisation computationnelle

La modélisation computationnelle est le fait de réaliser **un « modèle » mathématique décrivant un phénomène** (par exemple une fonction cognitive bien précise) par le biais d'une ou plusieurs équations, selon la complexité du phénomène. Elle est utilisée comme un outil complémentaire à la psychologie expérimentale et aux neurosciences cognitives pour tenter d'expliquer un phénomène (**Figure I.6**) (Rutledge & Adams, 2017). La modélisation permet de décrire les phénomènes observés dans la nature de façon synthétique, tout en bénéficiant de la précision des mathématiques. Ainsi, il est possible d'analyser les données d'une expérimentation en ***model-based*** , c'est-à-dire en confrontant directement les données aux variables du modèle. Les analyses sans modèle sont dites ***model-free***.

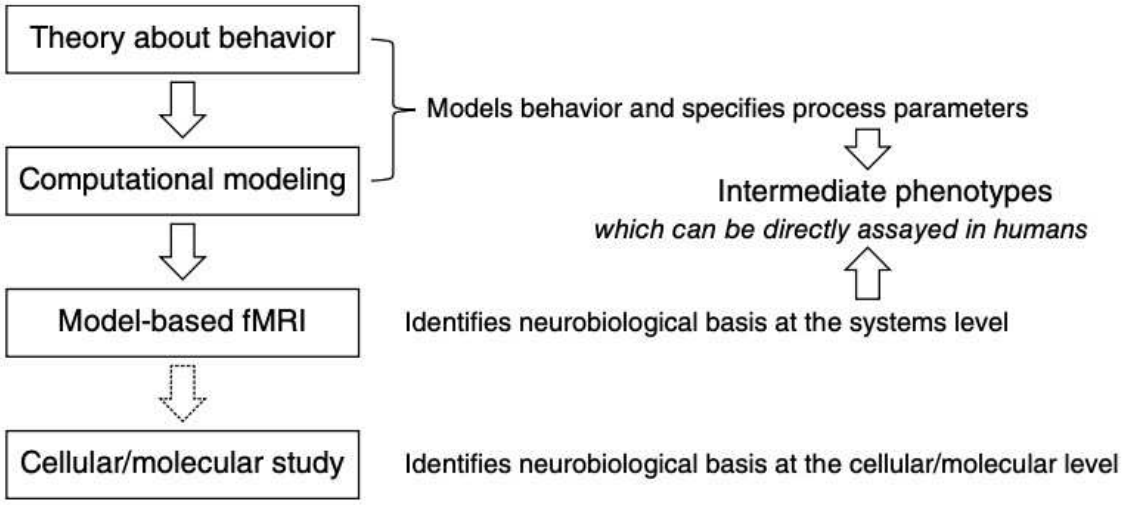

Figure I.6 : Illustration d'un exemple de quatre niveaux de l'approche computationnelle. Au 1er niveau, on trouve une théorie sur le comportement basée sur la littérature neurobiologique, qui spécifie les paramètres fondamentaux du comportement. Au 2ème niveau, la modélisation computationnelle impose une spécification explicite de la théorie en proposant des modèles mathématiques du comportement. Au 3ème niveau, en utilisant l'IRMf basée sur un modèle, le modèle le mieux adapté est régressé par rapport aux données IRMf des changements du signal BOLD au fil du temps pour comprendre le calcul

neuronal au niveau des systèmes. Au 4ème niveau, des études cellulaires et moléculaires identifient la base neurobiologique homogène sous-jacente des paramètres du processus afin de valider les preuves obtenues par l'IRMf basée sur un modèle au niveau des systèmes. *Figure issue de (Chen & Takahashi, 2017).*

Les modèles prédictifs (par rapport aux modèles descriptifs) vont chercher à décrire quelle est la **fonction sous-jacente** aux données observées, de façon à pouvoir expliquer pourquoi on s'attend à observer ces données dans telle situation et pas dans telle autre. On cherche donc à proposer une **explication mécanistique** du phénomène. C'est le neurobiologiste David Marr qui a fondé le champ des **neurosciences computationnelles** dans les années 80 (Marr, 2010). Il proposait 3 niveaux nécessaires à l'élaboration d'un modèle computationnel :

1. **Le niveau computationnel ou fonctionnel**, qui décrit quelle est la fonction du système étudié.
2. **Le niveau algorithmique**, qui décrit les représentations et les algorithmes que le système opère afin de réaliser sa fonction.
3. **Le niveau implémentationnel**, qui décrit comment ces algorithmes sont mis en œuvre dans un substrat naturel, donc en l'occurrence dans le substrat biologique qu'est le cerveau.

Un modèle computationnel est donc constitué (1) d'un **ensemble d'équations** qui décrivent le processus supposé sous-jacent au phénomène observé, (2) d'un **ensemble de paramètres** de ces équations et (3) d'un **ensemble de variables** qui représentent l'état du modèle à un instant *t* donné, ces variables étant ce qu'on peut lire comme étant le résultat des opérations.

c. Exemple pour l'apprentissage par renforcement

Par exemple, pour l'apprentissage par renforcement, il est possible de modéliser computationnellement l'expérience des chiens de Pavlov (Collins & Khamassi, 2021). Une **variable V décrit la force de l'association entre un SN et un SC**. A chaque fois que l'animal reçoit un SN après la présentation d'un SC, cette valeur d'association est augmentée

: $V_{(t+1)} = V_t + \boldsymbol{incrément}$**.** Il ne serait pas raisonnable de donner une valeur fixe à l'incrément, celle-ci pouvant augmenter infiniment (la quantité de salive produite par le chien atteint un plateau pendant l'apprentissage). Nous pouvons alors définir une **valeur maximale** pour la valeur d'association, par exemple 1, dont on considère qu'elle représente ce plateau. **Puis, à chaque essai, on incrémente l'association en proportion de la distance à cette valeur maximale**, $1 - V_t$. En d'autres termes, à mesure que V s'approchera de 1, on incrémentera V de plus en plus faiblement, de façon à la faire converger petit à petit vers 1. Cela nous donne l'équation suivante :

$$V_{(t+1)} = V_t + \alpha \times (1 - V_t)$$

Dans cette équation, **α** est un paramètre qui prend une valeur entre 0 et 1, et **indique à quelle vitesse on apprend** : si α = 0, on n'apprend pas du tout ; si α = 1, on atteint la valeur maximale après une seule expérience de la cloche suivie de la nourriture. Le problème d'une vitesse rapide d'apprentissage est qu'à la moindre absence fortuite de récompense on désapprendra d'un coup la valeur V. Le plus probable est une **valeur intermédiaire de α**. Il faudra alors **plusieurs répétitions du son de la cloche pour que le sujet soit confiant dans le fait que celui-ci annonce une récompense**.

Si le chien continue à être exposé à la cloche alors qu'elle n'est plus suivie de la délivrance de nourriture, il va se produire une phase dite **d'extinction**. L'association est petit à petit désapprise, le chien arrêtant progressivement de saliver lorsqu'il entend la cloche. Ce phénomène peut également être formalisé mathématiquement. Cette fois, **la valeur d'association diminue à chaque essai où le SN n'est pas présenté**, avec une valeur minimale fixée cette fois-ci à 0 : $V_{(t+1)} = V_t + \alpha \times (0 - V_t)$**.** Il faudra ainsi plusieurs absences de récompense après la présentation du son de cloche pour que l'animal considère que celui-ci n'est plus un prédicteur fiable de récompense. Ces deux équations peuvent être réécrites en une seule :

$$V_{(t+1)} = V_t + \alpha \times (r_t - V_t)$$

Où $\mathbf{r_t}$ **indique le renforcement**, ou la valeur appétitive de ce qui a été observé après le SN : 1 pour la nourriture, 0 pour l'absence de nourriture. Cette équation simple, parfois appelée ***delta-rule***, est au cœur des modèles d'apprentissage par renforcement. Elle décrit une variable importante : **l'erreur de prédiction** $\delta = r_t - V_t$ qui peut être interprétée computationnellement **comme la différence entre une récompense obtenue** $\mathbf{r_t}$ **et une récompense prédite** $\mathbf{V_t}$. Une forte erreur de prédiction conduit à changer la valeur d'association estimée plus qu'une faible erreur de prédiction.

En tant que généralisation du modèle de RW, les modèles d'apprentissage par renforcement (RL) n'estiment pas seulement l'association entre un état et le SN qui le suit juste après. **Ils essaient d'estimer la valeur future cumulée d'un état V(s)**, qui correspond à l'ensemble des renforcements que je m'attends à recevoir dans le futur, instant après instant, à partir de l'état actuel. Même lorsqu'elle arrive après un certain délai, **ceci permet au modèle d'apprendre à prédire la récompense** (Collins & Khamassi, 2021). Les récompenses immédiates sont pondérées plus fortement que celles des essais futurs, qui sont **dévaluées** (*discounted*) par un paramètre de dévaluation γ. On peut donc écrire $R_t = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \cdots$, avec $0 \leq \gamma \leq 1$. Si $\gamma = 0$, seules les récompenses immédiates sont prises en compte. Si $\gamma = 1$, une récompense dans dix jours aurait le même effet qu'une récompense immédiate. **En général, une valeur intermédiaire est appropriée.**

Pour estimer la valeur d'un état, les algorithmes de RL peuvent utiliser une simple extension du modèle RW. Comme dans le modèle RW, on incrémente la valeur estimée en proportion d'une **erreur de prédiction**, avec un **taux d'apprentissage** α contrôlant cette proportion :

$$V(s_t) \leftarrow V(s_t) + \alpha\delta_t$$

Pour tenir compte de l'estimation non seulement de la récompense immédiate, mais aussi de la récompense future, l'algorithme **TD-RL**, dit de **différence temporelle** (*temporal-difference reinforcement learning*), compare la prédiction $V(s_t)$ non seulement à la récompense

obtenue r_t, mais aussi à la valeur de l'état suivant $V(s_{t+1})$. Ainsi, l'erreur de prédiction TD est :

$$\delta_t = (r_t + \gamma\, V(s_{t+1})) - V(s_t)$$

Cette extension permet de tenir compte du futur, et résout par exemple les problèmes de RW pour le phénomène de conditionnement de deuxième ordre.

Si l'agent a une bonne estimation de la valeur des états, il peut sélectionner une action qui l'amènera au meilleur des états suivants - une stratégie appelée gloutonne (***greedy*** en anglais). Cependant, des théorèmes prouvent que l'algorithme ci-dessus n'est capable d'estimer correctement la valeur des états que si l'agent explore suffisamment : c'est-à-dire si l'agent choisit régulièrement une action qui a une valeur estimée non optimale. Cette exploration permet de s'assurer que tout l'espace des états est appris, et que l'agent ne s'est pas limité à un maximum local. Le degré d'exploration est souvent modélisé de manière **stochastique** : la stratégie *greedy* est choisie avec une certaine (forte) probabilité p, et d'autres actions sont choisies le reste du temps (c'est-à-dire avec une probabilité 1 − p). Plusieurs formules sont possibles, toutes paramétrisées par un certain degré de **stochasticité**.

Enfin, les algorithmes de **Q-learning** étendent l'apprentissage par différence temporelle à la prise de décision suivant une **mise à jour dynamique des associations état-action-conséquence** (on parle alors de **politique**), et en sélectionnant la plus avantageuse (Beaumont, 2018; Watkins & Dayan, 1992). La valeur d'une action repose alors à la fois sur les récompenses directes, mais aussi les **indirectes**. L'erreur de prédiction intègre alors la valeur de l'état-action espérée à l'instant suivant :

$$\delta_t = r + \gamma \times max_a Q_{t+1}(s,a) - Q_t(s,a)$$

Avec $Q_t(s,a)$ la valeur de l'association état *s* - action *a* à l'instant *t*, et $max_a Q_{t+1}(s,a)$ la valeur maximale espérée à l'instant *t+1*.

Ce type de modèle permet **d'extrapoler les opérations réalisées par un sujet en train de résoudre un problème similaire pour générer des hypothèses quantitatives** sur l'évolution de l'activité cérébrale à chaque nouvelle prise de décision. Entre le stimulus et l'action, l'agent doit manipuler les valeurs de chaque association stimulus-action-conséquences. Ces constructions abstraites sont appelées **variables cachées ou latentes**, par opposition aux observations expérimentales dont elles sont dérivées, et peuvent être utilisées pour identifier les structures cérébrales impliquées dans ces calculs (Beaumont, 2018; O'DOHERTY et al., 2007). La validité de ces modèles tient à leur capacité à **prédire le comportement**, mais également à rendre compte de l'activité neurale dans certains circuits, comme ceux impliquant l'activité des neurones dopaminergiques, d'abord chez le primate non humain (Bayer & Glimcher, 2005; Mirenowicz & Schultz, 1994; Suri & Schultz, 1999), puis chez l'Homme (Bray & O'Doherty, 2007; Pessiglione et al., 2006) (**Figure I.7**).

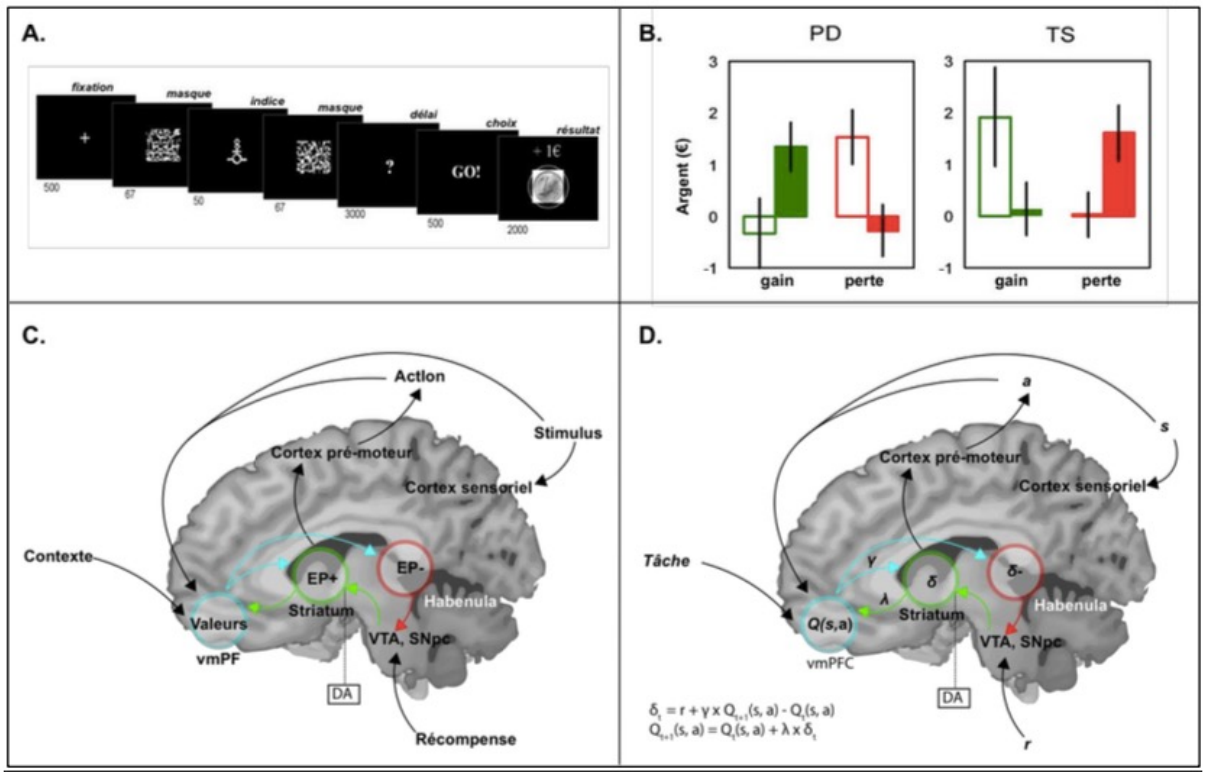

Figure I.7 : Exemple de paradigme d'apprentissage par renforcement et implémentation neurobiologique potentielle des algorithmes impliqués. Cette figure

présente un exemple de paradigme expérimental d'apprentissage par renforcement, avec des données chez des patients atteints de la maladie de Parkinson et de syndrome de Gilles de la Tourette, ainsi qu'une représentation simplifiée des régions cérébrales potentiellement impliquées dans l'apprentissage par renforcement. **A. Tache d'apprentissage par renforcement subliminal.** Dans cette tâche, les sujets devaient choisir l'action (Go ou No Go) permettant d'obtenir une récompense ou d'éviter une perte selon l'indice présenté de façon subliminale. Ils devaient donc apprendre par renforcement au cours des essais successifs les associations entre stimuli, actions, et récompenses *(Palminteri et al., 2009)*. **B. Résultats de la tâche présentée en (A) chez des patients atteints de la maladie de Parkinson et du syndrome de Gilles de la Tourette, avec et sans traitement.** La somme moyenne remportée par les sujets est présentée en ordonnée, selon la pathologie (PD : maladie de Parkinson, TS : syndrome de Gilles de la Tourette), la condition expérimentale (gain ou perte), et le traitement médicamenteux (L-Dopa pour la maladie de Parkinson, antagonistes dopaminergiques pour le syndrome de Gilles de la Tourette, barre vide : sans traitement, barre pleine : avec traitement). La non prise de L-Dopa ou la prise d'antagonistes dopaminergiques est associée à une altération de l'apprentissage par les gains et une augmentation de l'apprentissage par la perte, dans les deux groupes, et la prise de L-Dopa ou l'absence d'antagonistes dopaminergiques est associée à un rétablissement de l'apprentissage par les gains et une moindre sensibilité à la perte *(Palminteri et al., 2009)*. **C. Représentation simplifiée des régions encodant les variables essentielles au modèle d'apprentissage par renforcement**. Le cortex préfrontal ventro-médian (vmPFC) encode les valeurs subjectives associées aux actions étant donné le contexte et le stimulus. Le signal d'erreur de prédiction est encodé au niveau mésencéphalique par la dopamine, et projeté au niveau du striatum. L'habenula est associée aux erreurs de prédiction négatives et inhibe l'activité dopaminergique mésencéphalique. Au niveau du striatum, l'activité dopaminergique agit comme un signal de renforcement par des mécanismes de plasticité synaptique, qui facilite la sélection de l'action. **D. Représentation de l'algorithme de Q-learning sur les régions cérébrales potentiellement impliquées**. L'erreur de prédiction δ, est calculée comme la différence entre la récompense perçue, r, la valeur des états-actions à venir Qt+1(s, a), pondérée par un taux de décompte γ, et la valeur subjective Qt(s, a) de l'état-action actuel. Les valeurs Q(s, a) sont mises à jour selon un taux d'apprentissage λ. vmPFC = Cortex préfrontal ventro-médian, DA : dopamine, VTA : aire tegmentale ventrale, SNpc : substance noire pars compacta. *Figure issue de (Beaumont, 2018)*.

5. Résumé de la partie I

Nous avons vu dans cette partie que le **conditionnement instrumental** (ou conditionnement opérant) que nous étudions dans ce travail de thèse découle des travaux sur le conditionnement classique (ou pavlovien). Le conditionnement instrumental s'intéresse à **l'apprentissage en lien avec une action, en tenant compte des conséquences de cette dernière, rendant plus ou moins probable la reproduction de**

ce comportement (selon la **loi de l'effet**). Ainsi, ce type de conditionnement va dépendre à la fois du contexte dans lequel le comportement se produit, du comportement, ainsi que du renforcement qui suit le comportement, selon une certaine contingence, correspondant à la probabilité de la relation de cause à effet entre le comportement et le renforcement associé. Il faut noter que dans le conditionnement instrumental, **les réponses ne sont pas innées**, et que seules la force et la fréquence (probabilité) de la réponse, et non la forme, dépendent du renforçateur.

Il existe ainsi **quatre types de renforcement** : le renforcement appétitif positif, le renforcement appétitif négatif, le renforcement aversif positif, et le renforcement aversif négatif, en tenant compte de la valence (récompense vs. punition), et de l'ajout ou du retrait d'un stimulus contingent au renforcement. **Le renforcement va ainsi jouer un rôle dans l'apprentissage entre le stimulus et le résultat**. Pour cela, il faut toutefois que soient réunis trois conditions, à savoir la **contiguïté temporelle** (la vitesse à laquelle sont associés le stimulus et le résultat), le **caractère aléatoire** (avec la notion de **contingence**) et **l'erreur de prédiction** (la différence entre l'attente en lien avec un comportement et l'évaluation du résultat).

Parallèlement aux recherches de psychologie expérimentale, des **formalisations mathématiques** de ces comportements ont été développées initialement avec le modèle de Rescorla et Wagner. Ces modèles prédictifs tentent de formaliser mathématiquement un phénomène (ici l'apprentissage par renforcement), par le biais d'une ou plusieurs équations. Ils sont utilisés comme un outil complémentaire à la psychologie expérimentale et aux neurosciences cognitives pour tenter d'expliquer un phénomène (analyses en ***model-based***). Nous allons maintenant voir que l'apprentissage par renforcement se montre très pertinent pour étudier certaines pathologies neuropsychiatriques, en l'occurrence le **trouble bipolaire** pour ce travail de thèse, afin de mieux comprendre et caractériser les déficits cognitifs du trouble.

II

Le trouble bipolaire

Après avoir exposé les principales notions de l'apprentissage par renforcement, nous allons maintenant voir dans une deuxième sous-partie les enjeux de ce travail sur la population clinique que sont les personnes ayant un trouble bipolaire. Nous allons pour cela mettre en avant l'intérêt d'étudier ce processus cognitif dans cette population, notamment durant l'euthymie. Un résumé de cette partie sera également proposé.

1. Généralités

a. Epidémiologie

Le trouble bipolaire, anciennement connu sous le nom de psychose maniaco-dépressive (dont les nouvelles classifications internationales ont rendu le terme désuet), est un trouble de l'humeur chronique caractérisé par une alternance d'épisodes de **manie ou d'hypomanie** associés à des épisodes **dépressifs** (**Figure II.1**). Ces épisodes sont caractérisés par des atteintes de l'humeur et des émotions, mais également par des

altérations cognitives, **motivationnelles** et comportementales. Les troubles affectifs peuvent être classés le long d'un **spectre** défini par l'étendue et la gravité de l'élévation de l'humeur, du trouble unipolaire au trouble bipolaire de type I (BD-I) en passant par le trouble bipolaire de type II (BD-II) et les troubles bipolaires non spécifiés (BD NOS) selon les définitions du Manuel Diagnostique et Statistique des troubles mentaux (DSM) que nous allons voir (Grande et al., 2016).

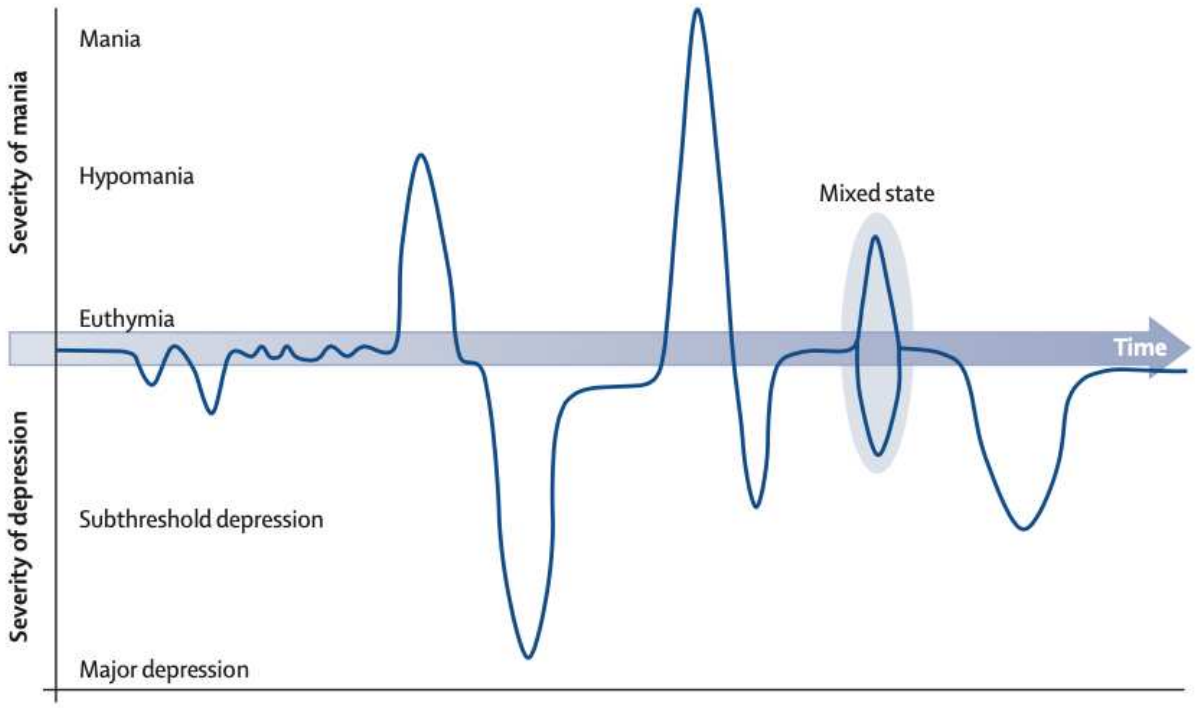

Figure II.1 : Exemple de l'évolution du trouble bipolaire au cours du temps. Le monitoring des variations de l'humeur au cours du temps sur un diagramme peut aider les cliniciens à diagnostiquer et à prendre en charge les patients atteints de trouble bipolaire. Selon l'intensité et le retentissement, les symptômes maniaques et hypomaniaques sont décrits au-dessus de l'état d'euthymie (état d'humeur « neutre »), tandis que les symptômes dépressifs sont décrits en dessous. *Figure issue de (Grande et al., 2016).*

Il toucherait environ **1 à 2% de la population** (Rowland & Marwaha, 2018) selon les définitions les plus restreintes, avec une prévalence sur une vie entière autour de 0,6% pour le trouble BD-I, de 0,4% pour le BD-II, et de 1,4% pour les BD NOS (Merikangas et al., 2011). Ces chiffres sont cependant très variables selon les études, passant **de 0,5% à environ 5%**, en lien avec une forte disparité du trouble selon les pays, selon les outils

psychométriques utilisés, les difficultés d'évaluation clinique, la prise en compte du **« spectre » bipolaire** (des formes sub-syndromiques et/ou pas clairement caractérisées) (**Figure II.2**), ainsi que les sur- et les sous-diagnostics (Ferrari et al., 2011). Il s'agit d'une pathologie **chronique** dont les décompensations aigues interviennent de manière **épisodique** (délimitées dans le temps) durant la vie entière et dont l'évolution est variable selon les individus. Le ratio serait de 2,5 épisodes dépressifs pour 1 épisode maniaque, hypomaniaque ou avec caractéristiques mixtes. Les épisodes durent de **4 à 13 mois** avec un **intervalle moyen de 12 mois** entre deux épisodes (Perlis et al., 2006).

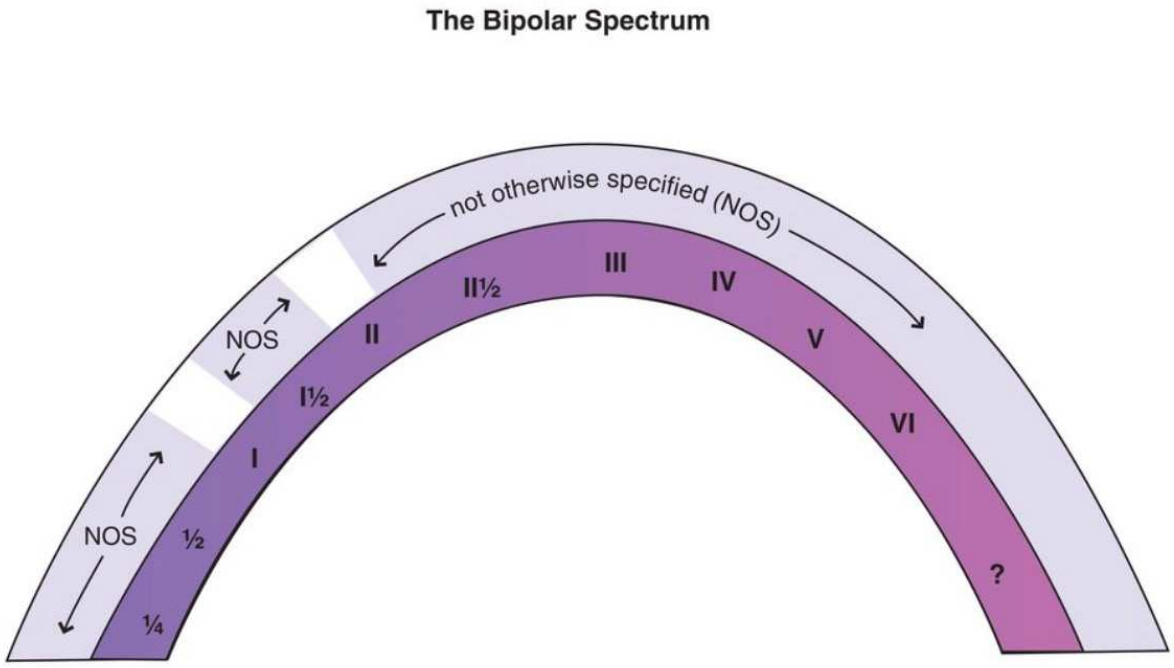

Figure II.2 : Le spectre bipolaire. La présentation des patients atteints de troubles bipolaires varie énormément. Historiquement, le trouble bipolaire a été catégorisé comme I, II ou non spécifié (NOS). Il peut être plus utile de considérer ces patients comme appartenant à un spectre bipolaire et d'identifier des sous-catégories de présentation, comme l'ont fait Akiskal et d'autres experts, et comme l'illustrent les figures suivantes. *Figure issue de (Stahl, 2013).*

Le **sex-ratio serait proche de 1**, à la différence du trouble unipolaire de l'humeur (ou trouble dépressif récurrent), plus fréquent chez les femmes (Diflorio & Jones, 2010), avec toutefois une disparité selon le sous-type, le BD-I touche autant les hommes que les femmes, tandis que le BD-II est plus fréquent chez les femmes (Ferrari et al., 2011). Il

entraine fréquemment des **troubles fonctionnels et cognitifs** de manière chronique, associés à une **réduction de la qualité de vie** (Grande et al., 2016). Comme le trouble bipolaire est principalement diagnostiqué chez les jeunes adultes, il affecte la population économiquement active et, par conséquent, entraine des coûts élevés pour la société (Grande et al., 2016).

b. Aspects cliniques et nosographiques

Aucun **biomarqueur diagnostique** n'est suffisamment performant pour être utilisé en pratique clinique courante pour diagnostiquer un quelconque trouble psychique, notamment pour le trouble bipolaire. Les **symptômes cliniques** (souvent regroupés en syndrome), associés à **l'élimination de divers diagnostics différentiels** psychiatriques et non psychiatriques (cliniquement et para-cliniquement), ainsi que la prise en compte du **retentissement fonctionnel** des symptômes, permettent de faire un diagnostic. La classification diagnostique la plus connue et utilisée à ce jour est la 5ème édition du Manuel Diagnostique et Statistique des troubles mentaux (**DSM-5**) (APA, s. d.).

Selon cette classification, les épisodes maniaques ou hypomaniaques sont caractérisés par une humeur élevée, expansive ou irritable de façon persistante, associée à une énergie anormalement et durablement élevée ainsi qu'une majoration des **comportements orientés vers un but**, d'une durée d'au moins 1 semaine pour la manie et de 4 jours consécutifs pour l'hypomanie. Ces symptômes doivent également être accompagnés d'un certain nombre d'autres symptômes (au moins 3, ou 4 si seulement irritabilité), tels qu'une augmentation de l'estime de soi, une diminution du besoin de sommeil, une augmentation du débit de parole, une sensation de fuite des idées (ou impression subjective que les pensées s'emballent), une distractibilité (l'attention est trop facilement attirée par des stimuli externes sans importance ou non pertinents), **une augmentation de l'activité dirigée vers un but** (soit socialement, au travail ou à l'école, soit sexuellement) ou une **agitation psychomotrice** (non dirigée vers un but), la **participation excessive à des activités agréables** qui ont un fort potentiel de conséquences délétères (par exemple, achats effrénés, conduites sexuelles à risque ou

investissements commerciaux insensés) (APA, s. d.). Un épisodes maniaque entrave très clairement le fonctionnement relationnel, social ou professionnel et peut être associé à des **symptômes psychotiques**, et conduire le plus souvent à une **hospitalisation** du fait de l'intensité et donc du retentissement des symptômes. Un épisode hypomaniaque est d'**intensité moindre**, et bien que la perturbation du fonctionnement puisse être perçue comme inhabituelle par l'entourage, celui-ci n'entraine généralement pas d'atteinte sévère du comportement et ne nécessite pas d'hospitalisation. Dans les deux cas, le patient présente un **mauvais insight** et ne se rend pas ou peu compte du caractère pathologique de son état, du moins au début de l'évolution du trouble et avant les interventions de psychoéducation (Grande et al., 2016).

Concernant l'épisode dépressif, l'individu doit présenter 5 symptômes ou plus au cours d'une **même période de deux semaines** et au moins un des symptômes doit être soit (1) une humeur triste, soit (2) une **perte d'intérêt ou de plaisir**. Parmi ces symptômes, on retrouve l'humeur dépressive, une **diminution de l'intérêt ou du plaisir pour tout**, une perte (le plus souvent) ou un gain de poids significatif, une diminution (le plus souvent) ou une augmentation de l'appétit, un ralentissement de la pensée et une **réduction des mouvements physiques** (observables par les autres, et pas seulement des sentiments subjectifs d'agitation ou de ralentissement), une **fatigue ou perte d'énergie**, un sentiment de dévalorisation ou de culpabilité excessive ou inappropriée, une réduction de la capacité de réflexion ou de concentration, ou **indécision**, des pensées récurrentes de mort, idées suicidaires ou tentative de suicide. Ici encore **selon l'intensité et le retentissement des symptômes**, l'épisode peut être qualifié de léger, modéré, ou sévère (APA, s. d.).

2. Etiopathogénie et physiopathologie du trouble bipolaire

a. L'étiopathogénie

Le mécanisme **étiopathogénique** des troubles bipolaires reste à ce jour incomplètement élucidé, ce mécanisme se montrant très probablement multifactoriel, la pathologie résultant d'une **interaction entre des facteurs génétiques et environnementaux** (Hamdani,

2012). Plusieurs gènes impliqués dans la genèse de la pathologie ont pu être identifiés comme facteurs de risque, avec une héritabilité et la transmission d'une vulnérabilité à présenter la pathologie et ainsi la « déclarer » en présence de facteurs environnementaux (Grunze, 2015). Les gènes impliqués codent pour la production de certaines protéines, de certains neurotransmetteurs ou encore pour des récepteurs notamment dopaminergiques. Toutefois, des génotypes similaires peuvent se manifester par des **phénotypes variés** indiquant une pénétrance incomplète des facteurs génétiques, d'où **l'hétérogénéité de la clinique** et la variabilité épidémiologique selon les études. Cela explique en partie que seuls certains membres de la famille développent la maladie, que d'autres témoignent de formes atténuées et que les derniers ne présentent aucun symptôme (Drevets et al., 1998; Grunze, 2015; Lichtenstein et al., 2009).

b. La physiopathologie

Concernant la **physiopathologie dans le trouble bipolaire**, plusieurs modèles coexistent pour tenter de l'expliquer en intégrant cette notion développementale de l'interaction gène et environnement, tout en rajoutant une sensibilité au stress environnemental favorisant les épisodes thymiques, pour **un modèle d'interaction gène-environnement-stress** (**Figure II.3**) (Maletic & Raison, 2014).

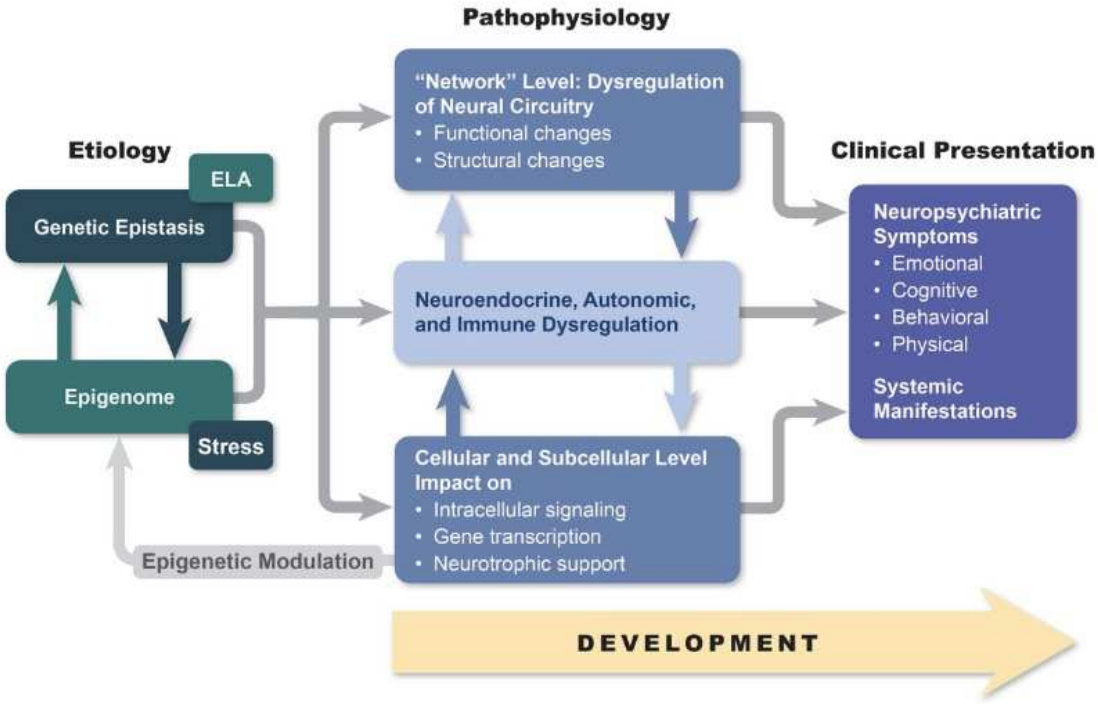

Figure II.3 : Une compréhension de la physiopathologie des troubles bipolaires basée sur l'étiopathogénie. Les modèles descriptifs des troubles de l'humeur n'offrent qu'une orientation minimale en matière de traitement. Un modèle reliant le génotype, les modifications épigénétiques et les altérations endo-phénotypiques à plusieurs niveaux à la présentation clinique peut ouvrir la voie à un meilleur succès thérapeutique. Ce modèle reconnaît la diversité physiopathologique des troubles bipolaires et offre la possibilité d'approches thérapeutiques individualisées basées sur le lien entre les constellations de symptômes, la génétique et les marqueurs endo-phénotypiques spécifiques. *Figure issue de (Maletic & Raison, 2014)*.

Ils ne seront pas tous abordés dans ce travail de thèse, mais il est intéressant d'évoquer **l'hypothèse dopaminergique des troubles bipolaires**, soutenant l'hypothèse d'un état d'hyperdopaminergie dans la manie, avec en particulier des élévations de la disponibilité des **récepteurs D2/D3 et un réseau hyperactif de traitement de la récompense** (Ashok et al., 2017; Pearlson, 1995; Whitton et al., 2015). Dans la dépression bipolaire, il existerait une **augmentation des transporteurs de la dopamine**, mais d'autres études sont nécessaires afin de confirmer cela. Les données pharmacologiques montrent que les agonistes de la dopamine et les anti-dopaminergiques peuvent améliorer les symptômes de la dépression bipolaire. Ces données suggèrent un modèle dans lequel une **augmentation de la disponibilité des récepteurs D2/3 au niveau du striatum** entraînerait une augmentation de la neurotransmission dopaminergique et de la manie,

tandis qu'une augmentation des niveaux du transporteur de la dopamine (DAT) au niveau du striatum entraînerait une réduction de la fonction dopaminergique et de la dépression (Ashok et al., 2017) (**Figure II.4**). Il est donc supposé qu'une **défaillance de l'homéostasie des récepteurs et des transporteurs de la dopamine** pourrait être à l'origine de la physiopathologie de ce trouble. Les limites de ce modèle sont sa dépendance à l'égard des preuves pharmacologiques, car ces études pourraient potentiellement affecter d'autres monoamines, ainsi que la rareté des preuves d'imagerie sur la fonction dopaminergique.

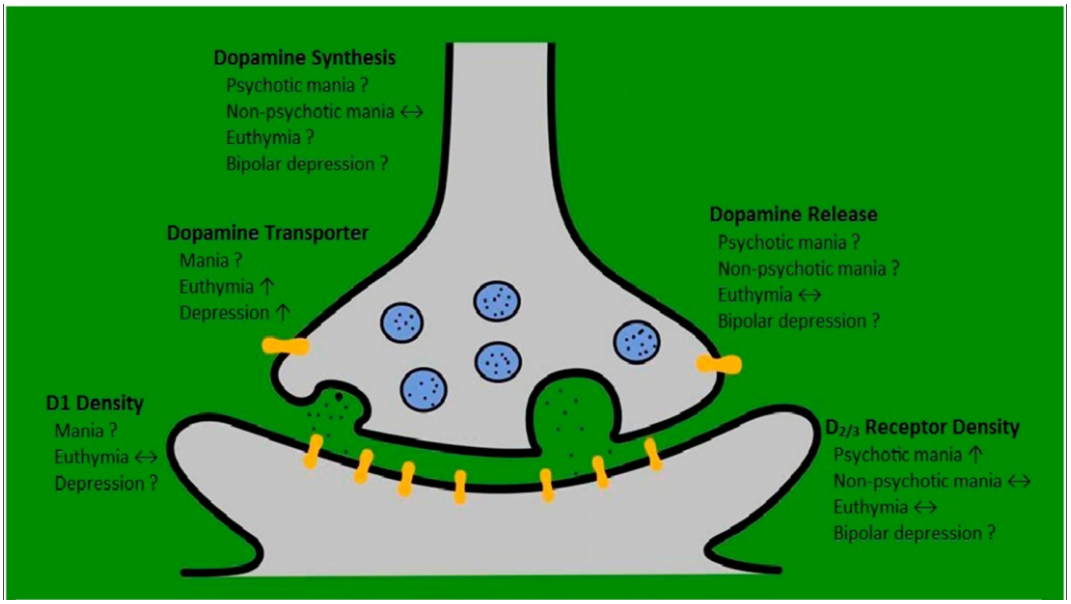

Figure II.4 : Résumé des résultats de l'imagerie moléculaire de la dopamine dans le trouble bipolaire. *Figure issue de (Ashok et al., 2017).*

De manière plus générale, de nombreux modèles s'accordent à donner une place importante aux **fonctions cognitives** tant leur expression est perturbée durant les différents états thymiques (S. de Sá et al., 2016), notamment en ce qui concerne le traitement des émotions et le **traitement de la récompense** (Miskowiak et al., 2019). Il reste toutefois difficile à dire si ces altérations participent à l'émergence de la pathologie et sont stables dans le temps (**marqueur-trait**), ou si elles sont une conséquence d'un état thymique donné, influencé par l'évolution de la pathologie et notamment en lien avec les différentes rechutes thymiques (**marqueur-état)**. Il apparait donc pertinent d'étudier les mécanismes

cognitifs impliqués dans le trouble bipolaire, notamment durant la **période euthymique** afin de mieux comprendre cette frontière entre marqueur trait et état, et ainsi mieux comprendre leur implication dans la physiopathologie du trouble.

3. L'euthymie : une période symptomatique

a. Les particularités cliniques de l'euthymie

Pendant longtemps, il a été considéré que le trouble bipolaire ne s'exprimait que par des épisodes dépressifs et maniaques ou hypomaniaques, avec un « retour à la normale » entre les épisodes, appelé **période euthymique** ou **euthymie**. Bien évidemment, les individus peuvent présenter des **fluctuations physiologiques de l'humeur** durant cette période, comme dans la population générale, ces fluctuations **n'apparaissant pas comme pathologiques**. En revanche, nous savons actuellement que les patients atteints de troubles bipolaires présentent certaines caractéristiques cliniques qui persistent durant l'euthymie par rapport à la population générale. En effet, ils peuvent présenter une **impulsivité** accrue (Etain et al., 2013), une plus grande **labilité émotionnelle** et une plus grande **réactivité émotionnelle** (Henry, 2012), ainsi que des perturbations des **fonctions cognitives** pour environ un tiers d'entre eux (Cullen et al., 2016), concernant l'attention soutenue, la mémoire verbale, la mémoire épisodique, et certaines fonctions exécutives. Ces symptômes persistants en dehors des épisodes thymiques aigus pourraient constituer des **biomarqueurs cognitifs** dans les troubles bipolaires. Un **biomarqueur** peut être défini comme une donnée objectivement mesurable et aisément accessible apportant une indication prédictive, diagnostique et/ou pronostique sur un processus pathologique et sa réponse à une stratégie thérapeutique (**Figure II.5**) (García-Gutiérrez et al., 2020).

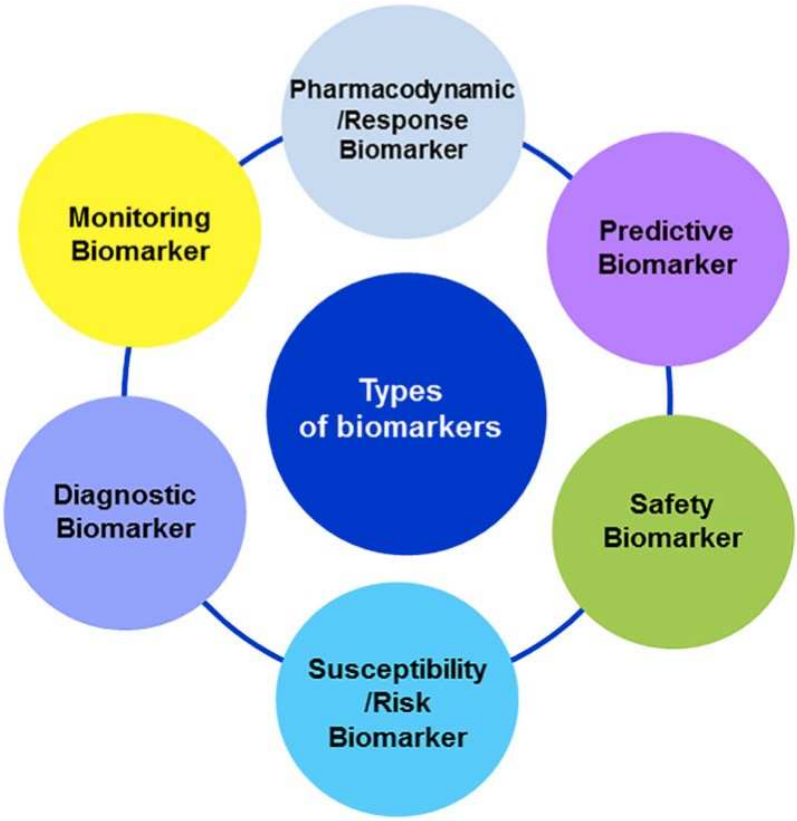

Figure II.5 : Classification des biomarqueurs en fonction de leur principale application clinique. *Figure issue de (García-Gutiérrez et al., 2020).*

b. L'objectif de la recherche de biomarqueurs-état du trouble bipolaire

La recherche sur le trouble bipolaire se concentre actuellement sur les stratégies de **prévention** primaire, secondaire et tertiaire, en plus de la **stadification** de la maladie, cette approche exigeant **d'affiner les méthodes de diagnostic** actuelles, notamment en développant des biomarqueurs. L'évolution vers une **médecine personnalisée** étant de plus en plus marquée, ceci incite à une amélioration des outils d'évaluation des risques et de sélection des traitements (Pasco et al., 2010; Teixeira et al., 2016). Les biomarqueurs peuvent ainsi être des outils utiles pour détecter l'activité de la maladie associée à un état thymique particulier (**marqueur-état**), ou pour identifier des caractéristiques spécifiques observées dans l'évolution à long terme de la maladie (**marqueur-trait**) (Frey et al., 2013). Ainsi, mieux comprendre la période de **rémission** permettrait donc de mieux diagnostiquer les troubles bipolaires, et de manière plus précoce en identifiant les **sujets à risque**, ainsi que de **personnaliser** les stratégies de prise en charge pour prévenir les rechutes, ce qui

reste un défi majeur pour les cliniciens. L'étude des **altérations cognitives telles que la motivation et le traitement de la récompense**, en cherchant à définir des phénotypes cognitifs et computationnels dans les troubles bipolaires en rémission, pourraient ainsi permettre de mieux prévenir ces rechutes (Pessiglione, Le Bouc, et al., 2018; Pessiglione, Vinckier, et al., 2018).

4. Pourquoi étudier l'apprentissage par renforcement dans le trouble bipolaire ?

a. L'apprentissage par renforcement et le programme *Research Domain of Criteria*

Comme évoqué précédemment, il a bien été montré que les patients atteints de trouble bipolaire présentaient des **troubles cognitifs durant les périodes euthymiques**, notamment dans le domaine des « **cognitions chaudes** » (faisant intervenir un traitement émotionnel). Selon l'approche matricielle du programme ***Research Domain of Criteria*** (**RDoC**), les différentes formes cliniques du trouble bipolaire peuvent être conceptualisées comme un « spectre » de troubles dans lequel la **sensibilité à la récompense** serait une dimension clé de la pathologie, bien que cela reste encore peu examiné (Insel et al., 2010). Cette approche a été développée en 2010, Thomas Insel et ses collègues expliquant que les catégories diagnostiques issues du DSM et de la CIM (Classification Internationale des Maladies), fondées sur des consensus cliniques reposant sur l'observation de signes et symptômes, ne correspondaient plus aux résultats des neurosciences cliniques et de la génétique. Ils prônent donc une **démarche translationnelle** : de la compréhension des mécanismes de fonctionnement du « mental » vers les expressions visibles des dysfonctionnements (« troubles ») de ces mécanismes, plutôt que l'inverse. Ils proposent ainsi non plus une approche *top-down* de troubles fondée sur des symptômes pour en chercher la psycho-physio-pathologie, mais **une approche *bottom-up***, les symptômes comme des disruptions dans le fonctionnement correct de mécanismes implémentant différentes fonctions. Ils proposent ainsi **une matrice** qui croise **7 unités d'analyse** (gènes, molécules, cellules, circuits, physiologie, comportements, descriptions subjectives) avec **6 domaines fonctionnels** (systèmes des valences négatives, systèmes des valences positives,

systèmes cognitifs, systèmes des processus sociaux, systèmes d'éveil et de modulation, systèmes sensorimoteurs) (**Figure II.6**).

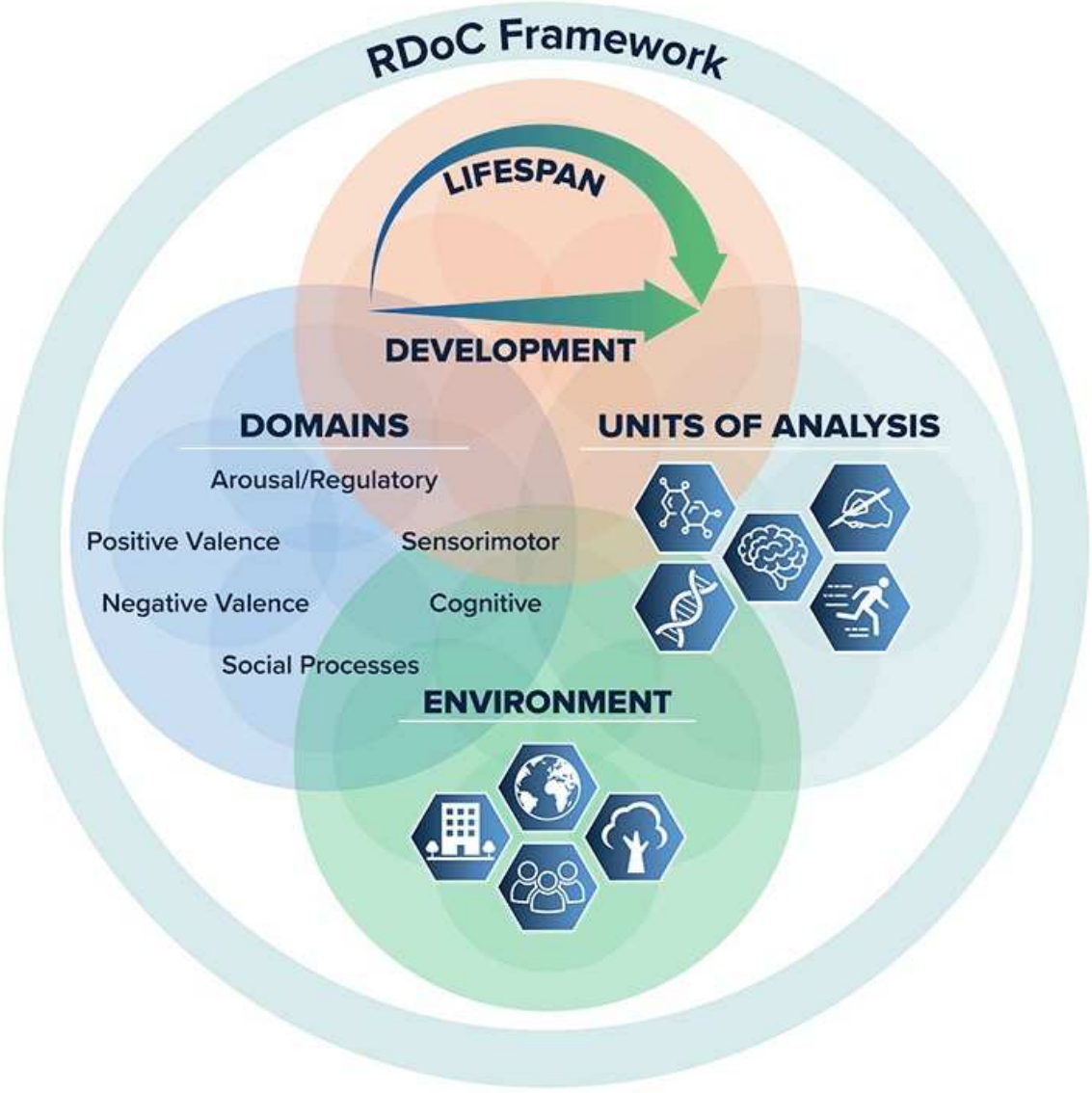

Figure II.6 : Le cadre de travail du programme *Research Domain of Criteria* (RDoC). Le cercle orange vise à souligner le rôle important du développement dans la santé mentale tout au long de la vie. Le cercle gris contient un exemple des unités d'analyse pour mesurer et comprendre les processus fonctionnels des troubles psychiques. Le cercle vert représente l'impact des facteurs environnementaux et sociaux sur les troubles psychiques. Le cercle bleu contient les 6 domaines fonctionnels qui peuvent être examinés dans les études. *Figure issue du site internet du programme Research Domain of Criteria (RDoC) (https://www.nimh.nih.gov/research/research-funded-by-nimh/rdoc/about-rdoc).*

Ces domaines fonctionnels sont eux-mêmes décomposés en un petit nombre de **construits théoriques**. Les construits sont des entités abstraites non directement observables à un instant donné mais auxquelles on suppose une **validité naturelle empiriquement testable**. Chaque case de la matrice (croisement construit X unité d'analyse) serait ensuite supposée faire l'objet de recherches empiriques documentant les systèmes biologiques impliqués et leur spectre de fonctionnement. Selon cette approche matricielle, **le traitement de la récompense** se situe donc au sein du domaine des **systèmes de valences positives**, auquel se subdivise les construits théoriques de la **réactivité (sensibilité) aux récompenses**, de **l'apprentissage par la récompense** (dont fait partie **l'apprentissage par renforcement**), et de l'évaluation de la récompense (Insel et al., 2010).

b. Les approches *top-down* et *bottom-up* du trouble bipolaire

L'étude du traitement de la récompense dans le trouble bipolaire peut donc certes se conceptualiser selon une approche ***top-down***, du fait de l'existence durant les différents états thymiques d'une modification des comportements dirigés vers un but ainsi que d'une recherche d'activités à caractère hédonique, mais également selon une approche ***bottom-up***, notamment du fait de l'implication de la dopamine dans la physiopathologie des troubles bipolaires. En effet, de nombreux patients peuvent présenter des symptômes psychotiques durant des épisodes maniaques et/ou dépressifs, les traitements antipsychotiques (anti-dopaminergiques D2) se montrant efficaces sur ces symptômes. De plus, il est fait état d'une **hypothèse physiopathologique dopaminergique** dans le trouble bipolaire comme évoqué précédemment, suite aux résultats d'études pharmacologiques, sur le modelé animal ou l'Homme, d'études post-mortem, en imagerie in-vivo, ou par neuromodulation (Ashok et al., 2017). Aussi, il a été retrouvé des altérations des performances lors de tâches de **prise de décision** chez des patients atteints de trouble bipolaire en phase euthymique sous **agonistes dopaminergiques** (Burdick et al., 2014).

5. Résumé de la partie II

Nous avons donc vu dans cette deuxième partie l'intérêt de faire de la recherche en neurosciences cliniques sur les troubles bipolaires, au regard de la fréquence et de l'impact de cette pathologie sur la population générale et les patients. Il apparait à ce jour pertinent d'essayer de **mieux comprendre la physiopathologie** des différents états thymiques du trouble bipolaire (incluant **l'euthymie**), dans le but à terme de trouver des **biomarqueurs**-état et surtout trait de la pathologie, dans le but de mieux diagnostiquer, mieux prévenir, et mieux traiter les épisodes.

Au-delà d'une simple alternance d'épisodes comprenant une altération de l'humeur, il existe durant les épisodes dépressifs et maniaques ou hypomaniaques une **altération du traitement de la récompense (et de la punition)** se traduisant par des **altérations motivationnelles**, ces caractéristiques étant moins claires durant la période euthymique. Il semble en effet persister des altérations cognitives durant l'euthymie, notamment concernant la prise de décision et le traitement de la récompense et motivation. Il est alors aisé de comprendre l'intérêt d'étudier le traitement de la récompense, tant l'aspect clinique (*top-down*) et la recherche en neurosciences sur l'implication de la récompense dans la physiopathologie (*bottom-up*). Nous allons voir dans cette troisième partie quels sont les déficits d'apprentissage par renforcement dans les troubles bipolaires connus à ce jour en commençant par voir comment les choix impactent notre humeur et comment notre humeur impacte nos choix.

III

Récompense et apprentissage par renforcement dans le trouble bipolaire

Maintenant que nous avons vu les grands principes de l'apprentissage par renforcement, et pourquoi il apparaissait intéressant de l'étudier dans le trouble bipolaire, nous allons voir dans cette troisième partie un état des lieux de la littérature scientifique sur l'apprentissage par renforcement et le traitement de la récompense spécifiquement dans le trouble bipolaire. Cela permettra de mieux comprendre les différentes questions de recherche que nous nous sommes posées et qui seront abordées dans la quatrième sous-partie.

1. Relation bidirectionnelle entre les choix et l'humeur

a. Impact des choix sur l'humeur

L'équipe de Robb Rutledge a développé un **modèle computationnel examinant en laboratoire les fluctuations de l'humeur** durant une tâche d'apprentissage probabiliste. Dans une de leurs études, les sujets devaient faire des choix de façon répétée lors de jeux de hasard dont les gains et les pertes potentiels variaient d'un essai à l'autre. La principale

conclusion de cette étude est que **les fluctuations de l'humeur dépendent principalement de la différence entre les résultats attendus et les résultats réels (l'erreur de prédiction)** (Rutledge et al., 2014). L'originalité de cette étude réside dans le fait que, grâce à la modélisation computationnelle, les auteurs ont pu **estimer la valeur théorique de l'humeur sur une base essai par essai**, pour chaque sujet, même en l'absence d'évaluation réelle de l'humeur. En particulier, le « bien-être subjectif » a été modélisé comme suit :

$$Happiness(t) = w_0 + w_1 \sum_{j=1}^{t} \gamma^{t-j} CR_j + w_2 \sum_{j=1}^{t} \gamma^{t-j} EV_j + w_3 \sum_{j=1}^{t} \gamma^{t-j} RPE_j$$

Où, pour chaque essai j (du premier essai à l'essai actuel t), si la récompense certaine a été choisie, elle est retrouvée dans l'équation sous la forme ***CRj***. Inversement, si le jeu a été choisi, deux termes sont entrés dans l'équation : ***EVj***, la valeur attendue du jeu, et ***RPEj***, la différence entre le résultat réel et l'EV. Les poids ω (dont un terme constant $\omega 0$) captent **l'influence des différents types d'évènements**. Enfin, ces influences décroissent exponentiellement dans le temps avec un **facteur d'oubli** $0 \leq \gamma \leq 1$ qui rend les évènements récents plus influents que les plus anciens.

b. Impact de l'humeur sur les choix

Ce travail marque un tournant dans l'étude des mécanismes par lesquels l'humeur influence le comportement en fournissant **une approche computationnelle de la façon dont les fluctuations de l'humeur peuvent découler du feedback qu'un individu reçoit.** Dans une autre étude, **l'humeur a été manipulée** à l'aide d'un jeu de roue de la fortune dans lequel les sujets gagnaient ou perdaient une somme d'argent relativement importante (Eldar & Niv, 2015). Il a été montré que le fait de gagner le tirage au sort augmentait non seulement l'humeur subjective mais aussi la valeur de récompense subjective perçue dans les choix ultérieurs. A l'inverse, perdre le tirage au sort a réduit l'humeur, ainsi que l'effet des récompenses sur les choix ultérieurs. **Ces résultats suggèrent que l'humeur influence la prise de décision en biaisant la perception des résultats, une humeur positive (négative) entrainant une évaluation plus élevée (plus faible).**

Une autre étude a **confirmé ces résultats** chez des sujets sains et avec un modèle computationnel similaire, en montrant que l'intégration dans le temps des feedbacks positifs et négatifs d'une tâche de quiz induisait des fluctuations de l'humeur qui, à leur tour, modulaient les poids relatifs attribués aux gains et aux pertes dans une tâche de choix (Vinckier et al., 2018). Plus précisément, **une bonne humeur favorisait la prise de risque en surpondérant les gains potentiels, tandis qu'une mauvaise humeur tempérait la prise de risque en surpondérant les pertes potentielles**.

Dans l'ensemble, ces résultats suggèrent qu'il existe une **influence bidirectionnelle entre l'humeur et le traitement des résultats pendant la prise de décision**. Alors que les fluctuations de l'humeur ont été modélisées avec précision par un processus d'intégration de l'historique du feedback et des attentes des sujets (Rutledge et al., 2014), des modèles ultérieurs ont inclus une **influence réciproque de l'humeur sur la perception du feedback** pour capturer le fait que l'humeur déforme également la façon dont les sujets perçoivent les résultats dans les contextes de décision (Eldar & Niv, 2015; Vinckier et al., 2018), en considérant les évènements comme plus positifs qu'ils ne le sont objectivement lorsqu'ils sont de bonne humeur et inversement lorsqu'ils sont de mauvaise humeur. De manière cruciale, **des dysfonctionnements spécifiques de cette boucle de rétroaction ont été avancés pour contribuer à l'émergence de troubles de l'humeur tels que la dépression ou l'instabilité de l'humeur** (**Figure III.1**) (Eldar et al., 2016).

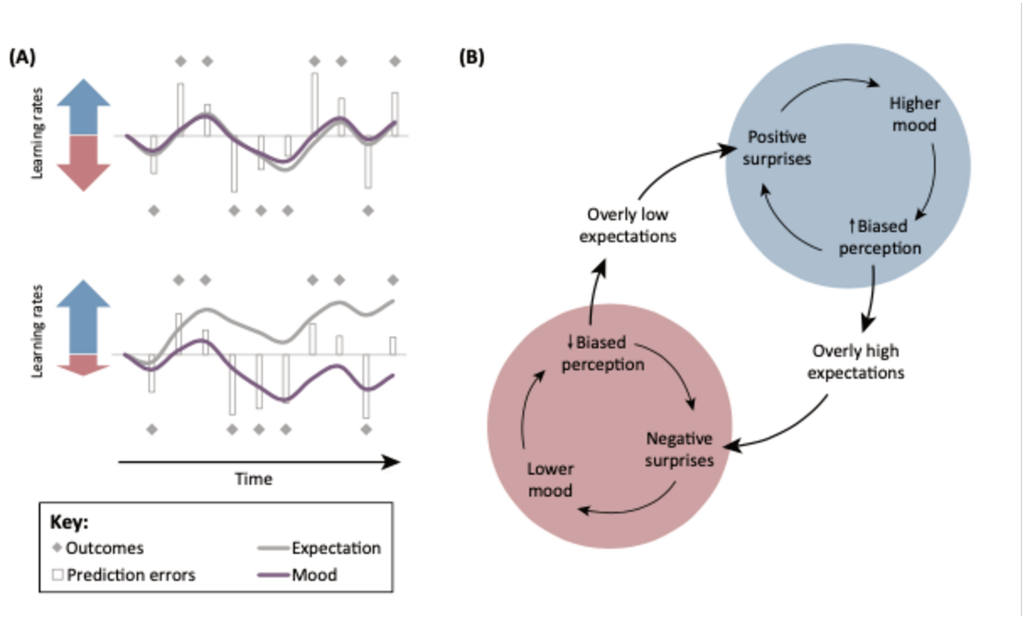

Figure III.1 : Schéma des dysfonctionnements possibles de l'humeur. (A) (En haut) Étant donné un taux d'apprentissage similaire en réponse à des résultats positifs et négatifs, un environnement dans lequel les résultats positifs et négatifs ont la même probabilité conduit à des attentes neutres et à une humeur neutre en moyenne. (En bas) Un taux d'apprentissage plus faible en réponse à des résultats négatifs conduit à des erreurs de prédiction négatives plus importantes et à une humeur négative persistante, un symptôme observé dans le **trouble dépressif majeur**. **(B)** La dynamique de rétroaction positive « escalatoire » pourrait transformer l'humeur en une « **prophétie autoréalisatrice** », conduisant à l'instabilité émotionnelle, un symptôme majeur du **trouble bipolaire**. Les surprises positives améliorent l'humeur, en biaisant la perception des résultats à la hausse, augmentant ainsi la fréquence et l'ampleur d'autres surprises positives. Les attentes optimistes se développent en raison de la perception biaisée des résultats. L'humeur se stabilise lorsque les attentes rattrapent les résultats perçus, mais les résultats ultérieurs, dont la perception n'est plus biaisée, ont tendance à ne pas répondre aux attentes optimistes. Ainsi, les surprises négatives s'enchainent, diminuant l'humeur et biaisant la perception des résultats à la baisse. Une dynamique de rétroaction positive similaire engendre alors des attentes pessimistes, ce qui permet au cycle de se répéter, oscillant indéfiniment entre bonne et mauvaise humeur, même si la distribution réelle des résultats ne change pas. *Figure issue de (Eldar et al., 2016).*

c. Lien entre choix, humeur, et erreur de prédiction

Plus récemment, il a été montré lors d'une étude que ce n'était finalement probablement pas le feedback en lui-même qui modulait l'humeur, mais **l'apprentissage en lien avec ce feedback** (Blain & Rutledge, 2020). En effet, il a été mis en évidence que **l'humeur est davantage influencée par des variables dépendantes de l'apprentissage** (erreur de prédiction sur la probabilité de gagner (PPE) par exemple) que par des variables ne

concernant pas l'apprentissage (erreur de prédiction sur les récompenses - RPE). **L'humeur serait donc plus fortement liée à la surprise d'une récompense (erreur de prédiction) plutôt qu'à sa valeur.** Ainsi, en confrontant différents modèles computationnels, celui qui rendait le mieux compte de l'humeur des participants était finalement le suivant :

$$Happiness(t) = w_0 + w_P \sum_{j=1}^{t} \gamma^{t-j}(P - O, 5) + w_{PPE} \sum_{j=1}^{t} \gamma^{t-j} PPE_j$$

Dans cette équation, w_0 est une constante, w_{PPE} et w_P sont respectivement l'influence des précédentes **erreurs de prédiction** (PPE) et des probabilités de choisir un des symboles (P estimées par le modèle additif), et $0 \leq \gamma \leq 1$ est le **facteur d'oubli** qui assure que les essais les plus récents aient davantage d'influence que les plus anciens. Le facteur d'oubli est un paramètre intrinsèque à chaque individu, néanmoins, ils ont trouvé qu'en moyenne ce sont les 6 derniers essais qui sont pris en compte pour noter l'humeur instantanée.

2. Apprentissage par renforcement dans le trouble bipolaire

Étant donné la **relation bidirectionnelle entre les fluctuations de l'humeur et l'apprentissage par renforcement** (Blain & Rutledge, 2020; Eldar et al., 2016; Rutledge et al., 2014), plusieurs études se sont concentrées sur l'identification des altérations d'apprentissage par renforcement dans le trouble bipolaire. Les études ont cependant montré des **résultats relativement disparates concernant les performances d'apprentissage par renforcement**.

a. Performances globales en apprentissage par renforcement

Si les performances en apprentissage par récompense ou par punition ne sont pas séparées, il est retrouvé de **moins bonnes performances** en apprentissage par renforcement dans les troubles bipolaires en comparaison des sujets sains dans **19 études** (Abohamza et al.,

2020; Adida et al., 2008, 2011, 2015; Akbari et al., 2019; Barch et al., 2017; Clark et al., 2001; Duek et al., 2014; Geana et al., 2022; Gomide Vasconcelos et al., 2014; Gu et al., 2020; Linke et al., 2012; Malloy-Diniz et al., 2011; Pizzagalli et al., 2008; Pratt et al., 2021; Roiser et al., 2009; Strauss et al., 2015; Van Enkhuizen et al., 2014; Whitton et al., 2021), mais **pas de différence** entre les deux groupes pour **12 études** (Brambilla et al., 2013; Caletti et al., 2013; Edge et al., 2013; Ibanez et al., 2012; Jogia et al., 2012; Lewandowski et al., 2016; Linke et al., 2011; Martino et al., 2011; Ono et al., 2015; Powers et al., 2013; Ryu et al., 2017; Yechiam et al., 2008).

b. Différence entre apprentissage par récompense et par punition

Si l'apprentissage par récompense est cette fois différencié de l'apprentissage par punition, il est retrouvé dans la **condition récompense 14 études** mettant en avant de **moins bonnes performances** en apprentissage par récompense dans les troubles bipolaires en comparaison des sujets sains (Abohamza et al., 2020; Adida et al., 2008, 2011, 2015; Akbari et al., 2019; Clark et al., 2001; Duek et al., 2014; Gomide Vasconcelos et al., 2014; Gu et al., 2020; Malloy-Diniz et al., 2011; Pizzagalli et al., 2008; Pratt et al., 2021; Van Enkhuizen et al., 2014; Whitton et al., 2021), contre **11 études ne montrant pas de différence** (Brambilla et al., 2013; Caletti et al., 2013; Edge et al., 2013; Ibanez et al., 2012; Jogia et al., 2012; Lewandowski et al., 2016; Martino et al., 2011; Ono et al., 2015; Powers et al., 2013; Ryu et al., 2017; Yechiam et al., 2008). Seulement 3 **études spécifient l'apprentissage par punition et retrouvent de moins bonnes performances** en apprentissage par évitement de la punition dans les troubles bipolaires en comparaison des sujets sains (Abohamza et al., 2020; Duek et al., 2014; Pratt et al., 2021). Les résultats de ces études sont décrits dans le **Tableau III.1**.

Tableau III.1 : Résumé des études expérimentales sur l'apprentissage par renforcement dans les troubles bipolaires.

	Trouble bipolaire			Sujets sains						
Source	No.	Homme, %	Age, moyenne (SD), année	No.	Homme, %	Age, moyenne (SD), année	État thymique	Tache	Apprentissage (type)	Résultat principal

Adida et al., 2008	45	51	37.8 (12.7)	45	51	37.3 (11.5)	Manie	IGT	Reward learning	BD < HC
Adida et al., 2011	167 (45 M, 32 D, 90 E)	41.3	40.3 (11.6)	150	50	38.8 (10.6)	Manie, euthymie, dépression	IGT	Reward learning	Les 3 groupes de BD < HC Pas de différence entre les 3 groupes
Adida et al., 2015	90 (34 avec lithium, 56 sans lithium)	36.65	38.9 (12.06)	152	50.7	39 (10.8)	Euthymie	IGT	Reward learning	Euthymiques avec lithium et HC > euthymiques sans Lithium Pas de différence entre euthymiques avec lithium et HC
Akbari et al., 2019	35 (BD-II)	51	28.8 (2.44)	30	60	25.98 (2.76)	Euthymie	IGT	Reward learning	BD < HC
Brambilla et al., 2013	70	53	44.6 (11.3)	140	51	43.9 (11.2)	Euthymie	IGT	Reward learning	BD = HC
Caletti et al., 2013	18	22	42.22 (11.72)	18	33	36.11 (14.51)	Euthymie	IGT	Reward learning	BD = HC
Clark et al., 2001	15	67	35.4 (13)	30	53	37.6 (11.3)	Manie	IGT	Reward learning	BD < HC
Edge et al., 2013	55 (BD-I)	35	36 (11.9)	39	41	33.5 (12.8)	Euthymie	IGT	Reward learning	BD = HC
Gomide Vasconcelos et al., 2014	50	54	33.09 (13.48)	256	54	33.09 (13.48)	Euthymie	IGT	Reward learning	BD < HC

Gu et al., 2020	29	62	35.72 (9.65)	34	65	33.79 (9.09)	NA	IGT	Reward learning	BD < HC
Ibanez et al., 2012	13	62	40.1 (9.4)	25	64	35.1 (11.2)	Euthymie	IGT	Reward learning	BD = HC
Jogia et al., 2012	36	47	42.5 (10.6)	37	57	37.6 (11.3)	Euthymie	IGT	Reward learning	BD = HC
Malloy-Diniz et al., 2011	95	34	41 (12)	94	47	32 (13)	Euthymie	IGT	Reward learning	BD < HC
Martino et al., 2011	85 (48 BD-I, 37 BD-II)	30.45	40.25 (15.7)	34	35.3	40.0 (12.9)	Euthymie	IGT	Reward learning	BD-I = BD-II = HC
Ono et al., 2015	13	46	38.4 (7.3)	15	53	32.9 (7.7)	Euthymie	IGT	Reward learning	BD = HC
Powers et al., 2013	98	47	40.4 (12.09)	95	56	38.29 (11.49)	NA	IGT	Reward learning	BD < HC
Van enkhuizen et al., 2014	16	56	33.8 (2.8)	17	29	33.9 (3.0)	Manie	IGT	Reward learning	BD < HC
Yechiam et al., 2008	28	36	44.05 (9.05)	25	32	39.2 (13.3)	Manie, euthymie, dépression	IGT	Reward learning	BD (tout état thymique) = HC
Lewandowski et al., 2016	42	45	29.6 (8.4)	29	41	31.0 (10.0)	NA	PRT	Probabilistic reward learning	BD = HC
Pizzagalli et al., 2008	13	62	38.77 (12.09)	25	56	38.36 (10.76)	Euthymie	PRT	Probabilistic reward learning	BD < HC
Ryu et al., 29017	44	43	34.65 (6.76)	24	46	31.9 (6.96)	Manie, euthymie	PRT	Probabilistic reward learning	BD (tout état thymique) = HC
Whitton et al., 2021	104	46.95	40.81 (13.37)	129	37.2	32.02 (12.13)	NA	PRT	Probabilistic reward learning	BD < HC
Duek et al., 2014	40	55	42 (11.73)	41	56	38.71 (11.22)	Euthymie	PCT	Probabilistic reward et punishment learning **séparés**	Ensemble : BD < HC

										Reward : BD < HC Punish : BD < HC
Pratt et al., 2019	43	NA	43.86 (4.56)	20	65	43.85 (6.8)	NA	PRPT	Probabilistic reward et punishment learning **séparés**	Ensemble : BD < HC Reward : BD < HC Punish : BD < HC
Roiser et al., 2009	49	23	33.6 (8.9)	55	35	34.9 (8.1)	Dépression	PRLT	Probabilistic reward et punishment learning avec du **reversal**	BD < HC que lors du reversal
Linke et al., 2012	19	42	45 (10)	22	50	28 (10)	Euthymie	PRLT	Probabilistic reward et punishment learning avec du **reversal**	BD < HC que lors du reversal
Barch et al., 2017	43	44	35.4 (10.0)	55	54	36.0 (11.1)	NA	EPILT	Probabilistic reward et punishment learning **séparés** (learning phase puis transfert phase)	Ensemble : BD < HC Pas de données séparés
Geana et al., 2021	60	NA (sup mat)	NA (sup mat)	72	NA (sup mat)	NA (sup mat)	NA	EPILT	Probabilistic reward et punishment learning **séparés** (learning phase puis transfert phase)	Ensemble : BD < HC Pas de données séparés
Pratt et al., 2021	62	35.5	38.3 (10.7)	75	54.7	37.4 (11.3)	NA	EPILT	Probabilistic reward et punishment learning **séparés** (learning phase puis transfert phase)	Ensemble : BD < HC Reward : BD < HC Punish : BD < HC (que pour contingence 90 :10)
Linke et al., 2011	23	48	44.1 (8.1)	19	47	43.1 (11.6)	Euthymie	PST	Probabilistic reward et punishment learning **séparés** (learning phase puis test phase)	Ensemble : BD = HC Différence si prise

										en compte du dernier épisode thymique
Strauss et al., 2015	47 (24 BD+ et 23 BD-)	32	35.75 (13.3)	24	45.8	36.1 (13.4)	Euthymie	PST	Probabilistic reward et punishment learning **séparés** (learning phase puis test phase)	BD+ et BD- < HC

Abréviations: IGT = Iowa Gambling Task ; PCT = Probabilistic Classification Task ; PRT = Probabilistic Reward Task ; PRPT = Probabilistic Reward Punishment Task ; PST = Probabilistic Selection Task ; EPILT = Explicit Probabilistic Incentive Learning Tasks ; PBLT = Probabilistic Learning Task. NA = Non disponible.

c. Pourquoi cette disparité dans les résultats ?

Ces divergences peuvent s'expliquer en partie par **l'hétérogénéité des tâches utilisées**. En effet, bien qu'il s'agisse de paradigmes concernant l'apprentissage lié à la récompense et/ou l'évitement de la punition, ils ne reflètent pas nécessairement le même processus cognitif sous-jacent (Balodis & Potenza, 2015; DePasque Swanson & Tricomi, 2014; Lutz & Widmer, 2014; Richards et al., 2013). Par exemple, l'Iowa Gambling Task (IGT), largement utilisée dans l'apprentissage de la récompense, impliquerait des processus cognitifs distincts de ceux des tâches d'apprentissage probabiliste. En outre, certains paradigmes n'incluent que l'apprentissage basé sur la récompense et non l'apprentissage basé sur l'évitement de la punition, ou mélangent les deux, de sorte que la spécificité des processus cognitifs modifiés dans le trouble bipolaire au cours de l'apprentissage n'est pas claire. Or, nous avons vu qu'il apparaissait intéressant **d'examiner distinctement l'apprentissage basé sur la récompense et l'apprentissage basé sur la punition**, du fait que les récompenses et les punitions sont également connues pour avoir un impact opposé sur l'humeur des sujets (Blain & Rutledge, 2020; Cecchi et al., 2022; Vinckier et al., 2018), de sorte qu'une asymétrie entre les processus d'apprentissage basés sur la récompense et la punition pourrait jouer un rôle dans la genèse de différents états thymiques et leur maintien (Eldar et al., 2016). Enfin, cela pourrait s'expliquer par d'autres facteurs tels que la prise en compte ou non de l'état thymique des patients. Par exemple, l'étude de Linke et

al. (2011) a mis en avant que les patients atteints de BD-I ayant connu un dernier épisode maniaque apprenaient mieux des feedbacks positifs que des feedbacks négatifs, alors que le schéma inverse était observé chez les patients atteints de BD-I ayant connu un dernier épisode dépressif (c'est-à-dire qu'ils apprenaient mieux des feedbacks négatifs) (Linke et al., 2011). Cet effet ne résulte ni d'un déficit général de la capacité d'apprentissage du patient, ni de symptômes d'humeur résiduels. Cette étude suggère donc que **la polarité du dernier épisode affectif biaise sélectivement la sensibilité des sujets aux résultats positifs/négatifs** pendant l'apprentissage par renforcement. Il semblerait donc que la capacité d'apprentissage en lien avec la récompense et la punition intervienne de façon importante dans les choix ultérieurs chez les patients atteints de trouble bipolaire. De même, la **sensibilité à la récompense** semble intervenir dans les comportements orientés vers un but et l'apprentissage qui en découle.

3. Sensibilité à la récompense et trouble bipolaire

a. Une hypersensibilité à la récompense dans le trouble bipolaire ?

L'hypothèse d'une **hypersensibilité à la récompense** dans le trouble bipolaire a longtemps était mise en avant comme **marqueur-trait**. Cette théorie provient notamment d'études sur le **système d'approche comportemental** (SAC ou BAS en anglais pour *Behavioral Approach System*), suggérant une hypersensibilité à la récompense liée aux symptômes hypomaniaques/maniaques et dépressifs chez les personnes atteintes de trouble bipolaire lorsqu'elles réagissent à des événements liés à la récompense (Alloy et al., 2012; Alloy & Abramson, 2010), et restant élevée en cas de rémission (Applegate et al., 2009; Mason et al., 2014; Meyer et al., 2001). L'hypersensibilité aux stimuli liés à la récompense pourrait être un élément clé de la dysrégulation émotionnelle, de la vulnérabilité et de la labilité affective dans le trouble bipolaire (Alloy et al., 2015; Urošević et al., 2008; Whitton et al., 2015). L'exploration des bases neurobiologiques des déficiences du traitement de la récompense chez les personnes atteintes de trouble bipolaire peut donc être fortement utile pour **améliorer le traitement et la prévention des rechutes**.

Certains auteurs proposent ainsi d'intégrer la **sensibilité à la récompense** dans un modèle physiopathologique du trouble bipolaire, dans lequel la sensibilité exacerbée ou diminuée associé à un évènement extérieur contribuerait à induire respectivement un épisode maniaque ou hypomaniaque et un épisode dépressif (Alloy & Nusslock, 2019; Nusslock & Alloy, 2017; Wessa et al., 2014) (**Figure III.2**). Sur la base de cette hypothèse d'hypersensibilité à la récompense, les sujets plus sensibles aux récompenses seraient susceptibles de connaître un épisode dépressif lorsque leur système de récompense est excessivement désactivé à la suite d'une exposition à des événements de vie impliquant un échec et une perte. De plus, **la sensibilité à la récompense pendant la période euthymique semble être influencée par l'effet du dernier épisode thymique**, avec une sensibilité émoussée après un épisode dépressif, et plutôt augmentée après un épisode (hypo)maniaque (Linke et al., 2011).

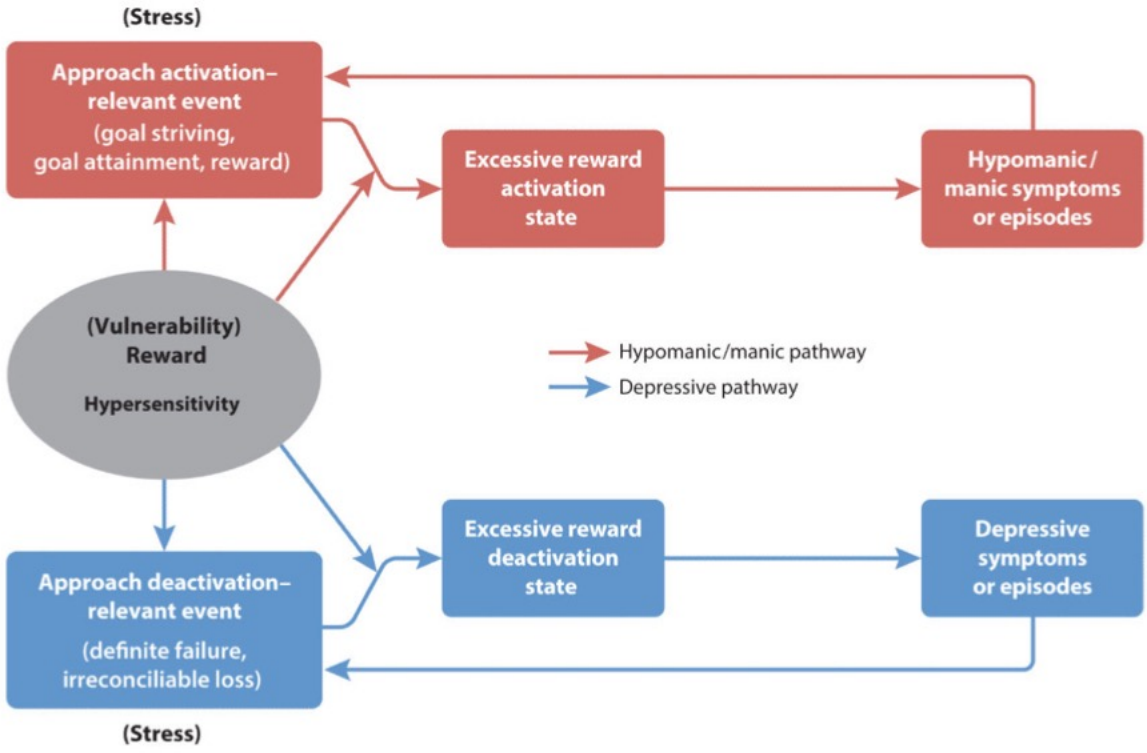

Figure III.2 : Modèle d'une hypersensibilté à la récompense dans le trouble bipolaire. *Figure issue de (Nusslock & Alloy, 2017).*

b. Ou une hyposensibilité à la récompense ?

Cependant, ce modèle a plus récemment été contredit par certaines études, notamment par une revue systématique de la littérature réalisée par l'ISBD (*International Society for Bipolar Disorders*) des études comportementales portant sur la sensibilité à la récompense dans les troubles bipolaires, observant que la plupart des études rapportent soit une **réduction de la sensibilité à la récompense**, soit une **sensibilité à la récompense inchangée** en période de rémission chez les patients par rapport aux sujets sains (Miskowiak et al., 2019). Les résultats en imagerie fonctionnelle sont cependant disparates, certaines études mettant en évidence une **hyperactivation** du striatum, du cortex orbitofrontal (OFC) et du cortex préfrontal (PFC) lors de **l'anticipation de la récompense** (Bermpohl et al., 2009; Caseras et al., 2013; Chase et al., 2013; Kollmann et al., 2017; Mason et al., 2014; Nusslock et al., 2012), d'autres montrant une **hypoactivation** du striatum, du cortex cingulaire antérieur (CCA) et de l'OFC (Chase et al., 2013; Dutra et al., 2015; Kollmann et al., 2017; Yip et al., 2015). Concernant la phase de **traitement de la récompense**, certains auteurs ont mis en évidence une **hyperactivation** du striatum et de l'OFC (Berghorst et al., 2016; Caseras et al., 2013; Dutra et al., 2015; Linke et al., 2012; Mason et al., 2014), tandis que d'autres ont mis en évidence une **hypoactivation** du striatum et de l'insula (Abler et al., 2008; Johnson et al., 2019; O'Sullivan et al., 2011; Redlich et al., 2015; Satterthwaite et al., 2015; Trost et al., 2014). Certains auteurs suggèrent ainsi un modèle **d'hyposensibilité à la récompense** dans le trouble bipolaire (Dutra et al., 2015; Johnson et al., 2019; Schreiter et al., 2016; Trost et al., 2014; Yip et al., 2015). Il a été suggéré que **le type de tâche expérimentale** utilisé, la prise en compte du **traitement pharmacologique** et le **sous-type de trouble bipolaire** pourraient expliquer au moins en partie ces différences (Johnson et al., 2019).

Concernant la **dépression bipolaire**, les résultats sont disparates, certaines études suggérant une sensibilité réduite à la récompense (Linke et al., 2011; Redlich et al., 2015; Satterthwaite et al., 2015), et d'autres une sensibilité accrue (Chase et al., 2013). Dans **l'(hypo)manie**, une hypersensibilité à la récompense est suggérée (Alloy et al., 2008; Bermpohl et al., 2009; Lozano & Johnson, 2001; Scott et al., 2000), mais certaines études suggèrent également une sensibilité inchangée par rapport aux sujets sains (Hägele et al., 2015) ou réduite (Abler et al., 2008).

Les résultats en matière de **punition** sont moins robustes, avec une **diminution de l'activité** du striatum et de l'OFC pour certains lors de **l'anticipation de la perte** (Bermpohl et al., 2009; Yip et al., 2015), sans différence par rapport aux sujets sains pour d'autres (Dutra et al., 2015; Kollmann et al., 2017; Nusslock et al., 2012; Redlich et al., 2015; Satterthwaite et al., 2015; Schreiter et al., 2016). Pendant la phase **d'évaluation de la punition**, Linke et al. (2012) ont mis en évidence une **hyperactivité de l'amygdale** (Linke et al., 2012), la plupart des études n'ayant trouvé aucune différence avec les sujets sains (Bermpohl et al., 2009; Dutra et al., 2015; Kollmann et al., 2017; Mason et al., 2014; Nusslock et al., 2012; Redlich et al., 2015; Satterthwaite et al., 2015; Schreiter et al., 2016; Yip et al., 2015).

4. Résumé de la partie III

Nous avons vu dans cette partie que de manière physiologique, nos choix influençaient notre humeur, mais également que notre humeur influençait nos choix ultérieurs, en lien avec **l'apprentissage qui découle des feedbacks**. Cela a été montré sur le plan expérimental mais également computationnellement. Une altération de cette boucle de rétroaction pourrait contribuer à **l'instabilité de l'humeur et à l'émergence des variations pathologiques de l'humeur dans les troubles bipolaires.**

Les études traitant de l'apprentissage par renforcement dans les troubles bipolaires ne sont toutefois pas univoques, notamment en lien avec des tâches expérimentales ne séparant pas l'apprentissage par récompense de celui de la punition, et ne prenant pas en compte l'impact du dernier épisode thymique. En effet, il a été montré que **la polarité du dernier épisode affectif biaise sélectivement la sensibilité des sujets aux résultats positifs/négatifs** pendant l'apprentissage par renforcement. La **sensibilité au renforçateur** pourrait également intervenir dans la disparité de ces résultats.

Il a été proposé qu'une **altération de la sensibilité à la récompense** dans le trouble bipolaire pouvait intervenir dans la genèse des épisodes thymiques, mais il n'est à ce jour

pas tranché s'il s'agirait davantage **d'une hypersensibilité ou d'une hyposensibilité à la récompense**. La modélisation computationnelle pourrait aider dans la compréhension de ce processus cognitif.

IV

Questions de recherche

Pour cette quatrième et dernière partie introductive, nous allons développer les différentes questions de recherche que nous nous sommes posées au cours de ce travail de thèse, en lien avec les apports théoriques que nous avons vus ci-dessus. Nous passerons ensuite en revue les études réalisées au cours de ce travail, avant de les discuter de façon générale au regard de la littérature scientifique existante sur le sujet.

1. Le sous-type de trouble bipolaire peut-il influencer les performances d'apprentissage par renforcement et la sensibilité à la récompense ?

a. L'apprentissage par renforcement est-il altéré ou non dans le trouble bipolaire en rémission ?

Nous avons vu qu'il apparaissait intéressant d'étudier le traitement de la récompense dans le trouble bipolaire, notamment durant la période de rémission. Nous avons également vu qu'au sein du processus cognitif de la motivation et du traitement de la récompense, il apparaissait plus précisément intéressant d'étudier **l'apprentissage en lien avec cette**

récompense, mais également en lien avec l'évitement de la punition. Nous avons cependant vu que la littérature à propos de l'apprentissage par renforcement dans le trouble bipolaire était très variée et qu'il n'était à ce jour **pas possible de trancher entre des performances altérées ou non en apprentissage par renforcement dans le trouble bipolaire en comparaison des sujets sains**. Nous proposons d'étudier cette première question au travers d'une **première étude** non-expérimentale mais **méta-analytique**, dans le but d'augmenter la puissance statistique de l'ensemble des études et tenter de répondre à la question de savoir si oui ou non l'apprentissage par renforcement apparait altéré dans le trouble bipolaire.

b. La sensibilité à la récompense pourrait-elle être différente entre le BD-I et le BD-II ?

Comme nous l'avons vu, certains auteurs émettent l'hypothèse **qu'une altération de l'apprentissage en lien avec le feedback pourrait induire et maintenir des épisodes thymiques pathologiques** (Eldar et al., 2016), **l'erreur de prédiction** semble être un élément essentiel dans l'apprentissage par renforcement, modulant potentiellement l'état affectif (Rutledge et al., 2014). Nous avons également vu que de part une approche *top-down* ou *bottom-up*, la principale différence pour poser le diagnostic entre un BD-I et un BD-II est l'absence d'épisode maniaque dans le BD-II par rapport au BD-I, au profit d'épisodes hypomaniaques, dont l'intensité (et l'impact) des symptômes en fait la principale différence (APA, s. d.). **Les sous-types de trouble bipolaire pourraient être conceptualisés comme un spectre dans lequel la sensibilité à la récompense pourrait être une dimension clé** (Nusslock & Alloy, 2017). En effet, bien que plusieurs études aient démontré des différences entre les BD-I et les BD-II en matière de génétique (Lee et al., 2010, 2011; Song et al., 2018), de neurobiologie (Chou et al., 2010) ou d'électrophysiologie (Ma et al., 2018), **seule une étude s'est intéressée à étudier leur différence selon le sous-type de trouble bipolaire lors du traitement de la récompense** (Caseras et al., 2013), retrouvant une activité du **striatum ventral** en ROI (*Region of interest*) **supérieure chez les BD-II** que les BD-I et les sujets sains lors de l'anticipation de la récompense. Nous nous demandons donc si les performances en apprentissage par renforcement et la sensibilité à la récompense pourraient être différentes entre les deux sous-types de trouble

bipolaire. Pour tenter de répondre à cette question, nous avons effectué une **deuxième étude** (première étude expérimentale), en comparant les performances d'apprentissage lors d'une tache **d'apprentissage probabiliste** précédemment utilisée dans de nombreuses études (Palminteri et al., 2012; Pessiglione et al., 2006), chez des patients en rémission ayant un BD-I, un BD-II, et des sujets sains contrôles.

2. L'agentivité pourrait-elle jouer un rôle dans les fluctuations de l'humeur pathologiques dans le trouble bipolaire ?

a. Définition du sens de l'agentivité

Comme nous l'avons vu, plusieurs études ont montré l'existence d'une **relation bidirectionnelle entre les choix que nous faisons et notre humeur**, pouvant intervenir dans le trouble bipolaire (Eldar et al., 2016; Eldar & Niv, 2015; Rutledge et al., 2014; Schultz, 2016; Vinckier et al., 2018). Parallèlement à cela, il a également été suggéré un **lien bidirectionnel entre le sens de l'agentivité (SdA) et l'humeur**.

Le SdA peut être défini comme la capacité à identifier que l'on est la cause d'une action ou d'une pensée, et de distinguer les conséquences des actions causées par soi de celles causées par les autres (Gallagher, 2000). Le SdA est donc un aspect fondamental des **représentations et du contrôle de l'action**. En effet, quelques études s'appuyant sur l'autodéclaration des sujets montrent que **les résultats d'actions positives par rapport aux résultats d'actions négatives augmentent le sentiment explicite d'agentivité** (voir (Kaiser et al., 2021) pour revue). En revanche, pour les études utilisant des mesures implicites, les résultats sont plus variés et en partie contradictoires. Il reste difficile de savoir si ces résultats contradictoires émanent de véritables différences dans l'expérience d'agentivité, ou bien de facteurs de confusion spécifiques aux mesures implicites elles-mêmes (Buehner & May, 2003; Kaiser & Schütz-Bosbach, 2018; Kok et al., 2012).

b. Sens de l'agentivité et traitement émotionnel

D'un autre côté, il a été suggéré que l**e SdA pouvait moduler les états émotionnels, ainsi que le traitement émotionnel** (**Figure IV.1**) (Kaiser et al., 2021). Plusieurs études ont ainsi regardé si le sentiment d'agentivité influençait les états émotionnels autodéclarés par les participants. La plupart de ces études ont révélé que le fait d'avoir un certain degré de choix sur ses actions et/ou un sentiment de contrôle sur les effets de l'action qui s'ensuit, entraine un affect plus positif (ou moins négatif) (Abelson et al., 2008; T. Li et al., 2021; Stolz et al., 2020; Thuillard & Dan-Glauser, 2017, 2021). Au niveau neuronal, il a été constaté que la simple **anticipation** de pouvoir faire un choix augmente l'activité dans les régions du cerveau qui sont liées au traitement de la récompense, comme le striatum ventral (Bjork & Hommer, 2007; Leotti & Delgado, 2014; Lorenz et al., 2015; Romaniuk et al., 2019; Stolz et al., 2020; Tricomi et al., 2004; Wang & Delgado, 2019). Dans l'ensemble, ces études suggèrent qu'un **sens accru de l'agentivité est communément vécu comme désirable**, et conduit à une augmentation de l'affect positif (Leotti et al., 2010). Cependant, les effets positifs de l'agentivité dans les choix peuvent potentiellement être diminués, **voire inversés**, dans des contextes où une trop grande liberté de choix augmente significativement la difficulté de la tâche (Greifeneder et al., 2010).

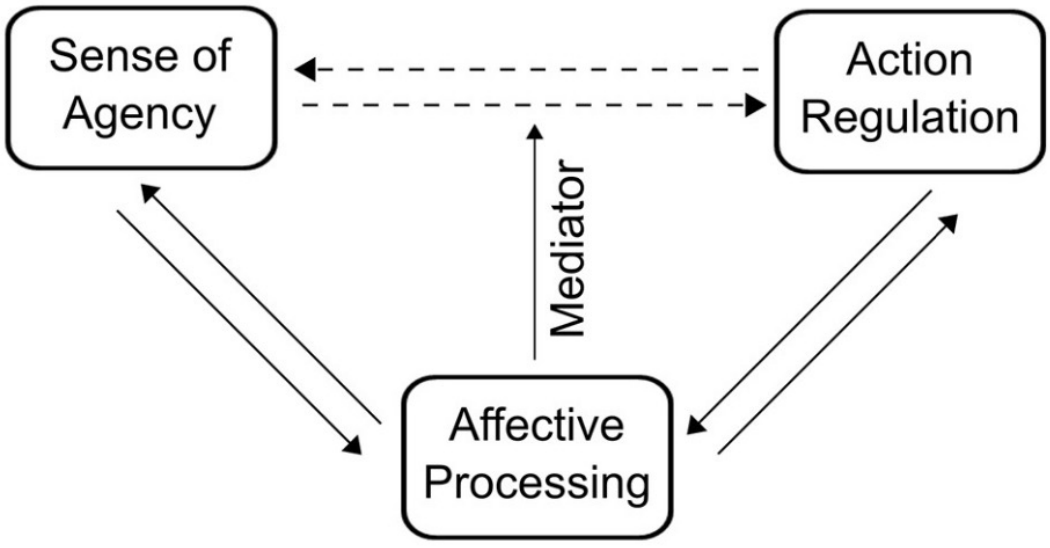

Figure IV.1 : Représentation schématique de la relation proposée entre le sentiment d'agentivité, le traitement affectif et la régulation de l'action. Le modèle suppose une relation bidirectionnelle entre le traitement affectif et le sentiment d'agentivité, ainsi qu'entre le traitement affectif et la régulation de l'action. Il est important de noter que le traitement

affectif joue un rôle de médiateur partiel dans l'influence entre le sentiment d'agentivité et la régulation de l'action. *Figure issue de (Kaiser et al., 2021).*

c. Sens de l'agentivité et feedback affectif

Plusieurs études ont cherché à savoir si le SdA augmente ou diminue la **sensibilité aux stimuli affectifs**. La plupart des études portant sur cette question ont manipulé le SdA des participants pour des feedbacks positifs ou négatifs lors de tâches d'apprentissage ou de jeu associées à de l'imagerie fonctionnelle (Kaiser & Schütz-Bosbach, 2019; Mühlberger et al., 2017; Polich, 2007; Proudfit, 2015). Ces résultats suggèrent que **le SdA augmente l'impact neuronal du feedback affectif**. Le renforcement neuronal pour les stimuli affectifs a le plus souvent été rapporté dans des études qui manipulaient l'agentivité au niveau du choix et/ou du résultat, par exemple en comparant des **tâches à choix libre avec des tâches à choix forcé** (P. Li et al., 2011; Mei et al., 2018; Mühlberger et al., 2017). Par conséquent, l'atténuation neuronale par rapport à l'amélioration neuronale pourrait être en partie liée au type d'agentivité en lien avec les **comportements vers un but** qui est manipulé (agentivité motrice / au niveau des choix / du résultat) (Hassall et al., 2019).

Alors que de nombreuses études montrent **qu'un SdA plus élevé augmente l'impact neuronal du feedback affectif**, il est moins clair si ces effets liés à l'agentivité sur le traitement affectif sont aussi forts pour les stimuli positifs que négatifs. Il apparait important de déterminer si l'expérience de l'agentivité conduit à une amélioration sélective du feedback positif ou négatif, car une telle conclusion impliquerait que le SdA génère un biais de traitement spécifique à la valence. Une étude a mis en avant un biais de positivité lié à l'agentivité dans une tâche d'apprentissage par renforcement pouvant traiter séparément récompense et punition. Un sentiment d'agentivité élevé par rapport à un faible sentiment d'agentivité a conduit à des **augmentations sélectives des taux d'apprentissage après un retour positif, mais pas négatif** (Chambon et al., 2020). L'agentivité (que le choix soit autodéterminé ou imposé) ainsi que la valence (punition ou récompense) joueraient ainsi un rôle important dans **l'apprentissage**. En revanche, **on ne sait toujours pas comment l'agentivité affecte l'humeur pendant une tache d'apprentissage**. C'est ce que nous proposons d'étudier dans une **troisième étude**

(deuxième étude expérimentale) dont les résultats sont encore préliminaires. Dans cette étude, nous avons fait passer une tâche d'apprentissage probabiliste à des personnes ayant un trouble bipolaire et des sujets sains contrôles, tout en manipulant l'agentivité et en évaluant régulièrement l'humeur des sujets.

ETUDES EXPERIMENTALES

V

Reinforcement learning in bipolar disorder: a systematic review and meta-analysis of behavioral studies

Arnaud Pouchon[a,b], Clément Dondé[a,b,c], Julien Bastin[d,1], Mircea Polosan[a,b,1]

a Univ. Grenoble Alpes, Inserm, U1216, CHU Grenoble Alpes, Grenoble Institut Neurosciences, 38000 Grenoble, France

b Department of Psychiatry, CHU Grenoble Alpes, 38000 Grenoble, France

c Department of Psychiatry, CH Alpes-Isère, 38000 Saint-Egrève, France

d Univ. Grenoble Alpes, Inserm, U1216, Grenoble Institut Neurosciences, 38000 Grenoble, France

1 These authors contributed equally to this work.

In preparation

Abstract

Introduction: Bipolar disorder (BD) appears to be characterized by impaired reinforcement learning (RL). However, there is no consensus whether patients with BD have impaired performance compared with healthy subjects (HC). Several factors may contribute to this ambiguity, including thymic state and the valence (reward and punishment). The aim is to decide whether or not RL is impaired in BD.

Method: For this systematic review and meta-analysis, the PubMed, Cochrane, Embase and PsychINFO databases were searched from inception to March 2023. Consensual criteria for inclusion were peer-reviewed studies published in English that used a computerized RL behavioral paradigm and compared individuals with BD with HC. Data were extracted and pooled using random-effects sizes. The main outcomes were performance on RL task, measured by Hedges g effect size.

Results: Twenty-six studies involving 2710 participants were included: 1309 individuals with BD and 1401 HC subjects. We put forward a significant and important impairment of reinforcement learning in BD ($k = 26$; effect size = -1.29; 95% CI = -1.90 to -0.69; $p < 0.001$), notably during euthymia ($k = 17$; effect size = -1.19; CI 95% = -1.86 to -0.52; $p < 0.01$). About the valence, we highlight a significant impairment of reward learning during euthymia ($k = 17$; effect size = -0.85; CI 95% = 1.59 to -0.11; $p < 0.05$), and no alteration for punishment learning ($k = 5$; effect size = -1.18; CI 95% = -2.63 to 0.28; $p = 0.11$).

Discussion: In this systematic review and meta-analysis, BD was associated with deficits in RL, notably during euthymic state, with a selective deficit in reward-based learning versus punishment-based learning. Understanding the cognitive and neurobiological mechanisms driving RL impairments may assist in developing novel interventions.

Keywords

Bipolar disorder; Reinforcement learning; Reward learning; Punishment learning; Euthymia; Systematic review; Meta-analysis

1. Introduction

Bipolar disorder (BD) is a chronic mood disorder characterized by alternating episodes of depression and mania or hypomania, for which the hedonic dimension is a key symptom for clinical diagnosis according to criteria A of the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) (APA, s. d.). Indeed, the hedonic dimension is characterized in depressive disorder by motivational anhedonia (reduced motivation to engage in activities of daily living) and consummatory anhedonia (reduced pleasure from usually pleasurable activities) (Ho & Sommers, 2013). In contrast, (hypo)manic episodes are characterized by hyperhedonia, with an increase in pleasure-seeking behaviors (APA, s. d.).

Reward processing therefore appears to be strongly impaired in BD (Whitton et al., 2015). Among the various stages of reward processing, the learning that results from these rewards (i.e. reinforcement learning), seems to be particularly affected and could largely contribute to relapse (Mason et al., 2017). Indeed, reinforcement learning (RL) describes the process by which an individual uses feedback to change their behavior in the future. Changes in behavior over time are assumed to reflect the updating of value expectations assigned to available behaviors (Maia & Frank, 2011).

Given the bidirectional relationship between mood fluctuations and RL (Blain & Rutledge, 2020; Eldar et al., 2016; Rutledge et al., 2014), several studies have focused on identifying reinforcement learning deficits in BD. However, these studies have shown disparate results regarding reinforcement learning performance (Adida et al., 2011; Brambilla et al., 2013; Duek et al., 2014; Lewandowski et al., 2016; Linke et al., 2012; Pizzagalli et al., 2008; Ryu et al., 2017; Yechiam et al., 2008). These discrepancies can be partly explained by the heterogeneity of the tasks used. Indeed, although these are paradigms concerning reward-related learning and/or punishment avoidance, they do not necessarily reflect the same underlying cognitive process (Johnson et al., 2019; Richards et al., 2013). Moreover, some paradigms include only reward-based learning and not punishment-avoidance learning, or mix the two, so that the specificity of the cognitive

processes altered in BD during learning is unclear. One reason to examine reward- and punishment-based learning distinctly is that rewards and punishments are also known to have an opposite impact on subjects' moods (Blain & Rutledge, 2020; Cecchi et al., 2022; Vinckier et al., 2018), so an asymmetry between reward- and punishment-based learning processes may play a role in the genesis of different thymic states and their maintenance (Eldar et al., 2016). Finally, this could be explained by a number of factors, first and foremost whether or not patients' thymic state is taken into account, given the hedonic differences observed between depressive disorder and (hypo)mania.

In this work, we propose to take stock of all the studies dealing with reinforcement learning in BD and to quantify their results, taking into account the various biases we have outlined. The aim is to decide whether or not reinforcement learning is impaired in BD.

2. Method

We followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines when conducting the review (Moher et al., 2009). The protocol was prospectively registered on PROSPERO (ID: CRD42021264430). All editors developed and approved the inclusion, exclusion, and non-inclusion standards (see corresponding section). Separate searches were performed in the various databases. The AP, and CD authors independently checked the title, abstract and full text of each article and assessed its eligibility. After the selection stage, disagreements concerning the eligibility of articles were discussed, and all disagreements were resolved by consulting a third author (JB or MP).

a. Search strategy

We used PubMed, Cochrane, Embase and PsychINFO databases regardless of year to conduct the systematic literature review before March 18, 2023. The search technique involved using a combination of specific keywords with Boolean operators to search for information on reward, bipolar disorder, and behavioral studies. The specific keywords

used to search the databases were: (deci* or reward* or motivat* or incentiv* or effort*) and (bipol*) and (task* or paradigm* or battery*). In order to find other relevant documents, reference lists of articles were also examined.

b. Inclusion and non-inclusion criteria

Studies were eligible if they met the following criteria: (1) articles that included adult patients (>18 years) diagnosed with bipolar disorder according to criteria based on the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV or DSM-5) or the International Classification of Diseases (ICD-10 or ICD-11), (2) articles that investigated the reward system at behavioral level by asking patients with BD to perform a task compared with healthy subjects. Exclusion criteria were: (1) case reports, book chapters, reviews, or meta-analyses, (2) studies not written in English or French, (3) participants under 18 years of age, studies not performing a task related to the reward system. Each study's quality was assessed according to the Newcastle-Ottawa Scale adapted for case-control studies (Ma et al., 2020; Wells G, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al., s. d.).

c. Data extraction

The outcome of the meta-analysis was defined and extracted as the mean and standard deviation of performance in learning tasks related to reward, and also related to punishment where possible (or analysis of variance [ANOVA] if these data were not available). Task performance measures differed according to the tasks chosen. For the Iowa Gambling Task, the "net score" was extracted, which is taken directly from the study data, or calculated by subtracting the number of disadvantageous choices (games A and B) from the number of advantageous choices (games C and D), i.e. (C+D)-(A+B) (Bechara et al., 1994). For the Probabilistic classification task (PCT), we extracted the optimal response scores (Duek et al., 2014). For the Probabilistic Reward Task (PRT), we extracted the "global response bias" (Lewandowski et al., 2016; Pizzagalli et al., 2008; Ryu et al., 2017). For the Probabilistic reward-punishment task in the study by Abohamza et al. (2020) (Abohamza et al., 2020), we extracted the percentage of correct answers, as we did for the Probabilistic

learning task in the study by Geana et al. (2022) (Geana et al., 2022). For the Probabilistic Selection Task (PST), we extracted the performance during the transfer phase (Linke et al., 2011; Strauss et al., 2015). Finally, for the Probabilistic Reinforcement Learning Task (PBLT), we extracted the percentage of performance in the reward condition during the test phase (Palminteri et al., 2015). Details of the various tasks are given in the supplementary materials. Where possible, we extracted the mean and standard deviation of behavioral performance directly from the article data or by requesting the raw data from the authors. Where these results were not available, we extracted the metrics on the figures provided in the articles, using the WebPlotDigitizer program (https://apps.automeris.io/wpd/). This program has shown high levels of reliability and validity between different coders (Drevon et al., 2017).

We also extracted demographic data (age, gender, and education level), total scores of psychometric scales assessing mood, but did not have time to explore the impact of antipsychotic treatments due to the risk of interference with reward processing (Pessiglione et al., 2006). Psychometric scales assessing depressive symptoms were converted to the equivalent of the Hamilton Depression Rating Scale (HAM-D) (Leucht et al., 2018).

d. Meta-analysis

All analyses were performed using the compute.es (Del Re, A. C., 2010) and metafor (Viechtbauer, 2010) packages in Rstudio, version 2023.6.1.524 (R Project for Statistical Computing) (RStudio Team, 2020), using random-effects models. The primary outcome of the meta-analysis was defined and analyzed as the mean and standard deviation of performance in reinforcement learning tasks (or the ANOVA result if not available). These values were transformed into Hedges g effect sizes for each included study. Next, a random-effects model was used to give an overall effect size. The significance level was set at $P < .05$ on a two-tailed basis. The effect size was interpreted in accordance with Cohen's guidelines (0.2, small; 0.5, medium; and 0.8, large) (Cohen, J., 1988).

A subgroup meta-analysis using mixed-effects models was carried out to determine whether the paradigm used could affect the overall effect size. Overall effects from independent meta-analyses were obtained by fitting separate random-effects models in each subset of studies. Effect sizes and SD from individual studies were combined within each model to obtain an overall effect size per model. Omnibus tests were performed to analyze whether the summarized effect sizes differed significantly.

Separate meta-analyses were carried out to compare group performance by separating tasks using IGT from other probabilistic learning tasks, by separating different thymic states (depression, euthymia, (hypo)mania) because of the definite influence of thymic state on performance (Adida et al., 2011), and by valence, with reward learning likely to be treated differently from punishment learning (Gueguen et al., 2021; Palminteri et al., 2012, 2015; Pessiglione et al., 2006). Mixed-effects meta-regression models were used to assess whether potential continuous confounders were associated with reinforcement learning performance, including the type of task, and the depressive and (hypo)manic symptoms.

Overall heterogeneity in summary effect size was quantified using Higgins' I^2 statistic, with values below 40%, 30% to 60%, 50% to 90%, and above 75% reflecting heterogeneity that might not be significant, moderate, substantial, and considerable, respectively (Deeks et al., 2019). Cochran's Q test was used to assess heterogeneity, and the significance level was set at P=.10. Baujat plots were performed to identify the studies that most influenced overall effect size and contributed to heterogeneity (Baujat et al., 2002). Publication bias was assessed by visual inspection of the funnel plot, and an Egger regression test was performed to estimate the magnitude of potential publication bias (Egger et al., 1997).

3. Preliminary results

a. Study selection

The initial search identified 533 studies published between 1975 and 2023. After removing duplicates and checking eligibility criteria, 33 articles were retained for the qualitative synthesis of the search. Of these, 2 were excluded due to overlapping samples between studies, one due to a lack of precision regarding the paradigm score used, and 4 dues to a lack of metrics to extract after contacting the authors. Overall, 26 studies contained extractable data and were included in the meta-analysis (**Figure V.V.1**).

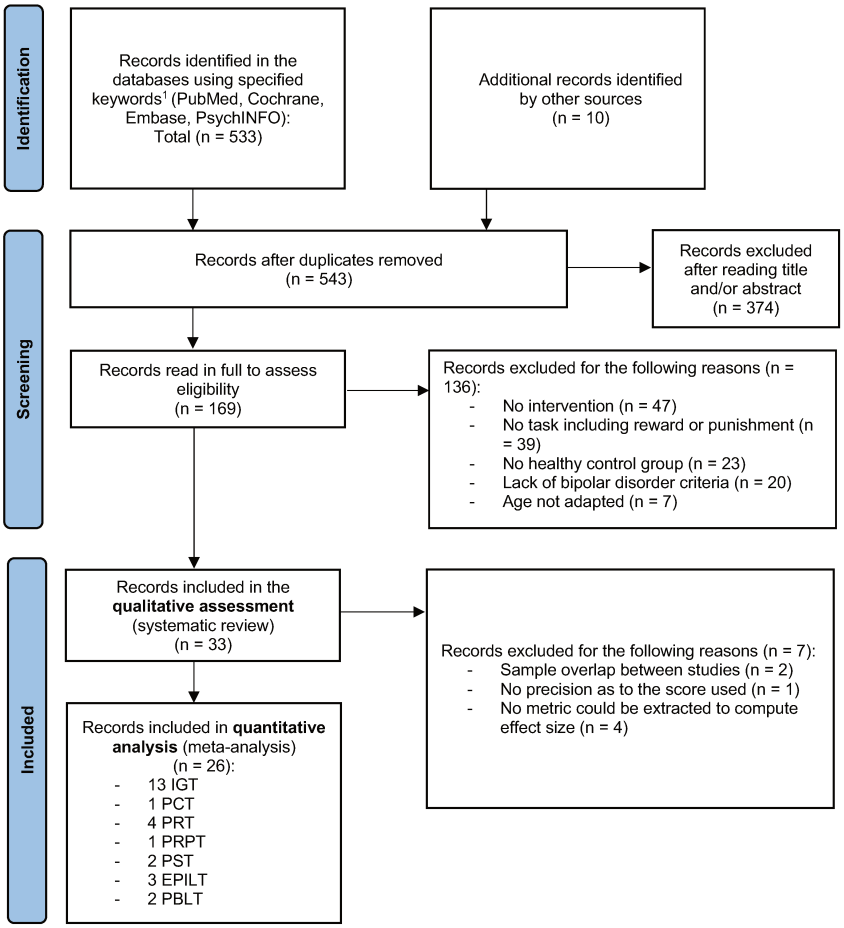

Figure V.V.1 : PRISMA flowchart of systematic literature review.

[1]Specified keywords used to search databases: (deci* or reward* or motivat* or incentiv* or effort*) and (bipol*) and (task* or paradigm* or battery*). IGT = Iowa Gambling Task; PCT = Probabilistic Classification Task; PRT = Probabilistic Reward Task; PRPT = Probabilistic Reward Punishment Task; PST = Probabilistic Selection Task; EPILT = Explicit Probabilistic Incentive Learning Tasks; PBLT = Probabilistic Learning Task.

The studies included in the meta-analysis represented a total of 2710 participants, including 1309 people with BD, and 1401 healthy control (HC) subjects, with an average of 48% male in the BD group and 52% male in the HC group, and an average age of 42.70 years old (SD = 12.50) in the BD group and 38.12 years old (SD = 12.51) in the HC group. Among patients with BD, the mean (SD) HAM-D score was 5.21 (3.80), and YMRS 6.38 (3.81). The types of reinforcement learning paradigms used were the Iowa Gambling Task (IGT) (13 studies) (Adida et al., 2011; Brambilla et al., 2013; Caletti et al., 2013; Gomide Vasconcelos et al., 2014; Gu et al., 2020; Ibanez et al., 2012; Jogia et al., 2012; Malloy-Diniz et al., 2011; Martino et al., 2011; Ono et al., 2015; Powers et al., 2013; Van Enkhuizen et al., 2014; Yechiam et al., 2008), the Probabilistic Classification (PCT) (1 study) (Duek et al., 2014), the Probabilistic Reward Task (PRT) (4 studies) (Lewandowski et al., 2016; Pizzagalli et al., 2008; Ryu et al., 2017; Whitton et al., 2021), the Probabilistic Reward Punishment Task (PRPT) (1 study) (Abohamza et al., 2020), the Probabilistic Selection Task (PST) (2 studies) (Linke et al., 2011; Strauss et al., 2015), the Explicit Probabilistic Incentive Learning Tasks (EPILT) (3 studies) (Barch et al., 2017; Geana et al., 2022; Pratt et al., 2021), and the Probabilistic Learning Task (PBLT) (2 studies) (Pouchon et al., 2023) ; Pouchon et al., in preparation) (**Table V.1**). Each study's quality assessment is available in **Table V.2** in supplementary information.

Table V.1 : Characteristics of Included Studies.

	Bipolar disorder			Heathy controls								
Source	**No.**	**Male, %**	**Age, mean (SD), years**	**No.**	**Male, %**	**Age, mean (SD), years**	**Thymic state**	**Task**	**Learning (type)**	**Outcome**	**Effect size (Hedges g)**	**P value**
Adida et al., 2011	167 (45 M, 32 D, 90 E)	41.3	40.3 (11.6)	150	50	38.8 (10.6)	Mania, euthymia, depression	IGT	Reward learning	Net score	-5,18	<0,01
Brambilla et al., 2013	70	53	44.6 (11.3)	140	51	43.9 (11.2)	Euthymia	IGT	Reward learning	Net score	0.19	0.18
Caletti et al., 2013	18	22	42.22 (11.72)	18	33	36.11 (14.51)	Euthymia	IGT	Reward learning	Net score	-0.17	0.61
Gomide Vasconce	50	54	33.09 (13.48)	256	54	33.09 (13.48)	Euthymia	IGT	Reward learning	Net score	-0.43	0.01

los et al., 2014												
Gu et al., 2020	29	62	35.72 (9.65)	34	65	33.79 (9.09)	NA	IGT	Reward learning	Net score	-0.69	0.01
Ibanez et al., 2012	13	62	40.1 (9.4)	25	64	35.1 (11.2)	Euthymia	IGT	Reward learning	Net score	-0,58	0.10
Jogia et al., 2012	36	47	42.5 (10.6)	37	57	37.6 (11.3)	Euthymia	IGT	Reward learning	Net score	-0.35	0.14
Malloy-Diniz et al., 2011	95	34	41 (12)	94	47	32 (13)	Euthymia	IGT	Reward learning	Net score	-0.69	<0.01
Martino et al., 2011	85 (48 BD -I, 37 BD -II)	30.45	40.25 (15.7)	34	35.3	40.0 (12.9)	Euthymia	IGT	Reward learning	Net score	0.21	0.29
Ono et al., 2015	13	46	38.4 (7.3)	15	53	32.9 (7.7)	Euthymia	IGT	Reward learning	Net score	-0.32	0.39
Powers et al., 2013	98	47	40.4 (12.09)	95	56	38.29 (11.49)	NA	IGT	Reward learning	Net score	-0.54	0.04
Van enkhuizen et al., 2014	16	56	33.8 (2.8)	17	29	33.9 (3.0)	Mania	IGT	Reward learning	Net score	0.13	
Yechiam et al., 2008	28	36	44.05 (9.05)	25	32	39.2 (13.3)	Mania, euthymia, depression	IGT	Reward learning	Net score	0.02	0.96
Lewandowski et al., 2016	42	45	29.6 (8.4)	29	41	31.0 (10.0)	NA	PRT	Probabilistic reward learning	Total response bias	-0.36	0.14
Pizzagalli et al., 2008	13	62	38.77 (12.09)	25	56	38.36 (10.76)	Euthymia	PRT	Probabilistic reward learning	Total response bias	-0.85	0.02
Ryu et al., 29017	44	43	34.65 (6.76)	24	46	31.9 (6.96)	Mania, euthymia	PRT	Probabilistic reward learning	Total response bias	-0.20	0.42
Whitton et al., 2021	104	46.95	40.81 (13.37)	129	37.2	32.02 (12.13)	NA	PRT	Probabilistic reward learning	Total response bias	-0.81	<0.01
Duek et al., 2014	40	55	42 (11.73)	41	56	38.71 (11.22)	Euthymia	PCT	Probabilistic reward et punishment learning **separated**	Optimal response scores	-2.43	<0.01
Abohamza et al., 2019	43	NA	43.86 (4.56)	20	65	43.85 (6.8)	NA	PRPT	Probabilistic reward et punishment learning **separated**	Probabilistic learning task performance	-4.24	<0.01
Barch et al., 2017	43	44	35.4 (10.0)	55	54	36.0 (11.1)	NA	EPIL T	Probabilistic reward et punishment learning **separated** (learning phase then transfert phase)	Accuracy (learning phase) reward and punishment	-0.66	<0.01
Geana et al., 2021	60	NA (sup mat)	NA (sup mat)	72	NA (sup mat)	NA (sup mat)	NA	EPIL T	Probabilistic reward et	Accuracy (learning	-5.18	<0.01

									punishment learning **separated** (learning phase then transfert phase)	phase) reward and punishment		
Pratt et al., 2021	62	35.5	38.3 (10.7)	75	54.7	37.4 (11.3)	NA	EPILT	Probabilistic reward et punishment learning **separated** (learning phase then transfert phase)	Accuracy (learning phase) reward and punishment	-1.99	<0.01
Linke et al., 2011	23	48	44.1 (8.1)	19	47	43.1 (11.6)	Euthymia	PST	Probabilistic reward et punishment learning **separated** (learning phase then test phase)	Accuracy (learning phase) reward and punishment	0.12	0.70
Strauss et al., 2015	47 (24 BD + et 23 BD -)	32	35.75 (13.3)	24	45.8	36.1 (13.4)	Euthymia	PST	Probabilistic reward et punishment learning **separated** (learning phase then test phase)	Accuracy (learning phase) reward and punishment	-2.14	<0.01
Pouchon et al. 2023	79 (45 BD -I et 34 BD -II)	58	46.39 (10.2)	30	67	44.35 (10.69)	Euthymia	PBLT	Probabilistic reward et punishment learning **separated**	Percentage of correct choices	-3.74	<0.01
Pouchon et al. (in prep)	32	56	45,09 (13,03)	32	59	44,38 (12,02)	Euthymia	PBLT	Probabilistic reward et punishment learning **separated**	Percentage of correct choices	-1.64	<0.01

Abbreviations: IGT = Iowa Gambling Task; PCT = Probabilistic Classification Task; PRT = Probabilistic Reward Task; PRPT = Probabilistic Reward Punishment Task; PST = Probabilistic Selection Task; EPILT = Explicit Probabilistic Incentive Learning Tasks; PBLT = Probabilistic Learning Task. NA = Not available.

b. Meta-analysis

Taken together, this meta-analysis highlights a significant and important impairment of reinforcement learning in BD (k = 26; effect size = -1.29; 95% CI = -1.90 to -0.69; $p < 0.001$). Heterogeneity between studies was significant and interpreted as very large (I2 = 97.8%; Q-test = $p < 0.01$). The Baujat plot indicates that two studies mainly influenced the overall effect size and contributed to heterogeneity (Abohamza et al., 2020; Pouchon et al.,

2023). After the elimination of these two studies, the magnitude of the overall effect size remains significant and large (k = 24; effect size = -1.08; 95% CI = -1.65 to -0.50; p < 0.001).

Subgroup meta-analysis revealed impaired reinforcement learning performance across all tasks except PST and EPILT, with effect sizes of -1.01 (k = 2; 95% CI = -3.22 to 1.20; p = 0.370) and -2.60 (k = 3; 95% CI = -5.21 to 0.02; p = 0.052) respectively. Performances appeared most impaired in PRPT (k = 1; effect size = -4.24; 95% CI = -5.14 to -3.34; p < 0.001), followed by PBLT (k = 2; effect size = -2.69; 95% CI = -4.74 to -0.63; p = 0.011), PCT (k = 1 ; effect size = -2.43; 95% CI = -2.98 to -1.88; p < 0.001), IGT (k = 13; effect size = -0.76; 95% CI = -1.50 to -0.01; p = 0.048), and finally PRT (k = 4; effect size = -0.56; 95% CI = -0.89 to -0.24; p = 0.001). These results can be seen in **Figure V.V.2**. Visual inspection of the funnel plot and Egger's regression test (intercept = -1.88; p = 0.06) suggest no publication bias (**Figure V.V.3**).

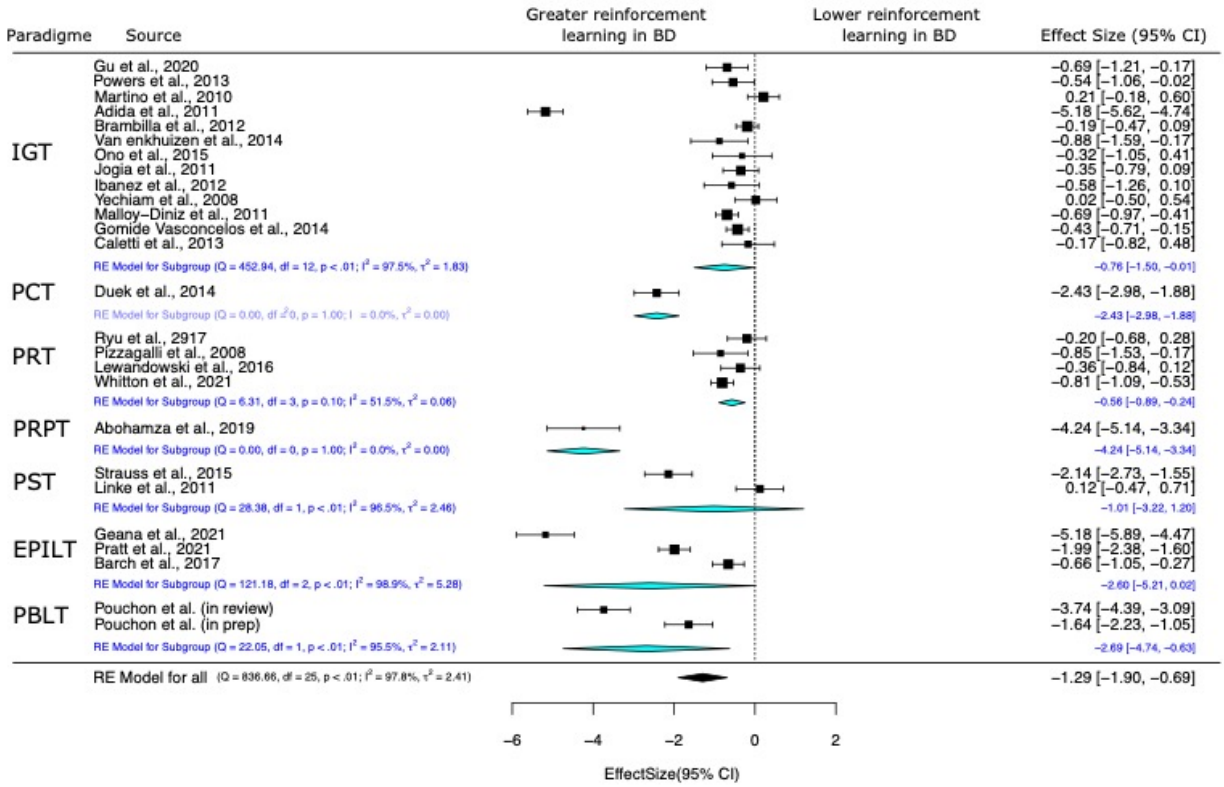

Figure V.V.2: Forest Plot of subgroup meta-analysis according to task type about reinforcement learning performances among patients with bipolar disorder (BD) and heathy controls. IGT = Iowa Gambling Task; PCT = Probabilistic Classification Task; PRT = Probabilistic Reward Task; PRPT = Probabilistic Reward Punishment Task; PST = Probabilistic Selection Task; EPILT = Explicit Probabilistic Incentive Learning Tasks; PBLT = Probabilistic Learning Task. NA = Not available.

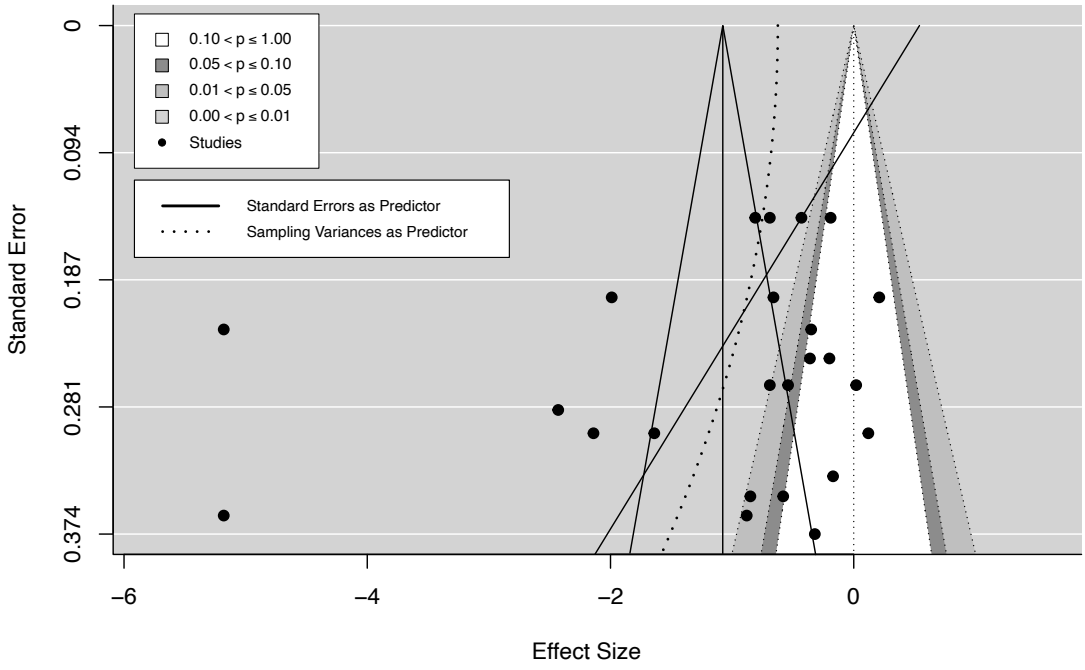

Figure V.V.3: Funnel plot of meta-analysis for visual inspection of publication bias. The solid black vertical line represents the observed overall summary effect. The unfilled funnel represents 95% confidence intervals for individual effect estimates, assuming no bias. The dashed black line represents the summary effect when including imputed studies (using the trim-and-fill method). The solid black non-vertical line represents the Egger regression line. The shades of grey funnel represent the differences in confidence intervals for individual effect estimates, including imputed studies.

Meta-regression revealed no significant association between reinforcement learning performance and depressive symptoms ($k = 16$; $\beta1 = -0.10$; $p = 0.36$), and (hypo)mania ($k = 19$; $\beta1 = -0.01$; $p = 0.80$). On the other hand, a significant association was found between reinforcement learning performance and the type of task used, with an influence of IGT on overall effect size ($k = 26$; $\beta1 = 1.82$; $p < 0.05$), but not of other tasks.

Because of the influence of IGT on the results, we carried out a separate meta-analysis with only the studies that used IGT, and one with the other studies. About studies using IGT, 13 studies were included (Adida et al., 2011; Brambilla et al., 2013; Caletti et al., 2013; Gomide Vasconcelos et al., 2014; Gu et al., 2020; Ibanez et al., 2012; Jogia et al., 2012; Malloy-Diniz et al., 2011; Martino et al., 2011; Ono et al., 2015; Powers et al., 2013; Van Enkhuizen et al., 2014; Yechiam et al., 2008). There was no impairment of RL performances

(i.e. by reward) compared with HC (k = 13; effect size = -0.48; CI 95% = -1.32 to 0.35; p = 0.259). The Baujat plot indicates that one study primarily influenced the overall effect size and contributed to heterogeneity (Martino et al., 2011). After the suppression of this outlier study, the magnitude of the overall effect size remains insignificant (k = 12; effect size = -0.65; CI 95% = -1.49 to 0.18; p = 0.127). Heterogeneity between studies was significant and interpreted as very large (I^2 = 98.20%; Q-test = p < 0.01). Visual inspection of the funnel plot and Egger's regression test (intercept = -0.63; p = 0.91) did not suggest publication bias.

About studies using other probabilistic learning tasks, 13 studies were included (Duek et al., 2014 ; Lewandowski et al., 2016; Pizzagalli et al., 2008; Ryu et al., 2017; Whitton et al., 2021; Abohamza et al., 2020; Linke et al., 2011; Strauss et al., 2015; Barch et al., 2017; Geana et al., 2022; Pratt et al., 2021; Pouchon et al., 2023; Pouchon et al., in preparation). This meta-analysis highlights a significant and important impairment of reinforcement learning in BD (k = 13; effect size = -1.71; 95% CI = -2.63 to -0.78; p < 0.001). Heterogeneity between studies was significant and interpreted as very large (I2 = 97.8%; Q-test = p < 0.01). The Baujat plot indicates that three studies mainly influenced the overall effect size and contributed to heterogeneity (Abohamza et al., 2020; Linke et al., 2011; Ryu et al., 2017). After the suppression of these three studies, the magnitude of the overall effect size remains significant and large (k = 10; effect size = -1.80; 95% CI = -2.65 to -0.95; p < 0.001). Visual inspection of the funnel plot and Egger's regression test (intercept = 1.07; p = 0.06) did not suggest publication bias.

We then conducted separate meta-analyses according to thymic state. With regard to depressive episodes, no meta-analysis was carried out, as only the study conducted by Adida et al. (2011) included patients with a diagnosis of bipolar depression, with the possibility of extracting the score (Adida et al., 2011). In this study, patients with depression performed less well than healthy subjects, with a large effect size (k = 1; effect size = -5.49; CI 95% = -5.98 to -5.01; p < 0.001). The study conducted by Yechiam et al. (2008), distinguished between euthymic and "acute" patients, but did not separate depression from mania (Yechiam et al., 2008).

About manic or hypomanic episodes, 3 studies were included (Adida et al., 2011; Ryu et al., 2017; Van Enkhuizen et al., 2014). No difference in RL performance was found between patients with a (hypo)manic episode and HC (k = 3; effect size = -2.53; CI 95% = -6.46 to 1.40; p = 0.21). Heterogeneity between studies was significant and interpreted as highly significant (I^2 = 99.29%; Q-test = p < 0.01). Visual inspection of the funnel plot and Egger's regression test (intercept = -4.20; p = 0.92) did not suggest publication bias.

Regarding euthymia, 17 studies were included (Adida et al., 2011; Brambilla et al., 2013; Caletti et al., 2013; Duek et al., 2014; Gomide Vasconcelos et al., 2014; Ibanez et al., 2012; Jogia et al., 2012; Linke et al., 2011; Malloy-Diniz et al., 2011; Martino et al., 2011; Ono et al., 2015; Pizzagalli et al., 2008; Pouchon et al., 2023; Pouchon et al., in preparation; Ryu et al., 2017; Strauss et al., 2015; Yechiam et al., 2008). A significant impairment in RL performance was found in euthymic patients compared with HC, with a large effect size (k = 17; effect size = -1.19; CI 95% = -1.86 to -0.52; p < 0.01) (**Figure V.V.4**). Heterogeneity between studies was significant and interpreted as very large (I^2 = 96.37%; Q-test = p < 0.01). Visual inspection of the funnel plot and Egger's regression test (intercept = -4.20; p = 0.92) did not suggest publication bias.

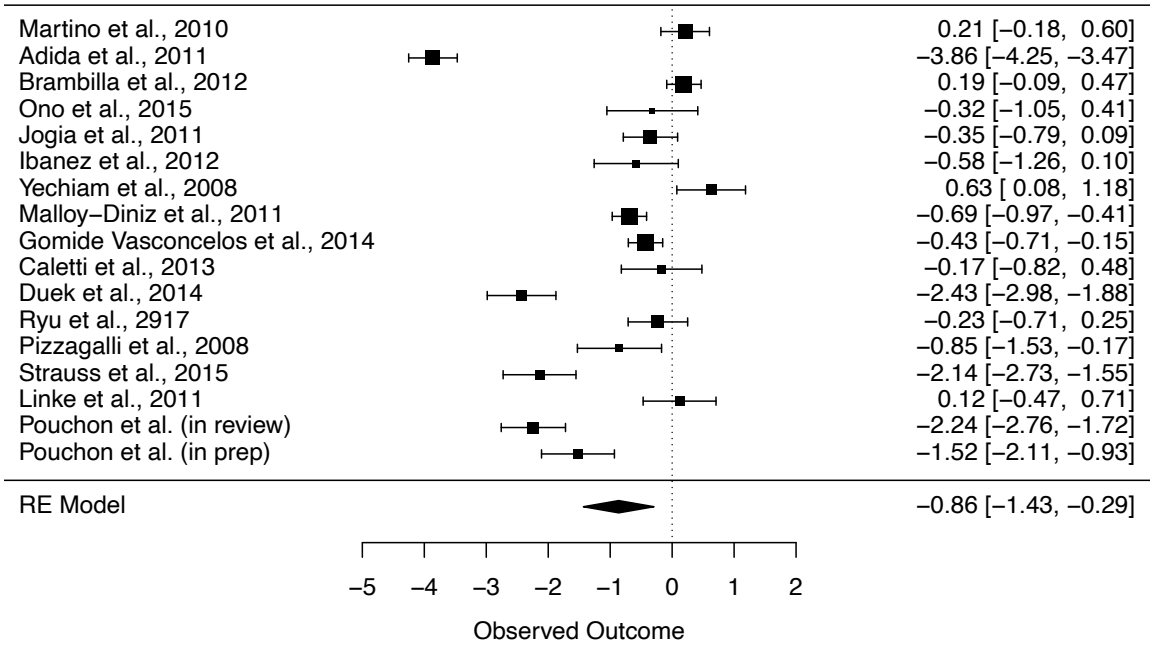

Figure V.V.4: Forest plot of meta-analysis about reinforcement learning performances among patients with bipolar disorder only during euthymic period and heathy controls.

Meta-analyses were then carried out, taking into account valence (reward vs. punishment learning). Concerning reward learning, 26 studies were included (Adida et al., 2011; Brambilla et al., 2013; Caletti et al., 2013; Gomide Vasconcelos et al., 2014; Gu et al., 2020; Ibanez et al., 2012; Jogia et al., 2012; Malloy-Diniz et al., 2011; Martino et al., 2011; Ono et al., 2015; Powers et al., 2013; Van Enkhuizen et al., 2014; Yechiam et al., 2008; Duek et al., 2014 ; Lewandowski et al., 2016; Pizzagalli et al., 2008; Ryu et al., 2017; Whitton et al., 2021; Abohamza et al., 2020; Linke et al., 2011; Strauss et al., 2015; Barch et al., 2017; Geana et al., 2022; Pratt et al., 2021; Pouchon et al., 2023; Pouchon et al., in preparation). A significant impairment of reward learning performance was found in patients compared to HC, with a large effect size (k = 26; effect size = -1.19; CI 95% = -1.86 to -0.52; $p < 0.01$). Heterogeneity between studies was significant and interpreted as very large (I^2 = 96.37%; Q-test = $p < 0.01$). The Baujat plot indicates that one study primarily influenced the overall effect size and contributed to heterogeneity (Geana et al., 2022). After the suppression of this outlier study, the magnitude of the overall effect size remains significant and large (k = 25; effect size = -1.07; CI 95% = -1.72 to -0.42; $p < 0.001$). Visual inspection

of the funnel plot and Egger's regression test (intercept = 1.17; p = 0.01) suggest publication bias. The trim-and-fill method was used to correct for this publication bias, with the result remaining significant (k = 26; effect size = -1.19; CI 95% = -1.86 to -0.52; $p < 0.01$).

In view of the impact of the thymic state on global reinforcement learning, a meta-analysis of only 17 studies was carried out on reward learning during the euthymic period (Adida et al., 2011; Brambilla et al., 2013; Caletti et al., 2013; Gomide Vasconcelos et al., 2014; Ibanez et al., 2012; Jogia et al., 2012; Malloy-Diniz et al., 2011; Martino et al., 2011; Ono et al., 2015; Yechiam et al., 2008; Linke et al., 2011; Strauss et al., 2015; Pouchon et al., 2023; Pouchon et al., in preparation). It found a significant impairment of reward learning during the euthymic phase compared to HC, with a large effect size (k = 17; effect size = -0.85; CI 95% = 1.59 to -0.11; $p < 0.05$) (**Figure V.V.5**). Heterogeneity between studies was significant and interpreted as highly significant ($I^2 = 97.85\%$; Q-test = $p < 0.01$). The Baujat plot indicates that one study primarily influenced the overall effect size and contributed to heterogeneity (Adida et al., 2011). After the suppression of this outlier study, the difference was no longer significant (k = 16; effect size = -0.66; CI 95% = -1.32 to -0.01; p = 0.54). Visual inspection of the funnel plot and Egger's regression test (intercept = 0.85; p = 0.13) do not suggest publication bias.

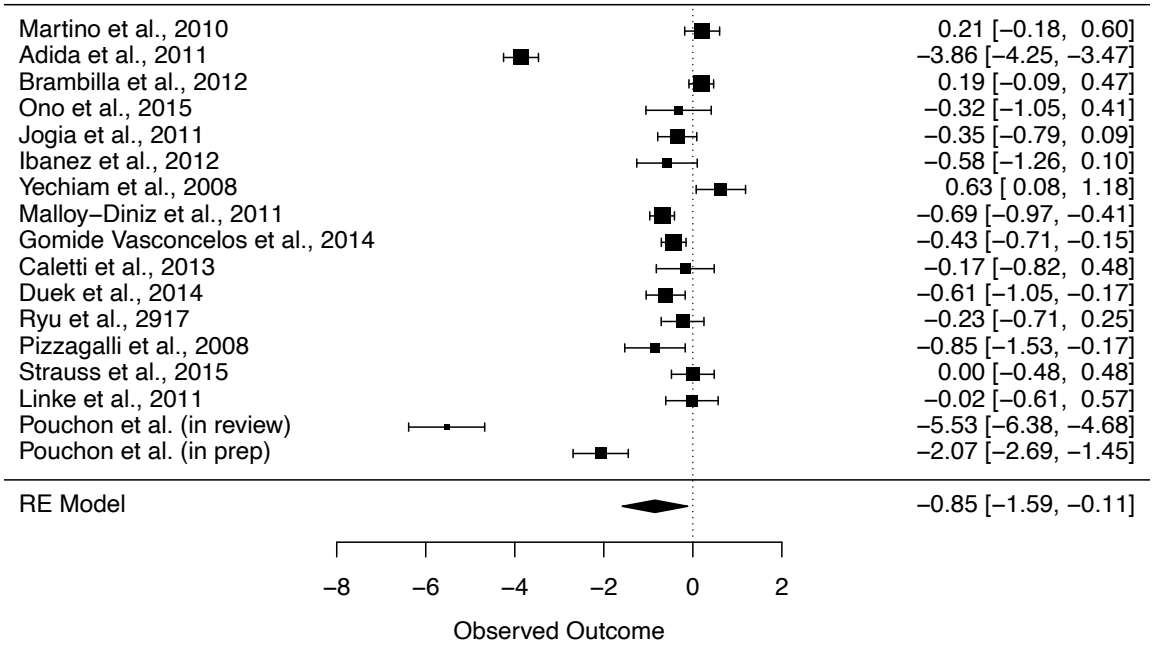

Figure V.V.5: Forest plot of meta-analysis only about reward learning performances during euthymic state among patients with bipolar disorder compared to heathy controls.

Concerning punishment learning, 9 studies were included in the meta-analysis (Duek et al., 2014; Abohamza et al., 2020; Linke et al., 2011; Strauss et al., 2015; Barch et al., 2017; Geana et al., 2022; Pratt et al., 2021; Pouchon et al., 2023; Pouchon et al., in preparation). There was significant impairment in punishment learning performances in patients compared with HC, with a large effect size (k = 9; effect size = -1.59; CI 95% = -2.60 to -0.59; $p < 0.01$). Heterogeneity between studies was significant and interpreted as very large ($I^2 = 97.11\%$; Q-test = $p < 0.01$). The Baujat plot indicates that 3 studies mainly influenced the overall effect size and contributed to heterogeneity (Duek et al., 2014; Geana et al., 2022; Linke et al., 2011). After the suppression of these studies, the magnitude of the overall effect size remains significant and large (k = 6; effect size = -1.10; CI 95% = -1.68 to -0.53; $p < 0.001$). Visual inspection of the funnel plot and Egger's regression test (intercept = 2.53; $p < 0.05$) suggest publication bias. The trim-and-fill method was used to correct for this publication bias, with the result remaining significant (k = 9; effect size = -1.59; CI 95% = -2.60 to -0.59; $p < 0.01$).

In view of the impact of the thymic state on global RL, a meta-analysis focusing solely on punishment learning conditions during the euthymic period was carried out in 5 studies (Duek et al., 2014; Linke et al., 2011; Strauss et al., 2015; Pouchon et al., 2023; Pouchon et al., in preparation) with 221 patients with BD and 146 HC. This time, no alteration was found in learning by punishment during the euthymic phase compared with healthy subjects (k = 5; effect size = -1.18; CI 95% = -2.63 to 0.28; p = 0.11) (**Figure V.V.6**). Heterogeneity between studies was significant and interpreted as very large (I^2 = 97.17%; Q-test = p < 0.01). The Baujat plot indicates that one study primarily influenced the overall effect size and contributed to heterogeneity (Linke et al., 2011). After the suppression of this outlier study, the difference remains insignificant (k = 4; effect size = -1.54; CI 95% = -3.18 to 0.10; p = 0.06). Visual inspection of the funnel plot and Egger's regression test (intercept = 4.87; p = 0.03) suggest publication bias. The trim-and-fill method was used to correct for this publication bias, with the result remaining insignificant (k = 5; effect size = -1.18; CI 95% = -2.63 to 0.28; p = 0.11).

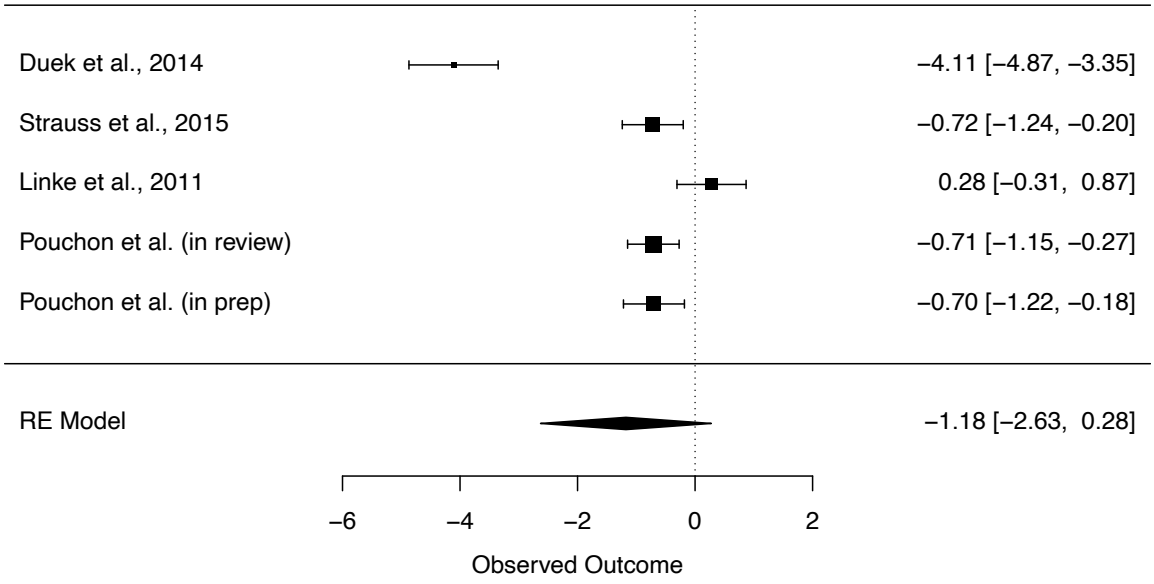

Figure V.V.6: Forest plot of meta-analysis only about punishment learning performances during euthymic state among patients with bipolar disorder compared to heathy controls.

4. Preliminary discussion

Our main findings highlight that the RL performance of patients with BD is impaired overall compared with HC. Despite the probable impact of the thymic episode on these results, there are too few studies to conclude concerning depression and mania. On the other hand, we highlight that this impairment persists during euthymia, with selective impairment of reward learning, but probably not of punishment learning.

Our results highlight an overall impairment of RL in BD compared to HC, except for 2 experimental tasks, with only 2 studies for one and 3 studies for the other. Meta-regressions showed an effect of IGT on our results. Taken separately, there seemed to be no impairment of RL (i.e. reward learning) for studies using this task, but a significant impairment for studies using other tasks, which seems to echo a previous meta-analysis on risky choices, rate of learning on IGT in patients with bipolar disorder type I in euthymic period, which showed no difference compared to healthy control subjects (Edge et al., 2013). The impact of IGT on these results probably stems from differences in task design, and therefore in the cognitive processes involved (Richards et al., 2013; Stocco et al., 2009). In the other tasks, the participant makes choices between symbols or shapes, with feedback and often varied contingency. He must gradually learn which of these symbols is considered correct and which is incorrect. In IGT, the participant chooses cards from 4 decks, 2 riskier and 2 less risky. For each card chosen, there is a 50% chance of having to pay a penalty. For decks A and B, the penalty is $250, while for decks C and D, it's $50. So, he needs to learn which packs offer the highest payout with the lowest penalty. As previously proposed, we emphasize that the type of paradigm influences the result and its interpretation and contributes to the disparity of results (Johnson et al., 2019; Richards et al., 2013; Stocco et al., 2009).

We also carried out meta-analyses according to thymic state. The main finding was an alteration in reward learning during the euthymic period, but not in punishment learning. These results echo a systematic review by the International Society for bipolar disorder on

reward processing and affective decision-making of 31 studies using mainly IGT and CGT. They concluded that a reduction in reward sensitivity in remitted patients among 15 studies, but no difference among 9 studies (Miskowiak et al., 2019). Their review takes into account many different tasks involving reward processing during euthymia, not necessarily in a learning context. Although sensitivity to reward during euthymia in bipolar disorder has not been quantified by meta-analysis, it is unclear to date whether there is hypersensitivity or hyposensitivity to reward, with some authors suggesting hypersensitivity to reward in BD BD (Alloy & Nusslock, 2019; Nusslock & Alloy, 2017; Wessa et al., 2014), others are more suggestive of hyposensitivity to reward (Dutra et al., 2015; Johnson et al., 2019; Schreiter et al., 2016; Trost et al., 2014; Yip et al., 2015). However, in view of our results, it may be that hyposensitivity to reward in BD leads to impaired performance in a learning context. It has been suggested that the type of experimental task used may be responsible for the different results, referring back to our discussion above about the impact of the task used on the results (Johnson et al., 2019).

With regard to learning by punishment, these results echo a study that found loss aversion in both a group of healthy subjects and a group of patients with BD, but with no difference between the two (Anderson et al., 2021). Caution should be exercised with these results, however, given the small number of studies taken into account and the heterogeneity of the tasks. In addition, our analyses show a high degree of heterogeneity in all our results, which needs to be considered when interpreting them. These results remain preliminary, however, and require further statistical analysis for their interpretation.

Whether or not valence is considered appears to be an important factor in the integration of results, and also contributes to the disparity of results. These results are consistent with the dissociation of reward and punishment processing in a learning context (Palminteri & Pessiglione, 2017).

5. References

Abohamza, E., Weickert, T., Ali, M., & Moustafa, A. A. (2020). Reward and punishment learning in schizophrenia and bipolar disorder. *Behavioural Brain Research*, *381*, 112298. https://doi.org/10.1016/j.bbr.2019.112298

Adida, M., Jollant, F., Clark, L., Besnier, N., Guillaume, S., Kaladjian, A., Mazzola-Pomietto, P., Jeanningros, R., Goodwin, G. M., Azorin, J.-M., & Courtet, P. (2011). Trait-Related Decision-Making Impairment in the Three Phases of Bipolar Disorder. *Biological Psychiatry*, *70*(4), 357-365. https://doi.org/10.1016/j.biopsych.2011.01.018

Alloy, L. B., & Nusslock, R. (2019). Future Directions for Understanding Adolescent Bipolar Spectrum Disorders : A Reward Hypersensitivity Perspective. *Journal of Clinical Child & Adolescent Psychology*, *48*(4), 669-683. https://doi.org/10.1080/15374416.2019.1567347

Anderson, Z., Fairley, K., Villanueva, C. M., Carter, R. M., & Gruber, J. (2021). No group differences in Traditional Economics Measures of loss aversion and framing effects in bipolar I disorder. *PLOS ONE*, *16*(11), e0258360. https://doi.org/10.1371/journal.pone.0258360

APA. (s. d.). *American Psychiatric Association. (2013). Cautionary statement for forensic use of DSM-5. In Diagnostic and statistical manual of mental disorders (5th ed.). Washington, DC: Author. Http://dx.doi.org/10.1176/appi.books.9780890425596 .CautionaryStatement. (N.d.).*

Barch, D. M., Carter, C. S., Gold, J. M., Johnson, S. L., Kring, A. M., MacDonald, A. W., Pizzagalli, D. A., Ragland, J. D., Silverstein, S. M., & Strauss, M. E. (2017). Explicit and implicit reinforcement learning across the psychosis spectrum. *Journal of Abnormal Psychology*, *126*(5), 694-711. https://doi.org/10.1037/abn0000259

Baujat, B., Mahé, C., Pignon, J.-P., & Hill, C. (2002). A graphical method for exploring heterogeneity in meta-analyses : Application to a meta-analysis of 65 trials: GRAPHICAL METHOD FOR EXPLORING HETEROGENEITY IN META-ANALYSES. *Statistics in Medicine*, *21*(18), 2641-2652. https://doi.org/10.1002/sim.1221

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, *50*(1-3), 7-15. https://doi.org/10.1016/0010-0277(94)90018-3

Blain, B., & Rutledge, R. B. (2020). Momentary subjective well-being depends on learning and not reward. *eLife*, *9*, e57977. https://doi.org/10.7554/eLife.57977

Brambilla, P., Perlini, C., Bellani, M., Tomelleri, L., Ferro, A., Cerruti, S., Marinelli, V., Rambaldelli, G., Christodoulou, T., Jogia, J., Dima, D., Tansella, M., Balestrieri, M., & Frangou, S. (2013). Increased salience of gains *versus* decreased associative learning differentiate bipolar disorder from schizophrenia during incentive decision making. *Psychological Medicine*, *43*(3), 571-580. https://doi.org/10.1017/S0033291712001304

Caletti, E., Paoli, R. A., Fiorentini, A., Cigliobianco, M., Zugno, E., Serati, M., Orsenigo, G., Grillo, P., Zago, S., Caldiroli, A., Prunas, C., Giusti, F., Consonni, D., & Altamura, A. C. (2013). Neuropsychology, social cognition and global functioning among bipolar, schizophrenic patients and healthy controls : Preliminary data. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00661

Cecchi, R., Vinckier, F., Hammer, J., Marusic, P., Nica, A., Rheims, S., Trebuchon, A., Barbeau, E. J., Denuelle, M., Maillard, L., Minotti, L., Kahane, P., Pessiglione, M., & Bastin, J. (2022). Intracerebral mechanisms explaining the impact of incidental feedback on mood state and risky choice. *eLife*, *11*, e72440. https://doi.org/10.7554/eLife.72440

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences. Hillsdale, NJ: Law-rence Erlbaum Associates. Inc, Publishers.*

Deeks, J. J., Higgins, J. P., Altman, D. G., & on behalf of the Cochrane Statistical Methods Group. (2019). Analysing data and undertaking meta-analyses. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Éds.), *Cochrane Handbook for Systematic Reviews of Interventions* (1re éd., p. 241-284). Wiley. https://doi.org/10.1002/9781119536604.ch10

Del Re, A. C. (2010). *Compute.es : Compute Effect Sizes. R package version 0.2. Http://CRAN.R-project.org/package=compute.es*.

Drevon, D., Fursa, S. R., & Malcolm, A. L. (2017). Intercoder Reliability and Validity of WebPlotDigitizer in Extracting Graphed Data. *Behavior Modification*, *41*(2), 323-339. https://doi.org/10.1177/0145445516673998

Duek, O., Osher, Y., Belmaker, R. H., Bersudsky, Y., & Kofman, O. (2014). Reward sensitivity and anger in euthymic bipolar disorder. *Psychiatry Research*, *215*(1), 95-100. https://doi.org/10.1016/j.psychres.2013.10.028

Dutra, S. J., Cunningham, W. A., Kober, H., & Gruber, J. (2015). Elevated striatal reactivity across monetary and social rewards in bipolar I disorder. *Journal of Abnormal Psychology*, *124*(4), 890-904. https://doi.org/10.1037/abn0000092

Edge, M. D., Johnson, S. L., Ng, T., & Carver, C. S. (2013). Iowa gambling task performance in euthymic bipolar I disorder : A meta-analysis and empirical study. *Journal of Affective Disorders*, *150*(1), 115-122. https://doi.org/10.1016/j.jad.2012.11.027

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*(7109), 629-634. https://doi.org/10.1136/bmj.315.7109.629

Eldar, E., Rutledge, R. B., Dolan, R. J., & Niv, Y. (2016). Mood as Representation of Momentum. *Trends in Cognitive Sciences*, *20*(1), 15-24. https://doi.org/10.1016/j.tics.2015.07.010

Geana, A., Barch, D. M., Gold, J. M., Carter, C. S., MacDonald, A. W., Ragland, J. D., Silverstein, S. M., & Frank, M. J. (2022). Using Computational Modeling to Capture Schizophrenia-Specific Reinforcement Learning Differences and Their Implications on Patient Classification. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *7*(10), 1035-1046. https://doi.org/10.1016/j.bpsc.2021.03.017

Gomide Vasconcelos, A., Sergeant, J., Corrêa, H., Mattos, P., & Malloy-Diniz, L. (2014). When self-report diverges from performance : The usage of BIS-11 along with neuropsychological tests. *Psychiatry Research*, *218*(1-2), 236-243. https://doi.org/10.1016/j.psychres.2014.03.002

Gu, Y., Zhou, C., Yang, J., Zhang, Q., Zhu, G., Sun, L., Ge, M., & Wang, Y. (2020). A transdiagnostic comparison of affective decision-making in patients with schizophrenia, major depressive disorder, or bipolar disorder. *PsyCh Journal*, *9*(2), 199-209. https://doi.org/10.1002/pchj.351

Gueguen, M. C. M., Lopez-Persem, A., Billeke, P., Lachaux, J.-P., Rheims, S., Kahane, P., Minotti, L., David, O., Pessiglione, M., & Bastin, J. (2021). Anatomical dissociation of intracerebral signals for reward and punishment prediction errors in humans. *Nature Communications*, *12*(1), 3344. https://doi.org/10.1038/s41467-021-23704-w

Ho, N., & Sommers, M. (2013). Anhedonia : A Concept Analysis. *Archives of Psychiatric Nursing*, *27*(3), 121-129. https://doi.org/10.1016/j.apnu.2013.02.001

Ibanez, A., Cetkovich, M., Petroni, A., Urquina, H., Baez, S., Gonzalez-Gadea, M. L., Kamienkowski, J. E., Torralva, T., Torrente, F., Strejilevich, S., Teitelbaum, J., Hurtado, E., Guex, R., Melloni, M., Lischinsky, A., Sigman, M., & Manes, F. (2012). The Neural Basis of Decision-Making and Reward Processing in Adults with Euthymic Bipolar Disorder or Attention-Deficit/Hyperactivity Disorder (ADHD). *PLoS ONE*, *7*(5), e37306. https://doi.org/10.1371/journal.pone.0037306

Jogia, J., Dima, D., Kumari, V., & Frangou, S. (2012). Frontopolar cortical inefficiency may underpin reward and working memory dysfunction in bipolar disorder. *The World Journal of Biological Psychiatry*, *13*(8), 605-615. https://doi.org/10.3109/15622975.2011.585662

Johnson, S. L., Mehta, H., Ketter, T. A., Gotlib, I. H., & Knutson, B. (2019). Neural responses to monetary incentives in bipolar disorder. *NeuroImage: Clinical*, *24*, 102018. https://doi.org/10.1016/j.nicl.2019.102018

Leucht, S., Fennema, H., Engel, R. R., Kaspers-Janssen, M., & Szegedi, A. (2018). Translating the HAM-D into the MADRS and vice versa with equipercentile linking. *Journal of Affective Disorders*, *226*, 326-331. https://doi.org/10.1016/j.jad.2017.09.042

Lewandowski, K. E., Whitton, A. E., Pizzagalli, D. A., Norris, L. A., Ongur, D., & Hall, M.-H. (2016). Reward Learning, Neurocognition, Social Cognition, and Symptomatology in Psychosis. *Frontiers in Psychiatry*, *7*. https://doi.org/10.3389/fpsyt.2016.00100

Linke, J., King, A. V., Rietschel, M., Strohmaier, J., Hennerici, M., Gass, A., Meyer-Lindenberg, A., & Wessa, M. (2012). Increased Medial Orbitofrontal and Amygdala Activation : Evidence for a Systems-Level Endophenotype of Bipolar I Disorder. *American Journal of Psychiatry*, *169*(3), 316-325. https://doi.org/10.1176/appi.ajp.2011.11050711

Linke, J., Sönnekes, C., & Wessa, M. (2011). Sensitivity to positive and negative feedback in euthymic patients with bipolar I disorder : The last episode makes the difference. *Bipolar Disorders*, *13*(7-8), 638-650. https://doi.org/10.1111/j.1399-5618.2011.00956.x

Ma, L.-L., Wang, Y.-Y., Yang, Z.-H., Huang, D., Weng, H., & Zeng, X.-T. (2020). Methodological quality (risk of bias) assessment tools for primary and secondary medical studies : What are they and which is better? *Military Medical Research*, *7*(1), 7. https://doi.org/10.1186/s40779-020-00238-8

Maia, T. V., & Frank, M. J. (2011). From reinforcement learning models to psychiatric and neurological disorders. *Nature Neuroscience*, *14*(2), 154-162. https://doi.org/10.1038/nn.2723

Malloy-Diniz, L. F., Neves, F. S., De Moraes, P. H. P., De Marco, L. A., Romano-Silva, M. A., Krebs, M.-O., & Corrêa, H. (2011). The 5-HTTLPR polymorphism, impulsivity and suicide behavior in euthymic bipolar patients. *Journal of Affective Disorders*, *133*(1-2), 221-226. https://doi.org/10.1016/j.jad.2011.03.051

Martino, D. J., Strejilevich, S. A., Torralva, T., & Manes, F. (2011). Decision making in euthymic bipolar I and bipolar II disorders. *Psychological Medicine*, *41*(6), 1319-1327. https://doi.org/10.1017/S0033291710001832

Mason, L., Eldar, E., & Rutledge, R. B. (2017). Mood Instability and Reward Dysregulation-A Neurocomputational Model of Bipolar Disorder. *JAMA Psychiatry*, *74*(12), 1275-1276. https://doi.org/10.1001/jamapsychiatry.2017.3163

Miskowiak, K. W., Seeberg, I., Kjaerstad, H. L., Burdick, K. E., Martinez-Aran, A., Bonnin, C., Bowie, C. R., Carvalho, A. F., Gallagher, P., Hasler, G., Lafer, B., López-Jaramillo, C., Sumiyoshi, T., McIntyre, R. S., Schaffer, A., Porter, R. J., Purdon, S., Torres, I. J., Yatham, L. N., … Vieta, E. (2019). Affective cognition in bipolar disorder : A systematic review by the ISBD targeting cognition task force. *Bipolar Disorders*, *21*(8), 686-719. https://doi.org/10.1111/bdi.12834

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses : The PRISMA Statement. *PLoS Medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Nusslock, R., & Alloy, L. B. (2017). Reward processing and mood-related symptoms : An RDoC and translational neuroscience perspective. *Journal of Affective Disorders*, *216*, 3-16. https://doi.org/10.1016/j.jad.2017.02.001

Ono, Y., Kikuchi, M., Hirosawa, T., Hino, S., Nagasawa, T., Hashimoto, T., Munesue, T., & Minabe, Y. (2015). Reduced prefrontal activation during performance of the Iowa Gambling Task in patients with bipolar disorder. *Psychiatry Research: Neuroimaging*, *233*(1), 1-8. https://doi.org/10.1016/j.pscychresns.2015.04.003

Palminteri, S., Justo, D., Jauffret, C., Pavlicek, B., Dauta, A., Delmaire, C., Czernecki, V., Karachi, C., Capelle, L., Durr, A., & Pessiglione, M. (2012). Critical Roles for Anterior Insula and Dorsal Striatum in Punishment-Based Avoidance Learning. *Neuron*, *76*(5), 998-1009. https://doi.org/10.1016/j.neuron.2012.10.017

Palminteri, S., Khamassi, M., Joffily, M., & Coricelli, G. (2015). Contextual modulation of value signals in reward and punishment learning. *Nature Communications*, *6*(1), 8096. https://doi.org/10.1038/ncomms9096

Palminteri, S., & Pessiglione, M. (2017). Opponent Brain Systems for Reward and Punishment Learning. In *Decision Neuroscience* (p. 291-303). Elsevier. https://doi.org/10.1016/B978-0-12-805308-9.00023-3

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., & Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*, *442*(7106), 1042-1045. https://doi.org/10.1038/nature05051

Pizzagalli, D. A., Goetz, E., Ostacher, M., Iosifescu, D. V., & Perlis, R. H. (2008). Euthymic Patients with Bipolar Disorder Show Decreased Reward Learning in a Probabilistic Reward Task. *Biological Psychiatry*, *64*(2), 162-168. https://doi.org/10.1016/j.biopsych.2007.12.001

Pouchon, A., Vinckier, F., Dondé, C., Gueguen, M. C., Polosan, M., & Bastin, J. (2023). Reward and punishment learning deficits among bipolar disorder subtypes. *Journal of Affective Disorders*, *340*, 694-702. https://doi.org/10.1016/j.jad.2023.08.075

Powers, R. L., Russo, M., Mahon, K., Brand, J., Braga, R. J., Malhotra, A. K., & Burdick, K. E. (2013). Impulsivity in bipolar disorder : Relationships with neurocognitive dysfunction and substance use history. *Bipolar Disorders*, *15*(8), 876-884. https://doi.org/10.1111/bdi.12124

Pratt, D. N., Barch, D. M., Carter, C. S., Gold, J. M., Ragland, J. D., Silverstein, S. M., & MacDonald, A. W. (2021). Reliability and Replicability of Implicit and Explicit Reinforcement Learning Paradigms in People With Psychotic Disorders. *Schizophrenia Bulletin*, *47*(3), 731-739. https://doi.org/10.1093/schbul/sbaa165

Richards, J. M., Plate, R. C., & Ernst, M. (2013). A systematic review of fMRI reward paradigms used in studies of adolescents vs. Adults : The impact of task design and implications for understanding neurodevelopment. *Neuroscience & Biobehavioral Reviews*, *37*(5), 976-991.

RStudio Team. (2020). *RStudio : Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/*. (2023.6.1.524) [Logiciel].

Rutledge, R. B., Skandali, N., Dayan, P., & Dolan, R. J. (2014). A computational and neural model of momentary subjective well-being. *Proceedings of the National Academy of Sciences*, *111*(33), 12252-12257. https://doi.org/10.1073/pnas.1407535111

Ryu, V., Ha, R. Y., Lee, S. J., Ha, K., & Cho, H.-S. (2017). Behavioral and Electrophysiological Alterations for Reinforcement Learning in Manic and Euthymic Patients with Bipolar Disorder. *CNS Neuroscience & Therapeutics*, *23*(3), 248-256. https://doi.org/10.1111/cns.12671

Schreiter, S., Spengler, S., Willert, A., Mohnke, S., Herold, D., Erk, S., Romanczuk-Seiferth, N., Quinlivan, E., Hindi-Attar, C., Banzhaf, C., Wackerhagen, C., Romund, L., Garbusow, M., Stamm, T., Heinz, A., Walter, H., & Bermpohl, F. (2016). Neural alterations of fronto-striatal circuitry during reward anticipation in euthymic bipolar disorder. *Psychological Medicine*, *46*(15), 3187-3198. https://doi.org/10.1017/S0033291716001963

Stocco, A., Fum, D., & Napoli, A. (2009). Dissociable processes underlying decisions in the Iowa Gambling Task : A new integrative framework. *Behavioral and Brain Functions*, *5*(1), 1. https://doi.org/10.1186/1744-9081-5-1

Strauss, G. P., Thaler, N. S., Matveeva, T. M., Vogel, S. J., Sutton, G. P., Lee, B. G., & Allen, D. N. (2015). Predicting psychosis across diagnostic boundaries : Behavioral and computational modeling evidence for impaired reinforcement learning in schizophrenia and bipolar disorder with a history of psychosis. *Journal of Abnormal Psychology*, *124*(3), 697-708. https://doi.org/10.1037/abn0000039

Trost, S., Diekhof, E. K., Zvonik, K., Lewandowski, M., Usher, J., Keil, M., Zilles, D., Falkai, P., Dechent, P., & Gruber, O. (2014). Disturbed Anterior Prefrontal Control of the

Mesolimbic Reward System and Increased Impulsivity in Bipolar Disorder. *Neuropsychopharmacology*, *39*(8), 1914-1923. https://doi.org/10.1038/npp.2014.39

Van Enkhuizen, J., Henry, B. L., Minassian, A., Perry, W., Milienne-Petiot, M., Higa, K. K., Geyer, M. A., & Young, J. W. (2014). Reduced Dopamine Transporter Functioning Induces High-Reward Risk-Preference Consistent with Bipolar Disorder. *Neuropsychopharmacology*, *39*(13), 3112-3122. https://doi.org/10.1038/npp.2014.170

Viechtbauer, W. (2010). Conducting Meta-Analyses in *R* with the **metafor** Package. *Journal of Statistical Software*, *36*(3). https://doi.org/10.18637/jss.v036.i03

Vinckier, F., Rigoux, L., Oudiette, D., & Pessiglione, M. (2018). Neuro-computational account of how mood fluctuations arise and affect decision making. *Nature Communications*, *9*(1), 1708. https://doi.org/10.1038/s41467-018-03774-z

Wells G, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al. (s. d.). *The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Https://www.ohri.ca/programs/clinical_epidemiology/oxford.asp*.

Wessa, M., Kanske, P., & Linke, J. (2014). Bipolar disorder : A neural network perspective on a disorder of emotion and motivation. *Restorative Neurology and Neuroscience*, *32*(1), 51-62. https://doi.org/10.3233/RNN-139007

Whitton, A. E., Lewandowski, K. E., & Hall, M.-H. (2021). Smoking as a Common Modulator of Sensory Gating and Reward Learning in Individuals with Psychotic Disorders. *Brain Sciences*, *11*(12), 1581. https://doi.org/10.3390/brainsci11121581

Whitton, A. E., Treadway, M. T., & Pizzagalli, D. A. (2015). Reward processing dysfunction in major depression, bipolar disorder and schizophrenia. *Current Opinion in Psychiatry*, *28*(1), 7-12. https://doi.org/10.1097/YCO.0000000000000122

Yechiam, E., Hayden, E. P., Bodkins, M., O'Donnell, B. F., & Hetrick, W. P. (2008). Decision making in bipolar disorder : A cognitive modeling approach. *Psychiatry Research*, *161*(2), 142-152. https://doi.org/10.1016/j.psychres.2007.07.001

Yip, S. W., Worhunsky, P. D., Rogers, R. D., & Goodwin, G. M. (2015). Hypoactivation of the Ventral and Dorsal Striatum During Reward and Loss Anticipation in Antipsychotic and Mood Stabilizer-Naive Bipolar Disorder. *Neuropsychopharmacology*, *40*(3), 658-666. https://doi.org/10.1038/npp.2014.215

6. Supplementary information

Table V.2 : Study Quality Assessment as Indexed by Newcastle-Ottawa Scale.

Study	Selection				Comparability	Exposure			Total
	Definition of the case	Representativeness of the cases	Selection of Controls	Definition of Controls	Comparability of cases and controls on the basis of the design or analysis	Ascertainment of exposure	Same method of ascertainment for cases and controls	Non-Response rate	
Adida et al., 2011	1	1	0	1	2	1	1	1	8
Brambilla et al., 2013	1	1	0	1	2	1	1	1	8
Caletti et al., 2013	1	1	0	0	1	1	1	1	6

Gomide Vasconcelos et al., 2014	1	1	0	1	2	1	1	1	8
Gu et al., 2020	1	1	0	1	2	1	1	1	8
Ibanez et al., 2012	1	1	0	1	2	1	1	1	8
Jogia et al., 2012	1	1	0	1	2	1	1	1	8
Malloy-Diniz et al., 2011	1	1	0	0	2	1	1	1	7
Martino et al., 2011	1	1	0	1	2	1	1	1	8
Ono et al., 2015	1	1	0	1	2	1	1	1	8
Powers et al., 2013	1	1	0	1	1	1	1	1	7
Van enkhuizen et al., 2014	1	1	0	1	2	1	1	1	8
Yechiam et al., 2008	1	1	0	1	2	1	1	1	8
Lewandowski et al., 2016	1	1	0	1	2	1	1	1	8
Pizzagalli et al., 2008	1	1	0	1	1	1	1	1	7
Ryu et al., 29017	1	1	0	1	2	1	1	1	8
Whitton et al., 2021	1	1	0	0	1	1	1	0	5
Duek et al., 2014	1	1	0	1	2	1	1	1	8
Abohamza et al., 2019	1	1	0	1	1	1	1	1	7
Barch et al., 2017	1	1	0	1	2	1	1	1	8
Geana et al., 2021	1	1	0	1	2	1	1	0	7
Pratt et al., 2021	1	1	0	1	2	1	1	0	7
Linke et al., 2011	1	1	0	1	2	1	1	1	8
Strauss et al., 2015	1	1	0	1	1	1	1	1	7
Pouchon et al. 2023	1	1	0	1	2	1	1	1	8
Pouchon et al. (in prep)	1	1	0	1	2	1	1	1	8

VI

Reward and punishment learning deficits among bipolar disorder subtypes

Arnaud Pouchon[a,b], Fabien Vinckier[c,d,e], Clément Dondé[a,b,f], Maëlle CM Gueguen[g,h], Mircea Polosan[a,b,1], Julien Bastin[i,1],

a Univ. Grenoble Alpes, Inserm, U1216, CHU Grenoble Alpes, Grenoble Institut Neurosciences, 38000 Grenoble, France

b Department of Psychiatry, CHU Grenoble Alpes, 38000 Grenoble, France

c Motivation, Brain & Behavior (MBB) lab, Institut du Cerveau (ICM), Hôpital Pitié-Salpêtrière, F-75013 Paris, France

d Université Paris Cité, F-75006 Paris, France

e Department of Psychiatry, Service Hospitalo-Universitaire, GHU Paris Psychiatrie & Neurosciences, F-75014 Paris, France

f Department of Psychiatry, CH Alpes-Isère, 38000 Saint-Egrève, France

g Department of Psychiatry, University Behavioral Health Care & the Brain Health Institute, Rutgers University—New Brunswick, Piscataway, USA

h Laureate Institute for Brain Research, Tulsa, OK 74136 USA

i Univ. Grenoble Alpes, Inserm, U1216, Grenoble Institut Neurosciences, 38000 Grenoble, France

1 These authors contributed equally to this work.

Published in "Journal of Affective Disorder"

Journal of Affective Disorders 340 (2023) 694–702

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research paper

Reward and punishment learning deficits among bipolar disorder subtypes

Arnaud Pouchon [a,b,*], Fabien Vinckier [c,d,e], Clément Dondé [a,b,f], Maëlle CM Gueguen [g,h], Mircea Polosan [a,b,1], Julien Bastin [i,1,**]

[a] Univ. Grenoble Alpes, Inserm, U1216, CHU Grenoble Alpes, Grenoble Institut Neurosciences, 38000 Grenoble, France
[b] Department of Psychiatry, CHU Grenoble Alpes, 38000 Grenoble, France
[c] Motivation, Brain & Behavior (MBB) lab, Institut du Cerveau (ICM), Hôpital Pitié-Salpêtrière, F-75013 Paris, France
[d] Université Paris Cité, F-75006 Paris, France
[e] Department of Psychiatry, Service Hospitalo-Universitaire, GHU Paris Psychiatrie & Neurosciences, F-75014 Paris, France
[f] Department of Psychiatry, CH Alpes-Isère, 38000 Saint-Egrève, France
[g] Department of Psychiatry, University Behavioral Health Care & the Brain Health Institute, Rutgers University—New Brunswick, Piscataway, USA
[h] Laureate Institute for Brain Research, Tulsa, OK 74136 USA
[i] Univ. Grenoble Alpes, Inserm, U1216, Grenoble Institut Neurosciences, 38000 Grenoble, France

ARTICLE INFO

Keywords:
Bipolar disorder$_1$
reinforcement$_2$
learning$_3$
reward$_4$
punishment$_5$
Computational biology$_6$

ABSTRACT

Background: Reward sensitivity is an essential dimension related to mood fluctuations in bipolar disorder (BD), but there is currently a debate around hypersensitivity or hyposensitivity hypotheses to reward in BD during remission, probably related to a heterogeneous population within the BD spectrum and a lack of reward bias evaluation. Here, we examine reward maximization vs. punishment avoidance learning within the BD spectrum during remission.

Methods: Patients with BD-I (n = 45), BD-II (n = 34) and matched (n = 30) healthy controls (HC) were included. They performed an instrumental learning task designed to dissociate reward-based from punishment-based reinforcement learning. Computational modeling was used to identify the mechanisms underlying reinforcement learning performances.

Results: Behavioral results showed a significant reward learning deficit across BD subtypes compared to HC, captured at the computational level by a lower sensitivity to rewards compared to punishments in both BD subtypes. Computational modeling also revealed a higher choice randomness in BD-II compared to BD-I that reflected a tendency of BD-I to perform better during punishment avoidance learning than BD-II.

Limitations: Our patients were not naive to antipsychotic treatment and were not euthymic (but in syndromic remission) according to the International Society for Bipolar Disorder definition.

Conclusions: Our results are consistent with the reward hyposensitivity theory in BD. Computational modeling suggests distinct underlying mechanisms that produce similar observable behaviors, making it a useful tool for distinguishing how symptoms interact in BD versus other disorders. In the long run, a better understanding of these processes could contribute to better prevention and management of BD.

1. Introduction

Bipolar disorder (BD) is a chronic mood disorder characterized by alternating episodes of depression and (hypo)mania, of which the hedonic dimension is a key symptom for clinical diagnosis according to the Diagnostic and Statistical Manual of Mental Disorders Fifth Edition (DSM-5) criteria A for these two episodes (APA, 2013). Depression is characterized by a reduction of hedonic dimension (anhedonia), whereas (hypo)mania is characterized by hyperhedonia. There are two BD subtypes according to the DSM-5: bipolar I disorder (BD-I) and bipolar II disorder (BD-II), with notably the absence of mania in favor of hypomania (and thus less hedonic dimension impairment) in BD-II.

* Correspondence to: A. Pouchon, Department of Psychiatry, CHU Grenoble Alpes, Allée de la Source, 38700 La Tronche, F-38000 Grenoble, France.
** Correspondence to: J. Bastin, Grenoble Institut Neurosciences, Site Santé, Bâtiment Edmond J. Safra, 31 Chem. Fortuné Ferrini, 38700 La Tronche, F-38000 Grenoble, France.
E-mail addresses: apouchon@chu-grenoble.fr (A. Pouchon), julien.bastin@univ-grenoble-alpes.fr (J. Bastin).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.jad.2023.08.075
Received 24 January 2023; Received in revised form 24 July 2023; Accepted 14 August 2023
Available online 15 August 2023

Within the dimensional orientation of the Research Domain Criteria (RDoC) (Insel et al., 2010), BD subtypes may be conceptualized as a spectrum in which reward sensitivity could be a key dimension (Nusslock and Alloy, 2017), as also supported by differential brain activity observed in the reward circuit between BD-I and BD-II (Caseras et al., 2013).

More generally, there is currently a debate regarding reward sensitivity in BD. During thymic periods, some studies found a reduced sensitivity to reward during bipolar depression (Redlich et al., 2015; Satterthwaite et al., 2015) while others found the opposite result (Chase et al., 2013). In (hypo)mania, results also exhibit some heterogeneity, ranging from a hypersensitivity to reward (Alloy et al., 2008; Bermpohl et al., 2009; Lozano and Johnson, 2001; Scott et al., 2000), to no alteration (Hägele et al., 2015) or a reduced reward sensitivity (Abler et al., 2008). During the remission period, some suggest hypersensitivity to reward (Alloy and Nusslock, 2019; Caseras et al., 2013; Linke et al., 2012; Wessa et al., 2014) whereas others suggest hyposensitivity (Schreiter et al., 2016; Trost et al., 2014; Yip et al., 2015). Interestingly, a systematic review found that, across studies, the most reliable finding was a reduction of sensitivity to reward (Miskowiak et al., 2019). Studying cognitive alterations during the remission period is essential for identifying specific trait markers of the disease (Frey et al., 2013) and could also help improve diagnosis, develop personalized treatment for patients and better prevent relapses (Guglielmo et al., 2021; McGorry, 2013; Teixeira et al., 2016).

Moreover, given the bidirectional relationship existing between mood fluctuations and reinforcement learning (Blain and Rutledge, 2020; Eldar et al., 2016; Rutledge et al., 2014), several studies focused on the identification of reinforcement learning (RL) deficits in BD during the remission period. Yet, these studies showed disparate results regarding RL performances during remission (Adida et al., 2011; Brambilla et al., 2013; Duek et al., 2014; Lewandowski et al., 2016; Linke et al., 2012, 2011; Pizzagalli et al., 2008; Ryu et al., 2017; Yechiam et al., 2008). These discrepancies may be partly explained by the polarity of the last affective episode experienced before the remission: BD-I patients who had experienced a manic episode learned better from positive feedback, whereas BD-I patients who had experienced a depressive episode learned better from negative feedback (Linke et al., 2011). Another issue regarding reinforcement learning deficits in BD is that a majority of studies only examined reward-based learning but not punishment avoidance learning, such that the specificity of the altered cognitive processes in BD during learning remains unclear. A clear distinction between reward vs. punishment learning is also important for the clinics, since BD-I and BD-II are distinguished by the intensity of the (hypo)manic episode, which are largely dopamine-dependent, whereas this is less true of bipolar depression (Ashok et al., 2017). Another reason to (separately) examine reward-based and punishment-based learning is that rewards and punishments are also known to impact subjects mood in opposite ways (Blain and Rutledge, 2020; Cecchi et al., 2021; Vinckier et al., 2018) such that an asymmetry between reward and punishment learning processes may play a role in the genesis of different thymic states and their maintenance (Eldar et al., 2016).

Here, we used a probabilistic learning task designed to dissociate reward learning from punishment avoidance learning (Gueguen et al., 2021; Pessiglione et al., 2006) to compare reinforcement learning performance across three groups of participants (euthymic BD-I and BD-II and a group of healthy subjects). We combined this behavioral analysis with computational modeling of RL, with the aim of investigating the possible underlying cognitive mechanisms involved, including learning rate, sensitivity to reward and punishment, and choice randomness. We also explored the impact of several moderators on learning performance, such as antipsychotic medications and the polarity of the last thymic episode before the remission period.

2. Methods

2.1. Participants

Remitted BD-I and BD-II patients were recruited at the Grenoble-Alpes University Hospital. Inclusion criteria were a BD-I or BD-II diagnosis defined by DSM-5 for patients, and remitted state defined by a Montgomery Asberg Depression Rating Scale (MADRS) score lower than 15 and Young Mania Rating Scale (YMRS) score lower than 12 since at least one month. Exclusions criteria were other psychiatric diagnoses than BD, history of a neurological disorder or a systemic illness with neurological complications, lack of fluency in French, recent history (6 months) of substance abuse or dependence except for tobacco, brain stimulation procedure, or head trauma with unconsciousness. The same criteria were applied for healthy controls, in addition to the presence of recent or past psychiatric or neurological disorders according to the DSM-5.

Based on previous studies with robust results, we included a larger number of subjects, namely eighty-eight (88) BD participants were recruited, including 49 patients with BD-I, 39 patients with BD-II, and 32 HC. 2 BD-I and 2 BD-II patients were excluded from the final analyses as they were experiencing an active mood episode on the day of the assessment. Two (2) BD-I patients, 3 BD-II patients, and 2 HC were excluded because they did not complete the task. The final analyzed sample included 79 BD patients (45 BD-I and 34 BD-II) and 30 HC. A clear oral and written explanation was also delivered to all participants. This study was approved by the French Institutional Ethic Committee (CPP Île de France V) and was compliant with international standards for tests involving human participants (Helsinki Declaration of 1975, as revised in 2008). The experiment was registered at ClinicalTrials.gov (NCT04237610).

2.2. Clinical assessment

We collected sociodemographic variables including sex (assigned at birth), age and education level in order to match patients and HC. For patients, we collected clinical data including age of BD onset, time in remission since the last relapse, polarity of first and last episode, number of past depressive, manic or hypomanic episodes, number of suicide attempts, number of hospitalizations, and medication. We converted antipsychotic medication into its chlorpromazine equivalent, employing a conversion of psychotropic treatments to a drug load (Davis and Chen, 2004; Hafeman et al., 2012; Phillips et al., 2008).

We assessed anhedonia because of its relationship to reward sensitivity. All participants completed the Snaith–Hamilton pleasure scale (SHAPS) (Snaith et al., 1995), a validated 14-item self-rating scale that explores hedonic responses in several situations related to leisure pursuit and interests, eating and drinking, social interactions and sensory experiences. A total score $\geq$3 indicates a significant reduction in hedonic capacity (anhedonia). We also assessed impulsivity, an important trait in BD which could interfere with reward sensitivity (Trost et al., 2014). For instance, preferential activity for high probability rewards correlates negatively with impulsivity in the dorsolateral prefrontal cortex and VS in euthymic BD patients (Mason et al., 2014). We used the UPPS Impulsive Behavior Scale (Whiteside and Lynam, 2001), a 59-item self-report questionnaire designed to measure five facets of impulsive behaviors: positive urgency, negative urgency, lack of perseverance, lack of premeditation and sensation seeking. Each statement is rated on a 5-point scale ranging from strongly agree (1) to strongly disagree (4).

2.3. Behavioral task

Participants performed a computer-based probabilistic instrumental learning task adapted from previous studies (Palminteri et al., 2012;

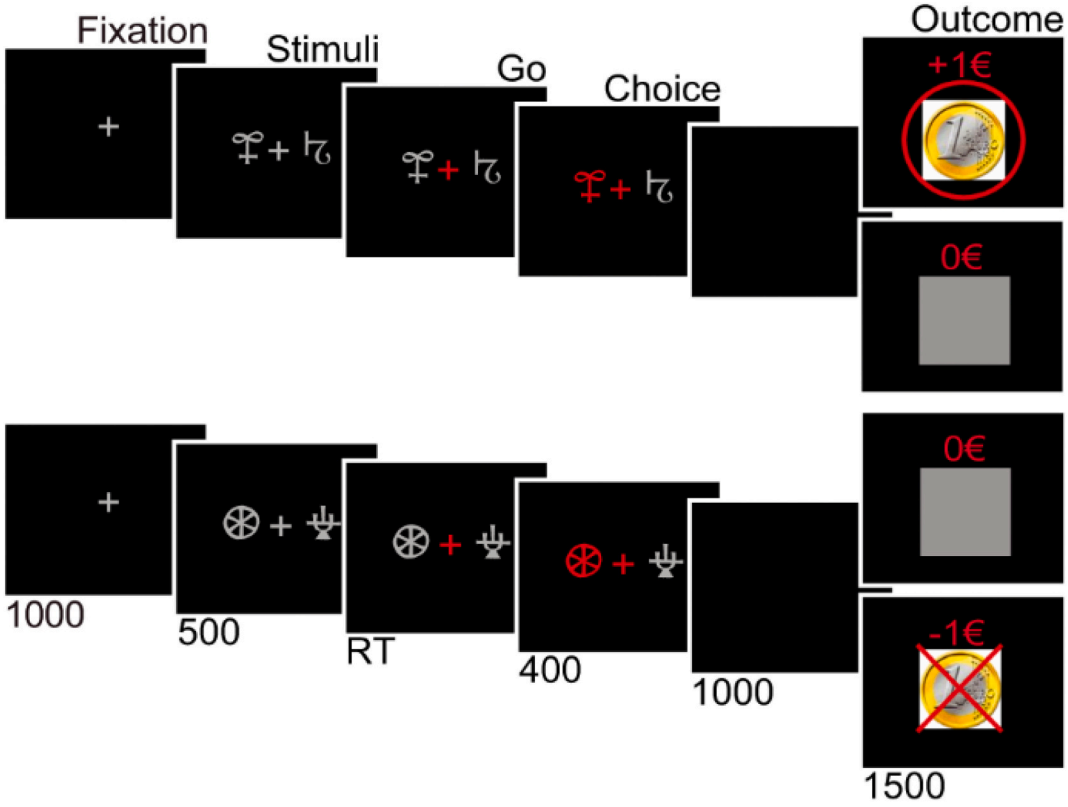

Fig. 1. Behavioral task. Successive screen of typical trials in the reward (top) and punishment (bottom) learning conditions. Participants had to select one abstract visual stimulus among the two presented on each side of a central visual fixation cross, and subsequently observed the outcome. Duration is given in milliseconds.

Pessiglione et al., 2006) (Fig. 1). Participants were provided with written instructions, which were reformulated orally if necessary. The aim of the task was to maximize the financial payoff and, to do so, participants had to consider reward seeking and punishment avoidance as equally important. Participants had initial short training sessions to familiarize themselves with task timing and responses. Each session was an independent task presenting 4 new pairs of cues to be learned. Cues were abstract visual stimuli taken from the Agathodaimon alphabet. Each pair of cues was presented 24 times for a total of 96 trials. The four cue pairs correspond to the two conditions (2 pairs of gain cues representing rewards and 2 pairs of loss cues representing punishments), which are respectively associated with different pairs of outcomes (winning 1€ versus nothing or losing 1€ versus nothing). Within each pair, the two cues are associated with the two possible outcomes with reciprocal probabilities (0.75/0.25 and 0.25/0.75). In each trial, a pair was randomly presented and the two cues were displayed on the screen on the left and right of a central fixation cross, their relative position being counterbalanced across trials. Participants were asked to choose the left stimulus or the right stimulus by using their left or right index fingers to press the corresponding button on a keyboard. Since the position on the screen was counterbalanced, response (left versus right) and value (good versus bad cue) were orthogonal. The chosen cue was colored in red for 250 ms and then the outcome (either "nothing", "gain", or "loss") was displayed on the screen. In order to win money, participants had to learn by trial and error the cue–outcome associations, so as to choose the most rewarding cue in the gain condition and the less punishing cue in the loss condition. Participants did 3 test sessions after the training. The training procedure included a very short session including two pairs of cues presented during 16 trials. This was followed by 2 to 3 short 5-minutes sessions during which the experimenter checked that participants correctly understood the task during both the reward and punishment conditions.

2.4. Statistical analysis

Statistics were performed using RStudio (RStudio Team (2016). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/) and MATLAB Statistical Toolbox (MATLAB R2017a, The MathWorks, Inc., USA) at a significance level $\alpha < 0.05$.

Model fit, as well as follow-up Bayesian analyses were implemented using the VBA toolbox (http://mbb-team.github.io/VBA-toolbox/) (Daunizeau et al., 2014; Rigoux et al., 2014). All models were inverted using a variational Bayes approach under the Laplace approximation. This algorithm not only inverts nonlinear models but also estimates their evidence, which represents a trade-off between accuracy (goodness of fit) and complexity (degrees of freedom). The significance of fitted parameters (means of posterior distributions) was tested using ANOVAs and Bonferroni post-hoc tests.

Demographic features for the three groups were analyzed for age and education level using analysis of variance (ANOVA). Chi-square tests were computed for gender/nominal clinical data. Normality was examined using the Shapiro–Wilk test. Then, we compared clinical features between BD-I and BD-II groups for residual mood symptoms and history of disease using Student's *t*-test.

For our first outcome, we used a two-way ANOVA with valence (reward and punishment) and group (BD-I, BD-II, HC) as study factors, after verification of ANOVA conditions, i.e. normality of distribution and homogeneity of variances. We performed post-hoc analyses on the various interactions, after Bonferroni correction. We also computed the reward bias (reward-learning performance subtracted from punishment-learning), which is the tendency for a participant to learn better by seeking rewards rather than avoiding punishment.

Regression models were used to evaluate if potential continuous confounders influenced average learning performance during the reward or the punishment learning (or the learning asymmetry between condition), including antipsychotic medication (GLM1–3, Chlorpro

mazine equivalent in mg), polarity of the last thymic episode (GLM4: depression vs. manic/hypomanic episodes) or residual mood symptoms (GLM5: MADRS and YMRS).

In a first general linear model (GLM1), average learning performance during the reward learning condition was regressed against:

- Groups (BD patients were coded as −1, pooling BD-I and BD-II patients together for this analysis; healthy subjects were coded as 1)
- And four confounding variables: medication, M (coded as Chlorpromazine equivalent in mg), polarity of the last thymic episode: depression was coded as −1, manic or hypomanic polarity of episodes were coded as 1) and MADRS and YMRS scores.

$$\text{Reward learning} = \beta_0 + \beta_{group} * GROUP + \beta_{med} * M + \beta_{last-episode} * last - episode + \beta_{MADRS} * MADRS + \beta_{YMRS} * YMRS + \epsilon \quad (1)$$

with $\beta 0$ corresponding to the intercept, β_{group} corresponding to the regression estimates of interest, β_{med}, $\beta_{last-episode}$, β_{MADRS} and β_{YMRS} corresponding to the continuous confounding factors and ε corresponding to the error term.

In a second GLM (GLM2), average learning performance during the punishment learning condition was regressed against BD groups (BD-I patients were coded as −1 and BD-II patients were coded as 1) and the four confounding variables.

$$\text{Punishment learning} = \beta_0 + \beta_{group} * GROUP + \beta_{med} * M + \beta_{last-episode} * last - episode + \beta_{MADRS} * MADRS + \beta_{YMRS} * YMRS + \epsilon \quad (2)$$

with $\beta 0$ corresponding to the intercept, β_{group} corresponding to the regression estimates of interest, β_{med}, $\beta_{last-episode}$, β_{MADRS} and β_{YMRS} corresponding to the continuous confounding factors and ε corresponding to the error term.

In a third GLM (GLM3), the reward bias was regressed against BD-I and HC (BD-I patients were coded as −1 and HC were coded as 1) and the four confounding variables.

$$\text{Reward bias} == \beta_0 + \beta_{group} * GROUP + \beta_{med} * M + \beta_{last-episode} * last - episode + \beta_{MADRS} * MADRS + \beta_{YMRS} * YMRS + \epsilon \quad (3)$$

with $\beta 0$ corresponding to the intercept, β_{group} corresponding to the regression estimates of interest, β_{med}, $\beta_{last-episode}$, β_{MADRS} and β_{YMRS} corresponding to the continuous confounding factors and ε corresponding to the error term.

2.5. Model fitting and model simulations

To dissociate learning abilities from differential sensitivity to reward and punishment, we used a standard RL model. For each pair of cues, the model predicts the probability of choosing cues A and B (P_A and P_B) for a given trial (t), through a softmax function of their expected value (Q_a and Q_b):

$$P_A(t) = \frac{e^{Qa(t)/\beta}}{e^{Qa(t)/\beta} + e^{Qb(t)/\beta}} \quad (5)$$

where β is the temperature, a positive free parameter capturing choice stochasticity.

The expected values were set at 0 before learning and, after each outcome, the value of the chosen stimulus (say A) was updated in proportion to the prediction error, according to the following "delta rule":

$$Q_{chosen}(t+1) = Q_{chosen}(t) + \alpha.(Out(t) - Q_{chosen}(t)) \quad (6)$$

where α is the learning rate, a free parameter between 0 and 1. The variables across trials t are choice probability (P), expected value (Q) and outcome (Out). Depending on models (see below). Out could be encoded as 1/−1 or R/−1, where R is a free parameter capturing the asymmetry between reward and punishment.

We compared several variants of a RL algorithm according to three factors: 1) updating or not of the unchosen cue value based on the counterfactual outcome, 2) one single or two separate learning rates for reward and punishment pairs of cues, 3) same or different weighting or reward and punishment outcomes (i.e. R fixed to 1 or not). The full factorial model space makes a total of 8 variants, but we removed the combination of separate learning rates with separate weighting of reward and punishment, to preserve parameter identifiability, resulting in 6 different models. Bayesian model comparison was used to compare models. In the best model, the free parameters were the asymmetry between reward and punishment R, the temperature β (weight on expected value difference), and the learning rate α (weight on prediction error).

3. Results

3.1. Demographics

Regarding demographic features, we found no difference between the groups (BD-I, BD-II and HC groups). Furthermore, BD-I and BD-II patients did not significantly differ in residual mood symptoms (MADRS and YMRS), impulsivity (UPPS) and anhedonia (SHAPS). Regarding history of disease, we found no significant difference between BD-I and BD-II regarding the last episode type. Yet, BD-II patients had a higher number of past depressive ($t_{(77)} = -4.49$; $p < 0.001$) and hypomanic episodes than BD-I ($t_{(77)} = -4.53$; $p < 0.001$). BD-I patients experienced more psychotic symptoms during thymic episodes than BD-II ($Chi^2 = 20.58$; $p < 0.001$). No differences were observed between BD-I and BD-II for other variables (Table 1).

3.2. Behavioral performances

We compared the average percentage of correct choices across valence and groups (Fig. 2) using a two-way analysis of variance (mixed ANOVA with valence -reward and punishment-, and group -BD-I, BD-II, HC- as study factors). We found a significant group effect ($F_{(2,106)} = 10.93$; $p < 0.0001$), and an interaction between group and valence ($F_{(2,106)} = 7.77$; $p = 0.0007$). To further specify the source of the group by valence interaction, we used Bonferroni multiple comparison tests.

As can be seen in Fig. 2A, learning performance was lower in patients with BD during the reward-learning condition compared to HC (BD-I vs. HC: $p < 0.001$; BD-II vs. HC: $p < 0.001$). There was no significant difference between BD-I and BD-II learning performance during the reward condition ($p > 0.99$).

The difference between BD patients and HC during reward learning was neither explained by antipsychotic medication, subsyndromal depressive symptoms or by the polarity of the last episodes. This was demonstrated by using a general linear model (see methods, GLM1). The regression estimates were significant regarding the group effect (BD patients vs. HC) on reward-learning ($\beta_{group} = 0.071$; $p = 1.35 \times 10^{-5}$) and for the MADRS ($\beta_{MADRS} = -0.009$; $p = 0.02$) but there was no significant effect of medication ($\beta_{med} = 1.6 \times 10^{-7}$; $p = 0.99$), polarity of the last episode ($\beta_{last\text{-}episode} = -3.8 \times 10^{-5}$; $p = 0.53$) and YMRS ($\beta_{YMRS} = 0.004$; $p = 0.63$). Thus, despite the fact that some of the patients had subsyndromal depressive symptoms that significantly altered reward-learning performance, the difference between BD patients and HC remained significant even when taking into account three important confounds.

Going back to the specification of the group by valence interaction on reinforcement learning performance, we next investigated more

Table 1
Socio-demographic and clinical characteristics for healthy control (HC), bipolar I disorder (BD-I) and bipolar II disorder (BD-II) patients.

	BD-I (n = 45)	BD-II (n = 34)	HC (n = 30)	Statistics	p-Value
Demographics					
Sex (female/male)	24/21	22/12	20/10	$X^2_{(2)} = 1.57$	0.47
Age, years, mean (SD)	46.36 (10.92)	46.42 (9.48)	44.35 (10.69)	$F_{(2,106)} = 1.18$	0.31
Years of education post-BAC, mean (SD)	3.92 (2.19)	3.18 (2.00)	4.24 (2.78)	$F_{(2,106)} = 1.13$	0.33
Clinical characteristics, mean (SD)					
YMRS	1.13 (1.50)	0.90 (1.67)		$t_{(77)} = 0.40$	0.69
MADRS	2.24 (3.35)	5.16 (4.04)		$t_{(77)} = -1.60$	0.11
UPPS	114.66 (11.85)	115.14 (11.57)		$t_{(76)} = -0.05$	0.96
SHAPS	0.84 (0.94)	1.35 (1.46)		$t_{(76)} = -1.35$	0.18
History of disease					
Last depressive episode, n (%)	24 (53.33)	24 (70.59)		$X^2_{(1)} = 1.75$	0.19
Last manic episode, n (%)	13 (28.89)				
Last hypomanic episode, n (%)	8 (17.78)	10 (29.41)		$X^2_{(1)} = 0.90$	0.34
Past depressive episodes, mean (SD)	4.57 (2.90)	9.46 (4.84)		$t_{(77)} = -4.49$	**<0.001**
Past manic episodes, mean (SD)	4.09 (3.05)				
Past hypomanic episodes, mean (SD)	2.05 (2.25)	6.49 (3.78)		$t_{(77)} = -4.53$	**<0.001**
Psychotic bipolar disorder, n (%)	26 (57.78)	2 (5.88)		$X^2_{(1)} = 20.58$	**<0.001**
Antipsychotic, n (%)	20 (44.44)	12 (35.29)		$X^2_{(1)} = 0.35$	0.56
Medication load in CPZ equivalent, mean (SD)	0.82 (0.86)	0.59 (0.76)		$t_{(77)} = 1.27$	0.21
Attempted suicide (lifetime), n (%)	15 (33.33)	10 (29.41)		$X^2_{(1)} = 0.02$	0.9
No. of suicide attempts (lifetime), mean (SD)	0.4 (0.54)	0.58 (1.04)		$t_{(77)} = -1.06$	0.29
Age of onset, years, mean (SD)	23.51 (7.43)	25.44 (7.74)		$t_{(77)} = -0.90$	0.37
No. of previous hospitalizations, mean (SD)	3.60 (2.36)	3 (3.09)		$t_{(77)} = 0.82$	0.41
Time in remission since the last relapse (months), mean (SD)	11.83 (11.31)	11.27 (7.89)		$t_{(77)} = -0.12$	0.91

YMRS = Young Mania Rating Scale; MADRS = Montgomery Asberg Depression Rating Scale; UPPS = UPPS Impulsive Behavior Scale; SHAPS = Snaith–Hamilton pleasure scale; CPZ = chlorpromazine.
Bold characters represent a significance level of the statistical test.

specifically punishment learning using post-hoc Bonferroni tests (Fig. 2B). There was no significant differences between the three groups (HC vs. BD-I: $p > 0.99$; HC vs. BD-II: $p = 0.6339$, BD-I vs. BD-II: $p = 0.23$; still, we noticed that BD-I tended to learn better from punishments than BD-II when using a less conservative statistical post-hoc test: BD-I vs. BD-II: $t_{(77)} = 2.5$; $p < 0.05$).

The difference between BD-I and BD-II patients during punishment learning was neither explained by antipsychotic medication, subsyndromal depressive symptoms nor by the polarity of the last episodes (GLM 2). Hence, the difference between BD-I and BD-II patients during punishment learning remained significant ($\beta_{BDIvsII} = 0.021$; $p = 0.03$) while none of the four confounds were significant ($\beta_{med} = -1.07 \times 10^{-5}$; $p = 0.77$; $\beta_{last\text{-}episode} = -0.0075$; $p = 0.45$; $\beta_{MADRS} = -0.002$; $p = 0.33$; $\beta_{YMRS} = 0.004$; $p = 0.41$).

Finally, to assess whether possible behavioral performance asymmetries between reward and punishment-based learning could explain the identified group by valence interaction on learning performance, we investigated with post-hoc comparisons the difference between the average percentage of correct choices in the reward and punishment conditions (reward bias, Fig. 2C). Post-hoc comparisons revealed that HC were better at learning from reward than from punishments ($p = 0.0103$) whereas patients with BD-I were better at learning from punishments than from rewards ($p = 0.0439$). Reward bias was also significantly higher in HC than BD-I ($p = 0.0005$) and BD-II ($p = 0.0223$) while there was no significant difference between BD-I and BD-II ($p = 0.9233$). We checked that these differences of reward bias between HC and patients with BD were not driven by antipsychotic medication, subsyndromal depressive symptoms or by the polarity of the last episodes (GLM 3: $\beta_{HCIvsBD} = 0.052$; $p = 0.02$; $\beta_{med} = -2.79 \times 10^{-5}$; $p = 0.70$; $\beta_{last\text{-}episode} = 0.0019$; $p = 0.92$; $\beta_{MADRS} = -0.005$; $p = 0.26$; $\beta_{YMRS} = -0.002$; $p = 0.86$).

3.3. Computational results

To further specify and quantify the possible mechanisms through which learning performances differed in BD patients, we used a model-based approach. We compared several variants of a Q-learning model. The most plausible model included updating of the unchosen cue (as if participants inferred the counterfactual outcome) and a sensitivity to reward that was different from sensitivity to punishment (as if subjectively perceived outcomes differed from the objectively symmetrical monetary gain and loss of +1€ and − 1€; xp = 1). Thus, there were 3 free parameters: the asymmetry between reward and punishment R, the temperature β (weight on expected value difference), and the learning rate α (weight on prediction error).

We then tested for group effect using an ANOVA on each of these parameters (Fig. 3). We found a significant group effect for the asymmetry between gains and losses (R; $F_{(2,106)} = 10.2$; $p = 9.10^{-5}$). Post hoc comparisons showed that the R was lower (and actually closer to 1) in both patient groups compared to HC (BD-I vs. HC: $p < 0.01$; BD-II vs. HC: $p < 0.01$). We also found a significant effect of group on temperature ($F_{(2,106)} = 4.52$; $p = 0.01$), with a more pronounced choice stochasticity found in BD-II patients than in the two other groups (both $p < 0.01$). There was no difference between groups regarding the learning rate ($p > 0.05$). In summary, the computational analysis indicated that the observed reward-based learning deficit in both BD subtypes compared to HC was captured by a lower reward/punishment asymmetry (and thus a lower sensitivity to reward compared to punishments) in both BD groups, while choice stochasticity was increased specifically in the BD-II group, probably accounting for the tendency towards a lower performance of BD-I in the punishment condition compared to BD-II.

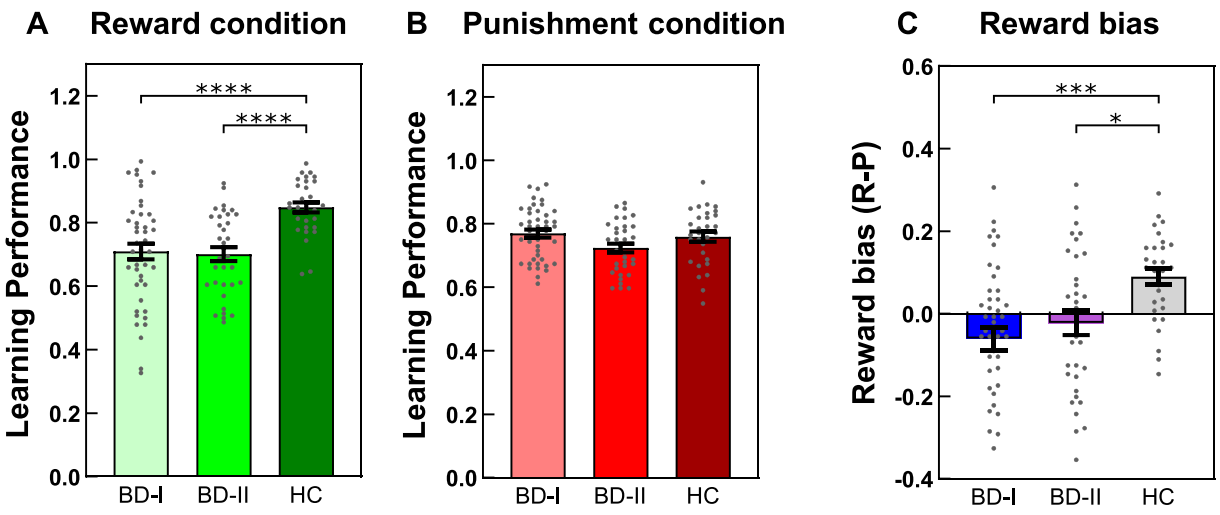

Fig. 2. Behavioral results. Learning performance during the task for reward condition (A, green) and punishment condition (B, red) across groups. C. Reward/punishment bias (i.e. punishment learning subtracted from reward learning performance). A positive reward bias indicates better learning by reward than by punishment (HC group), and a negative result indicates better learning by punishment than by reward (BD-I group). Dots represent data points from individual participants. BD-I = patients with bipolar I disorder; BD-II = patients with bipolar II disorder; HC = healthy controls. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

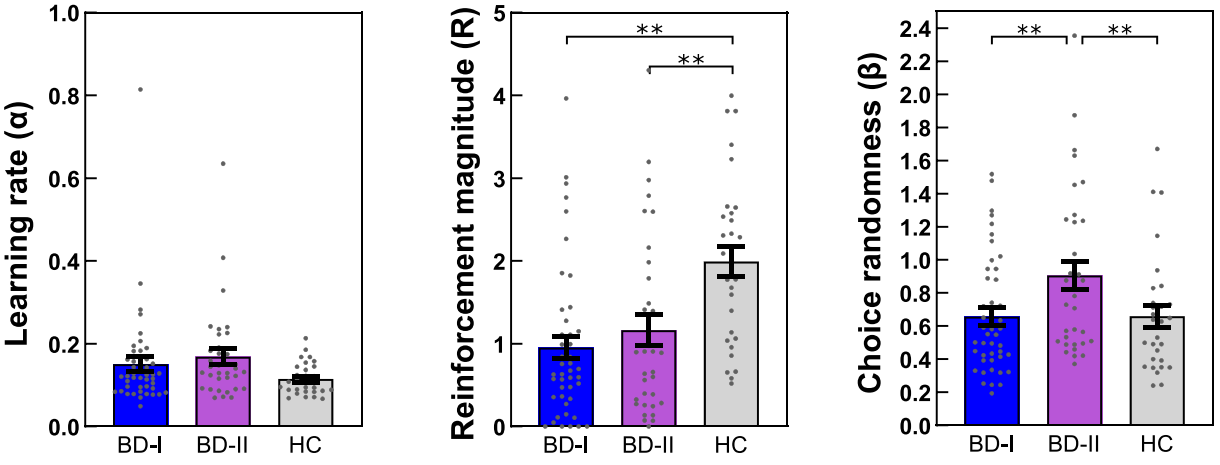

Fig. 3. Computational analysis. Adjusted free parameters (α: learning rate; *R*: reinforcement magnitude; ß: choice randomness) are illustrated for the different groups (BD-I, BD-II, HC). BD-I = bipolar I disorder; BD-II = bipolar II disorder; HC = healthy controls.

4. Discussion

In the present study, we used an instrumental learning task combined with computational modeling to explore reward and punishment learning alterations in BD. We found that 1) BD patients performed worse than HC during reward-based learning, which was captured at the computational level by reduced sensitivity to rewards in BD patients, and that 2) learning performance of BD-II patients in the punishment avoidance condition tended to be reduced compared to BD-I patients. This difference was captured at the computational level by a heightened level of choice stochasticity in patients with BD-II. In what follows, we first examine the impact of BD on reward-based learning, then on punishment-based learning.

4.1. BD and reward-based reinforcement learning

Our main result is that BD patients performed worse than HC during reward-based learning. Computational modeling suggests that this learning deficit was captured by a reduced sensitivity to reward magnitude compared to punishment magnitude in BD compared to HC. These results are more consistent with the hypodopaminergic hypothesis of patients with BD in remission for the following reasons.

First, since pharmacological manipulation in healthy subjects has shown that dopamine selectively affects reward learning during a similar task than the one used in this study (Pessiglione et al., 2006), a lower performance during reward learning might be explained by a blunted dopaminergic system activity in BD. A second argument is that computational modeling revealed that BD patients had a lower R parameter (reflecting an asymmetric sensitivity between rewards and punishments). Thus, the reduced sensitivity to reward in BD identified at the computational level could also indirectly reflect hypodopaminergic activity in the brain system involved during reward-related processes. Third, our results are also in line with a recent meta-analysis of fMRI studies on reward sensitivity in BD (Long et al., 2022). This meta-analysis showed that across eight neuroimaging studies from euthymic BD patients, the ventral striatum exhibited reduced activity during reward-related tasks in BD compared to HC. Finally, our behavioral observation is also consistent with two previous behavioral studies of reward learning in BD in remission (Pizzagalli et al., 2008; Ryu et al., 2017).

Our results also highlight a negative correlation between the intensity of residual depressive symptoms represented by MADRS scores and reward learning. Although this negative correlation did not account for the learning performance difference between patients with BD and HC, it is consistent with a previous report found in BD in remission (Abohamza et al., 2020) suggesting a negative correlation between reward learning performance and the Hamilton Depression Scale.

Clinically, our results suggest that BD patients learn more from their failures than from their successes in remission than healthy subjects do. It could be that this hyposensitivity to rewards (in comparison to punishments) leads to behaviors that are more oriented towards punishment avoidance than reward seeking, and therefore contributes to the perpetuation of certain residual symptoms of the motivational sphere in remission (Caruso et al., 2020). Our results imply that it may be advantageous to concentrate the cognitive behavioral therapy strategy on this reward sensitivity deficit when motivational symptoms remain in individuals who are in remission.

4.2. BD and punishment-based reinforcement learning

We did not find any difference regarding punishment learning in BD compared to HC. Yet, at the computational level, we found that BD-II exhibited a higher choice variability compared to BD-I. This increased choice variability was probably explained by a tendency towards better punishment learning in BD-I than in BD-II (uncorrected $p < 0.05$).

Few studies have addressed punishment learning separately from reward learning in BD during remission. Interestingly, we replicate the results of Duek et al.'s (2014) study in which they found a similar asymmetry between reward and punishment learning. In our study, we further show that patients with BD-II exhibited a lower punishment learning performance than BD-I which was explained at the computational level by an increased choice randomness parameter in BD-II. Our findings also echo another study that tested reward and punishment learning during a probabilistic learning task by comparing psychotic (with history of psychotic symptoms) and non-psychotic BD with schizophrenia and HC (Abohamza et al., 2020; Duek et al., 2014; Linke et al., 2012, 2011). They found that non-psychotic BD performed significantly worse on punishment learning task than psychotic BD and HC (but not between the psychotic subtype and HC). The authors interpreted their results in relation to dopamine since the non-psychotic patients showed less dopamine reduction. Yet, this interpretation is unlikely to explain our results since, as stated above, the task employed in our study was shown to be affected by the dopaminergic state selectively during reward learning, but not during punishment learning (Pessiglione et al., 2006). Since in our study sample the percentage of psychotic bipolar patients was significantly lower in BD-II (<10 %) than in BD-I (>55 %), this suggests that punishment learning deficits in BD would specifically concern non-psychotic/BD-II patients.

At the computational level, we found that the choice randomness parameter, coding for the degree to which choices were reward-maximizing (punishment-minimizing) versus random between available options was increased in BD-II compared to both BD-I and HC. It remains difficult to derive strong cognitive conclusions from such a parameter, because it could also simply capture proportion of trials that are not well explained by the model due to patients' lapses of attention for example (Geana et al., 2021) or, alternatively, could correspond to an increased internal neural variability (Wilson et al., 2014) or to sub-optimal brain inference mechanisms (Gershman, 2019). It would therefore be interesting in the future to carry out targeted studies on the noise parameter in BD in remission, using a dedicated paradigm on this parameter combined with computational modeling, with the aim of better characterizing this deficit within bipolar spectrum disorders. Longitudinal fMRI studies could also help to better characterize patients' evolutionary trajectories (Macoveanu et al., 2022).

4.3. Strengths and limitations

One of the strengths of this study is that we were able to recruit a large number of participants (n = 45 BD-I; n = 34 BD-II) compared to previous studies such as Linke et al., 2011 (n = 23) (Linke et al., 2011), Trost et al., 2014 (n = 16) (Trost et al., 2014), Yip et al., 2015 (n=20) (Yip et al., 2015), Schreiter et al., 2016 (n = 20) (Schreiter et al., 2016), Pizzagalli et al., 2008 (n = 18) (Pizzagalli et al., 2008), Ryu et al., 2017 (n = 20) (Ryu et al., 2017). That said, our study also has several limitations.

First, we found a significant learning bias towards reward in HC, indicating that healthy subjects were better at maximizing rewards than at avoiding punishments. This learning asymmetry probably reflects the extra cognitive step that is necessary during punishment avoidance learning since subjects have to first identify the cue associated with a high probability of punishment and then select the other cue to avoid punishments whereas during the reward learning condition subjects simply have to identify and select the most rewarding cue. While this type of learning asymmetry was not always detected (Palminteri et al., 2012; Pessiglione et al., 2006), it is not uncommon and has been previously reported by different groups (Cheng et al., 2022; Kobza et al., 2012; Xu et al., 2021).

Secondly, it is difficult to exclude antipsychotic medication as a possible confounding factor. It is indeed difficult, and even unethical, to recruit untreated patients (i.e., risking destabilizing the BD by modifying the usual medical treatment). It is also possible to recruit patients at the beginning of their disease and before any significant clinical intervention to mitigate this bias (Yip et al., 2015), but this would exclude very severe patients and thus may induce a potential underestimation of the magnitude of the observed effects. To address this potential bias, a conversion of psychotropic treatments to a drug load is often used (Hafeman et al., 2012; Phillips et al., 2008). This is what we did to account for possible effects of the antipsychotic treatments. Because our sample size was large enough to do so, we also replicated our analyses after excluding patients on antipsychotic treatment (Caseras et al., 2013). This is consistent with the little evidence for the effects of medication on probabilistic RL in BD (Abohamza et al., 2020). Another limitation regarding the potential impact of antipsychotics on our findings is that in this study we could not differentiate partial agonist (only n = 7 patients were treated with Aripiprazole) from the anti-dopaminergic activity of other antipsychotics.

Another limitation of this study may be the presence of sub-syndromic depressive symptoms in our patients, although we recruited patients in intercritical phase. However, these symptoms are the rule rather than the exception in bipolar populations in remission (Bourne et al., 2013) and our patients met the criteria of the ISBD for the definition of symptomatic remission (Tohen et al., 2009).

Finally, we failed to replicate the effect of the last episode on reward and punishment learning performance. According to Linke's study (Linke et al., 2011), we predicted that patients with a "last depressive episode" would show increased learning performance in the punishment condition, whereas patient with a "last (hypo)manic episode" would show increased learning performance in the reward condition. This replication failure could be due to the difference in patient recruitment (only BD-I patients were recruited in the study of Linke et al. (2011)). Critically, the paradigm employed was different between studies: the effect of the last thymic episode on the RL process per-se was not tested in the study of Linke et al. since the learning phase in their paradigm corresponded to the training phase, while all analyses were performed using binary choices between stimuli whose values had been previously learned during training. Thus, we hypothesize that the last episode has an impact on the value-comparison process associated with binary choices in a context where the value of each cue has already been learned but must be compared (during conflicting choices such as those tested in Linke's study). In contrast, there appeared to be no effect of last episode during the learning phase, when subjects were required to learn Q values of each cue to maximize monetary gains and to minimize monetary losses.

5. Conclusion

We found impaired performances in reward learning that were explained at the computational level by blunted reward sensitivity. Conversely, during punishment avoidance, we found a difference between BD-I and BD-II patients which was explained by a higher choice randomness computational parameter in BD-II. Thus, such a computational psychiatry approach to BD could be of interest in clinical practice, by leading to a more precise and personalized cognitive behavior therapy (Nair et al., 2020) focused on patient-specific bias.

Funding

This work benefited from the program from Université Grenoble Alpes, within the program 'Investissements d'Avenir' (ANR-17-CE37-0018; ANR-18-CE28-0016; ANR-22-CE17-0057).

CRediT authorship contribution statement

- Design and conduct of the study: JB, MP, AP
- Collection, management, analysis, and interpretation of the data:
- AP, JB, MP, FV
- Preparation, review, or approval of the manuscript: AP, JB, MP, FV, CD, MG
- Decision to submit the manuscript for publication: AP, JB, MP, FV, CD, MG.

Declaration of competing interest

FV has been invited to scientific meetings, consulted and/or served as speaker and received compensation by Lundbeck, Servier, Recordati, Janssen, Otsuka, LivaNova, and Chiesi. He has received research support by Lundbeck and LivaNova. None of these links of interest are related to this work. AP, CD, MG, MP, JB declare that no competing interests exist.

Acknowledgments

The authors gratefully acknowledge the participants. The authors also acknowledge the assistance of A. Bertrand, C. Gauld, T. George, L. Djennaoui, C. Lasserre, L. Bony, C. Baratin and M. Baumgaertner for their technical support.

References

Abler, B., Greenhouse, I., Ongur, D., Walter, H., Heckers, S., 2008. Abnormal reward system activation in mania. Neuropsychopharmacol 33, 2217–2227. https://doi.org/10.1038/sj.npp.1301620.

Abohamza, E., Weickert, T., Ali, M., Moustafa, A.A., 2020. Reward and punishment learning in schizophrenia and bipolar disorder. Behav. Brain Res. 381, 112298 https://doi.org/10.1016/j.bbr.2019.112298.

Adida, M., Jollant, F., Clark, L., Besnier, N., Guillaume, S., Kaladjian, A., Mazzola-Pomietto, P., Jeanningros, R., Goodwin, G.M., Azorin, J.-M., Courtet, P., 2011. Trait-related decision-making impairment in the three phases of bipolar disorder. Biol. Psychiatry 70, 357–365. https://doi.org/10.1016/j.biopsych.2011.01.018.

Alloy, L.B., Nusslock, R., 2019. Future directions for understanding adolescent bipolar spectrum disorders: a reward hypersensitivity perspective. J. Clin. Child Adolesc. Psychol. 48, 669–683. https://doi.org/10.1080/15374416.2019.1567347.

Alloy, L.B., Abramson, L.Y., Walshaw, P.D., Cogswell, A., Grandin, L.D., Hughes, M.E., Iacoviello, B.M., Whitehouse, W.G., Urosevic, S., Nusslock, R., Hogan, M.E., 2008. Behavioral Approach System and Behavioral Inhibition System sensitivities and bipolar spectrum disorders: prospective prediction of bipolar mood episodes. Bipolar Disord. 10, 310–322. https://doi.org/10.1111/j.1399-5618.2007.00547.x.

APA, 2013. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5). American Psychiatric Press, Inc., Arlington, VA (n.d.).

Ashok, A.H., Marques, T.R., Jauhar, S., Nour, M.M., Goodwin, G.M., Young, A.H., Howes, O.D., 2017. The dopamine hypothesis of bipolar affective disorder: the state of the art and implications for treatment. Mol. Psychiatry 22, 666–679. https://doi.org/10.1038/mp.2017.16.

Bermpohl, F., Kahnt, T., Dalanay, U., Hägele, C., Sajonz, B., Wegner, T., Stoy, M., Adli, M., Krüger, S., Wrase, J., Ströhle, A., Bauer, M., Heinz, A., 2009. Altered representation of expected value in the orbitofrontal cortex in mania. Hum. Brain Mapp. 31, 958–969. https://doi.org/10.1002/hbm.20909.

Blain, B., Rutledge, R.B., 2020. Momentary subjective well-being depends on learning and not reward. eLife 9, e57977. https://doi.org/10.7554/eLife.57977.

Bourne, C., Aydemir, Ö., Balanzá-Martínez, V., Bora, E., Brissos, S., Cavanagh, J.T.O., Clark, L., Cubukcuoglu, Z., Dias, V.V., Dittmann, S., Ferrier, I.N., Fleck, D.E., Frangou, S., Gallagher, P., Jones, L., Kieseppä, T., Martínez-Aran, A., Melle, I., Moore, P.B., Mur, M., Pfennig, A., Raust, A., Senturk, V., Simonsen, C., Smith, D.J., Bio, D.S., Soeiro-de-Souza, M.G., Stoddart, S.D.R., Sundet, K., Szöke, A., Thompson, J.M., Torrent, C., Zalla, T., Craddock, N., Andreassen, O.A., Leboyer, M., Vieta, E., Bauer, M., Worhunsky, P.D., Tzagarakis, C., Rogers, R.D., Geddes, J.R., Goodwin, G.M., 2013. Neuropsychological testing of cognitive impairment in euthymic bipolar disorder: an individual patient data meta-analysis. Acta Psychiatr. Scand. 128, 149–162. https://doi.org/10.1111/acps.12133.

Brambilla, P., Perlini, C., Bellani, M., Tomelleri, L., Ferro, A., Cerruti, S., Marinelli, V., Rambaldelli, G., Christodoulou, T., Jogia, J., Dima, D., Tansella, M., Balestrieri, M., Frangou, S., 2013. Increased salience of gains *versus* decreased associative learning differentiate bipolar disorder from schizophrenia during incentive decision making. Psychol. Med. 43, 571–580. https://doi.org/10.1017/S0033291712001304.

Caruso, D., Meyrel, M., Krane-Gartiser, K., Benard, V., Benizri, C., Brochard, H., Geoffroy, P.-A., Gross, G., Maruani, J., Prunas, C., Yeim, S., Palagini, L., Dell'Osso, L., Leboyer, M., Bellivier, F., Etain, B., 2020. Eveningness and poor sleep quality contribute to depressive residual symptoms and behavioral inhibition in patients with bipolar disorder. Chronobiol. Int. 37, 101–110. https://doi.org/10.1080/07420528.2019.1685533.

Caseras, X., Lawrence, N.S., Murphy, K., Wise, R.G., Phillips, M.L., 2013. Ventral striatum activity in response to reward: differences between bipolar I and II disorders. AJP 170, 533–541. https://doi.org/10.1176/appi.ajp.2012.12020169.

Cecchi, R., Vinckier, F., Hammer, J., Marusic, P., Nica, A., Rheims, S., Trebuchon, A., Barbeau, E., Denuelle, M., Maillard, L., Minotti, L., Kahane, P., Pessiglione, M., Bastin, J., 2021. Intracerebral mechanisms explaining the impact of incidental feedback on mood state and risky choice. Neuroscience. https://doi.org/10.1101/2021.06.01.446610 (preprint).

Chase, H.W., Nusslock, R., Almeida, J.R., Forbes, E.E., LaBarbara, E.J., Phillips, M.L., 2013. Dissociable patterns of abnormal frontal cortical activation during anticipation of an uncertain reward or loss in bipolar versus major depression. Bipolar Disord. 15, 839–854. https://doi.org/10.1111/bdi.12132.

Cheng, X., Wang, L., Lv, Q., Wu, H., Huang, X., Yuan, J., Sun, X., Zhao, X., Yan, C., Yi, Z., 2022. Reduced learning bias towards the reward context in medication-naive first-episode schizophrenia patients. BMC Psychiatry 22, 123. https://doi.org/10.1186/s12888-021-03682-5.

Daunizeau, J., Adam, V., Rigoux, L., 2014. VBA: a probabilistic treatment of nonlinear models for neurobiological and behavioural data. PLoS Comput. Biol. 10, e1003441 https://doi.org/10.1371/journal.pcbi.1003441.

Davis, J.M., Chen, N., 2004. Dose response and dose equivalence of antipsychotics. J. Clin. Psychopharmacol. 24, 192–208. https://doi.org/10.1097/01.jcp.0000117422.05703.ae.

Duek, O., Osher, Y., Belmaker, R.H., Bersudsky, Y., Kofman, O., 2014. Reward sensitivity and anger in euthymic bipolar disorder. Psychiatry Res. 215, 95–100. https://doi.org/10.1016/j.psychres.2013.10.028.

Eldar, E., Rutledge, R.B., Dolan, R.J., Niv, Y., 2016. Mood as representation of momentum. Trends Cogn. Sci. 20, 15–24. https://doi.org/10.1016/j.tics.2015.07.010.

Frey, B.N., Andreazza, A.C., Houenou, J., Jamain, S., Goldstein, B.I., Frye, M.A., Leboyer, M., Berk, M., Malhi, G.S., Lopez-Jaramillo, C., Taylor, V.H., Dodd, S., Frangou, S., Hall, G.B., Fernandes, B.S., Kauer-Sant'Anna, M., Yatham, L.N., Kapczinski, F., Young, L.T., 2013. Biomarkers in bipolar disorder: a positional paper from the International Society for Bipolar Disorders Biomarkers Task Force. Aust. N. Z. J. Psychiatry 47, 321–332. https://doi.org/10.1177/0004867413478217.

Geana, A., Barch, D.M., Gold, J.M., Carter, C.S., MacDonald, A.W., Ragland, J.D., Silverstein, S.M., Frank, M.J., 2021. Using computational modeling to capture schizophrenia-specific reinforcement learning differences and their implications on patient classification. Biol. Psychiatry Cogn. Neurosci. Neuroimaging, S2451902221001154. https://doi.org/10.1016/j.bpsc.2021.03.017.

Gershman, S.J., 2019. Uncertainty and exploration. Decision 6, 277–286. https://doi.org/10.1037/dec0000101.

Gueguen, M.C.M., Lopez-Persem, A., Billeke, P., Lachaux, J.-P., Rheims, S., Kahane, P., Minotti, L., David, O., Pessiglione, M., Bastin, J., 2021. Anatomical dissociation of intracerebral signals for reward and punishment prediction errors in humans. Nat. Commun. 12, 3344. https://doi.org/10.1038/s41467-021-23704-w.

Guglielmo, R., Miskowiak, K.W., Hasler, G., 2021. Evaluating endophenotypes for bipolar disorder. Int. J. Bipolar Disord. 9, 17. https://doi.org/10.1186/s40345-021-00220-w.

Hafeman, D.M., Chang, K.D., Garrett, A.S., Sanders, E.M., Phillips, M.L., 2012. Effects of medication on neuroimaging findings in bipolar disorder: an updated review: medication effects on neuroimaging in bipolar disorder. Bipolar Disord. 14, 375–410. https://doi.org/10.1111/j.1399-5618.2012.01023.x.

Hägele, C., Schlagenhauf, F., Rapp, M., Sterzer, P., Beck, A., Bermpohl, F., Stoy, M., Ströhle, A., Wittchen, H.-U., Dolan, R.J., Heinz, A., 2015. Dimensional psychiatry: reward dysfunction and depressive mood across psychiatric disorders. Psychopharmacology 232, 331–341. https://doi.org/10.1007/s00213-014-3662-7.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D.S., Quinn, K., Sanislow, C., Wang, P., 2010. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. AJP 167, 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379.

Kobza, S., Ferrea, S., Schnitzler, A., Pollok, B., Südmeyer, M., Bellebaum, C., 2012. Dissociation between active and observational learning from positive and negative feedback in Parkinsonism. PLoS One 7, e50250. https://doi.org/10.1371/journal.pone.0050250.

Lewandowski, K.E., Whitton, A.E., Pizzagalli, D.A., Norris, L.A., Ongur, D., Hall, M.-H., 2016. Reward learning, neurocognition, social cognition, and symptomatology in psychosis. Front. Psychiatry 7. https://doi.org/10.3389/fpsyt.2016.00100.

Linke, J., Sönnekes, C., Wessa, M., 2011. Sensitivity to positive and negative feedback in euthymic patients with bipolar I disorder: the last episode makes the difference: the last episode makes the difference. Bipolar Disord. 13, 638–650. https://doi.org/10.1111/j.1399-5618.2011.00956.x.

Linke, J., King, A.V., Rietschel, M., Strohmaier, J., Hennerici, M., Gass, A., Meyer-Lindenberg, A., Wessa, M., 2012. Increased medial orbitofrontal and amygdala activation: evidence for a systems-level endophenotype of bipolar I disorder. AJP 169, 316–325. https://doi.org/10.1176/appi.ajp.2011.11050711.

Long, X., Wang, X., Tian, F., Cao, Y., Xie, H., Jia, Z., 2022. Altered brain activation during reward anticipation in bipolar disorder. Transl. Psychiatry 12, 300. https://doi.org/10.1038/s41398-022-02075-w.

Lozano, B.E., Johnson, S.L., 2001. Can personality traits predict increases in manic and depressive symptoms? J. Affect. Disord. 63, 103–111. https://doi.org/10.1016/s0165-0327(00)00191-9.

Macoveanu, J., Stougaard, M.E., Kjærstad, H.L., Knudsen, G.M., Vinberg, M., Kessing, L. V., Miskowiak, K.W., 2022. Trajectory of aberrant reward processing in patients with bipolar disorder - a longitudinal fMRI study. J. Affect. Disord. 312, 235–244. https://doi.org/10.1016/j.jad.2022.06.053.

Mason, L., O'Sullivan, N., Montaldi, D., Bentall, R.P., El-Deredy, W., 2014. Decision-making and trait impulsivity in bipolar disorder are associated with reduced prefrontal regulation of striatal reward valuation. Brain 137, 2346–2355. https://doi.org/10.1093/brain/awu152.

McGorry, P.D., 2013. Early clinical phenotypes, clinical staging, and strategic biomarker research: building blocks for personalized psychiatry. Biol. Psychiatry 74, 394–395. https://doi.org/10.1016/j.biopsych.2013.07.004.

Miskowiak, K.W., Seeberg, I., Kjaerstad, H.L., Burdick, K.E., Martinez-Aran, A., Bonnin, C., Bowie, C.R., Carvalho, A.F., Gallagher, P., Hasler, G., Lafer, B., López-Jaramillo, C., Sumiyoshi, T., McIntyre, R.S., Schaffer, A., Porter, R.J., Purdon, S., Torres, I.J., Yatham, L.N., Young, A.H., Kessing, L.V., Van Rheenen, T.E., Vieta, E., 2019. Affective cognition in bipolar disorder: a systematic review by the ISBD targeting cognition task force. Bipolar Disord. 21, 686–719. https://doi.org/10.1111/bdi.12834.

Nair, A., Rutledge, R.B., Mason, L., 2020. Under the hood: using computational psychiatry to make psychological therapies more mechanism-focused. Front. Psychiatry 11, 140. https://doi.org/10.3389/fpsyt.2020.00140.

Nusslock, R., Alloy, L.B., 2017. Reward processing and mood-related symptoms: an RDoC and translational neuroscience perspective. J. Affect. Disord. 216, 3–16. https://doi.org/10.1016/j.jad.2017.02.001.

Palminteri, S., Justo, D., Jauffret, C., Pavlicek, B., Dauta, A., Delmaire, C., Czernecki, V., Karachi, C., Capelle, L., Durr, A., Pessiglione, M., 2012. Critical roles for anterior insula and dorsal striatum in punishment-based avoidance learning. Neuron 76, 998–1009. https://doi.org/10.1016/j.neuron.2012.10.017.

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R.J., Frith, C.D., 2006. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. Nature 442, 1042–1045. https://doi.org/10.1038/nature05051.

Phillips, M.L., Travis, M.J., Fagiolini, A., Kupfer, D.J., 2008. Medication effects in neuroimaging studies of bipolar disorder. AJP 165, 313–320. https://doi.org/10.1176/appi.ajp.2007.07071066.

Pizzagalli, D.A., Goetz, E., Ostacher, M., Iosifescu, D.V., Perlis, R.H., 2008. Euthymic patients with bipolar disorder show decreased reward learning in a probabilistic reward task. Biol. Psychiatry 64, 162–168. https://doi.org/10.1016/j.biopsych.2007.12.001.

Redlich, R., Dohm, K., Grotegerd, D., Opel, N., Zwitserlood, P., Heindel, W., Arolt, V., Kugel, H., Dannlowski, U., 2015 Oct. Reward Processing in Unipolar and Bipolar Depression: A Functional MRI Study. Neuropsychopharmacology 40 (11), 2623–2631. https://doi.org/10.1038/npp.2015.110. Epub 2015 Apr 16. PMID: 25881114; PMCID: PMC4569953.

Rigoux, L., Stephan, K.E., Friston, K.J., Daunizeau, J., 2014. Bayesian model selection for group studies — revisited. NeuroImage 84, 971–985. https://doi.org/10.1016/j.neuroimage.2013.08.065.

Rutledge, R.B., Skandali, N., Dayan, P., Dolan, R.J., 2014. A computational and neural model of momentary subjective well-being. Proc. Natl. Acad. Sci. U. S. A. 111, 12252–12257. https://doi.org/10.1073/pnas.1407535111.

Ryu, V., Ha, R.Y., Lee, S.J., Ha, K., Cho, H.-S., 2017. Behavioral and electrophysiological alterations for reinforcement learning in manic and euthymic patients with bipolar disorder. CNS Neurosci. Ther. 23, 248–256. https://doi.org/10.1111/cns.12671.

Satterthwaite, T.D., Kable, J.W., Vandekar, L., Katchmar, N., Bassett, D.S., Baldassano, C. F., Ruparel, K., Elliott, M.A., Sheline, Y.I., Gur, R.C., Gur, R.E., Davatzikos, C., Leibenluft, E., Thase, M.E., Wolf, D.H., 2015 Aug. Common and Dissociable Dysfunction of the Reward System in Bipolar and Unipolar Depression. Neuropsychopharmacology 40 (9), 2258–2268. https://doi.org/10.1038/npp.2015.75. Epub 2015 Mar 13. PMID: 25767910; PMCID: PMC4613620.

Schreiter, S., Spengler, S., Willert, A., Mohnke, S., Herold, D., Erk, S., Romanczuk-Seiferth, N., Quinlivan, E., Hindi-Attar, C., Banzhaf, C., Wackerhagen, C., Romund, L., Garbusow, M., Stamm, T., Heinz, A., Walter, H., Bermpohl, F., 2016. Neural alterations of fronto-striatal circuitry during reward anticipation in euthymic bipolar disorder. Psychol. Med. 46, 3187–3198. https://doi.org/10.1017/S0033291716001963.

Scott, J., Stanton, B., Garland, A., Ferrier, I.N., 2000. Cognitive vulnerability in patients with bipolar disorder. Psychol. Med. 30, 467–472. https://doi.org/10.1017/s0033291799008879.

Snaith, R.P., Hamilton, M., Morley, S., Humayan, A., Hargreaves, D., Trigwell, P., 1995. A scale for the assessment of hedonic tone the Snaith-Hamilton Pleasure Scale. Br. J. Psychiatry 167, 99–103. https://doi.org/10.1192/bjp.167.1.99.

Teixeira, A.L., Salem, H., Frey, B.N., Barbosa, I.G., Machado-Vieira, R., 2016. Update on bipolar disorder biomarker candidates. Expert. Rev. Mol. Diagn. 16, 1209–1220. https://doi.org/10.1080/14737159.2016.1248413.

Tohen, M., Frank, E., Bowden, C.L., Colom, F., Ghaemi, S.N., Yatham, L.N., Malhi, G.S., Calabrese, J.R., Nolen, W.A., Vieta, E., Kapczinski, F., Goodwin, G.M., Suppes, T., Sachs, G.S., Chengappa, K.R., Grunze, H., Mitchell, P.B., Kanba, S., Berk, M., 2009. The International Society for Bipolar Disorders (ISBD) Task Force report on the nomenclature of course and outcome in bipolar disorders. Bipolar Disord. 11, 453–473. https://doi.org/10.1111/j.1399-5618.2009.00726.x.

Trost, S., Diekhof, E.K., Zvonik, K., Lewandowski, M., Usher, J., Keil, M., Zilles, D., Falkai, P., Dechent, P., Gruber, O., 2014. Disturbed anterior prefrontal control of the mesolimbic reward system and increased impulsivity in bipolar disorder. Neuropsychopharmacol 39, 1914–1923. https://doi.org/10.1038/npp.2014.39.

Vinckier, F., Rigoux, L., Oudiette, D., Pessiglione, M., 2018. Neuro-computational account of how mood fluctuations arise and affect decision making. Nat. Commun. 9, 1708. https://doi.org/10.1038/s41467-018-03774-z.

Wessa, M., Kanske, P., Linke, J., 2014. Bipolar disorder: a neural network perspective on a disorder of emotion and motivation. Restor. Neurol. Neurosci. 32, 51–62. https://doi.org/10.3233/RNN-139007.

Whiteside, S.P., Lynam, D.R., 2001. The five factor model and impulsivity: using a structural model of personality to understand impulsivity. Personal. Individ. Differ. 30 (4), 669–689 (n.d.).

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

Xu, S., Sun, Y., Huang, M., Huang, Y., Han, J., Tang, X., Ren, W., 2021. Emotional state and feedback-related negativity induced by positive, negative, and combined reinforcement. Front. Psychol. 12, 647263 https://doi.org/10.3389/fpsyg.2021.647263.

Yechiam, E., Hayden, E.P., Bodkins, M., O'Donnell, B.F., Hetrick, W.P., 2008. Decision making in bipolar disorder: a cognitive modeling approach. Psychiatry Res. 161, 142–152. https://doi.org/10.1016/j.psychres.2007.07.001.

Yip, S.W., Worhunsky, P.D., Rogers, R.D., Goodwin, G.M., 2015. Hypoactivation of the ventral and dorsal striatum during reward and loss anticipation in antipsychotic and mood stabilizer-naive bipolar disorder. Neuropsychopharmacol 40, 658–666. https://doi.org/10.1038/npp.2014.215.

VII

Asymmetric influence of agency on mood during reward and punishment learning

Arnaud Pouchon[a,b], Fabien Vinckier[c,d,e], Marc Benhamou[c], Clément Dondé[a,b,f], Julien Bastin[g,1], Mircea Polosan[a,b,1]

a Univ. Grenoble Alpes, Inserm, U1216, CHU Grenoble Alpes, Grenoble Institut Neurosciences, 38000 Grenoble, France

b Department of Psychiatry, CHU Grenoble Alpes, 38000 Grenoble, France

c Motivation, Brain & Behavior (MBB) lab, Institut du Cerveau (ICM), Hôpital Pitié-Salpêtrière, F-75013 Paris, France

d Université Paris Cité, F-75006 Paris, France

e Department of Psychiatry, Service Hospitalo-Universitaire, GHU Paris Psychiatrie & Neurosciences, F-75014 Paris, France

f Department of Psychiatry, CH Alpes-Isère, 38000 Saint-Egrève, France

g Univ. Grenoble Alpes, Inserm, U1216, Grenoble Institut Neurosciences, 38000 Grenoble, France

1 These authors contributed equally to this work.

In preparation

Abstract

The outcomes of recent events influence how we feel, but does it play a role in our mood if those outcomes are the consequence of our own choices? We explored this question of agency in healthy subjects (HC) and subjects with altered mood regulation (bipolar disorder, BD) alike, using a probabilistic reinforcement learning task. In this task, participants were either free to choose themselves or forced to confirm a pre-made choice. This choice between two simple cues resulted in rewarding or punishing outcomes, and their subjective well-being was subsequently probed. We found that reward learning was impeded in bipolar patients compared to healthy controls and that actively making a choice resulted in more efficient learning in both groups (HC/BD) and both outcome conditions (Reward/Punishment). Most importantly, we found that being an agent increased subjective well-being only when the outcome of a choice was rewarding and did not significantly influence subjective well-being when the outcome was punishing. Agency in the context of reinforcement learning seems to play a more important role in mood after rewards than after punishments.

Keywords

Mood; Bipolar disorder; agency; reinforcement learning; reward; punishment

1. Introduction

Our mood is influenced by our recent experiences and represents the cumulation of recent short-term changes in our affective state, with the last received outcome of an event often having the most weight (Rutledge et al., 2014; Vinckier et al., 2018). Due to this, the mood has been proposed to represent environmental momentum (Eldar et al., 2016), indicating if our environment is currently changing for the better, or for the worse. Supporting this idea, changes in mood were shown to be more dependent on the expectation of an outcome, rather than its absolute value (Rutledge et al., 2014). This benchmarking of outcomes against previous recent experiences is more suited to encode change in the environment, rather than the absolute state of the environment. The difference between the previous, thus expected, and the received outcomes is Prediction Error (PE). PE is essential for learning adaptive behaviors through reinforcing mechanisms (Blain and Rutledge, 2020), since they inform about environmental fluctuations and incentivize new behaviors which can lead to e.g. higher-than expected rewards. This creates inferred knowledge from past events, on the basis of which improved decisions can be taken. This process of refining decisions towards the highest gains is vital to an organism's fitness in a given environment.

It has been shown that the neural substrate for the calculation of reward-related PE are midbrain dopaminergic neurons (Pessiglione et al., 2006; Zaghloul et al., 2009). They can encode reward-prediction errors (RPE) during simple associative reinforcement-learning tasks (Pessiglione et al., 2006) by varying the timing and frequency of firing bursts (Zaghloul et al., 2009), though their role in punishment-learning is debated. These reinforcement learning tasks, which consist of repeated simple decisions between reward or punishment related cues, have since then helped to understand the functioning of dopamine and cortico-basal ganglia-thalamo-cortical (CBGTC) circuits. Due to this, those tasks have been proven increasingly useful for the understanding of neurological disorders involving disturbances of the dopaminergic system, which can result in altered regulation of mood, and/or learning (Maia and Frank, 2011).

Indeed, it has been shown that mood is intricately linked to learning (mood influences learning (Vinckier et al., 2018) and learning influences mood (Rutledge et al., 2014)). Additionally, it has been shown that agency plays a major role in learning: learning is more efficient from outcomes resulting from freely chosen options, as opposed to passively experienced outcomes (forced choices) (Chambon et al., 2020; Murayama et al., 2015).

Since agency plays a major role in learning, and learning is closely interlinked with mood, could agency have a direct link to mood? Agency's effects on mood have never been researched, neither in healthy subjects nor in subjects with mood disorders. The currently used computational models of subjective well-being are thus lacking the input of important agency-related parameters. Beyond the fundamental question of agency's effects on mood in the context of adaptive reinforcement learning, clinical research within this topic might provide insights into maladaptive behaviors and mood dysregulations in patients with neurological disorders. One neurological disorder where dopamine plays a major pathophysiological role is bipolar disorder (BD) (Ashok et al., 2017). In BD, decision-making and mood change drastically between (hypo-) maniac and depressive episodes which can last for weeks, with some cognitive symptoms even persisting through euthymic (in-between) phases (Grande et al., 2016).

A possible hypothesis for this episodic change is an altered sensitivity to reward (Mason et al., 2014), namely hyper- or hyposensitivity, which would trigger dysfunctional positive feedback loops (Eldar et al., 2016) leading to mood instability and causing the observed episodes. For some time, the reward-hypersensitivity hypothesis has been present in the scientific consensus, especially during mania (Johnson et al., 2012) but recent findings increasingly report the presence of reward hyposensitivity in BD patients (Miskowiak et al., 2019; Pouchon et al., 2023; Schreiter et al., 2016; Trost et al., 2014; Yip et al., 2015).

These heterogeneous results might be due to a lack of differentiation between reward and punishment learning, which are hypothesized to take place in different brain regions, like the ventromedial prefrontal cortex (vmPFC) and ventral striatum (VS) for

reward, and the anterior insula and the dorsomedial prefrontal cortex for punishment (Palminteri et al., 2015). Alternatively, they could be due to the possible differences between the two clinically described types of BD. Generally, bipolar type-I includes more manic episodes, while bipolar type-II includes more depressive episodes (Grande et al., 2016). Since altered mood regulation is integral to bipolar disorder, and mood is tightly linked to learning, the question if bipolar patients show altered learning performances naturally arises. Indeed, some studies show that bipolar patients seem to be less efficient in learning from rewards than healthy controls (HC) (Pizzagalli et al., 2008; Pouchon et al., 2023). This result raises a second question: could this reward-learning hyposensitivity in BD patients also be reflected in the short-term changes in their mood? Together with the question of agency's general influence on mood, the answers could help untangle the links between agency, mood, and learning further, as well as hint at physiological differences between BD and HC dopaminergic systems.

To investigate the impact of agency on mood and learning in HC and BD, we designed a study with a computer-based probabilistic reinforcement-learning task. Subjects are either agent (free choice) or non-agent (forced choice) and learn which choices to make through reward- or punishment-associated cues, after which they are asked to report their current mood. We tested the influences of the independent variables of outcome (reward vs. punishment), agency (free vs. forced), and pathology (HC vs. BD) on the dependent variables of learning and mood.

2. Method

a. Participants

64 subjects were included in this study. 32 BD patients were recruited from Grenoble-Alpes University Hospital, and 32 matched healthy subjects were recruited. Patients were matched to healthy controls using sociodemographic variables including sex, age, and education level (**Table VII.1**). The Snaith-Hamilton Pleasure Scale (SHAPS, (Snaith et al., 1995)) was completed by all subjects to assess their hedonic capacity since anhedonia is linked to blunted dopamine-related reward responses (Huys et al., 2013). Additionally, the UPPS

(Impulsive Behaviour Scale) questionnaire was filled out by each subject since impulsivity also interferes with reward sensitivity (Mason et al., 2014). Patient's clinical data including age at BD onset, duration of the present euthymic phase, the polarity of their first and last episode, number of past (hypo)maniac, thymic and depressive episodes, number of hospitalizations and type/dosage of medications were collected. 16 BD patients were diagnosed with BD-I, 14 were diagnosed with BD-II, and 2 were of unknown type. For patients, inclusion criteria were BD diagnosis as defined by DSM-5 and remitted (euthymic) state defined by Montgomery Asberg Depression Rating Scale (MADRS) lower than 15. Young Mania Rating Scale (YMRS) had to be rated lower than 12 for at least one month. All Bipolar patients were in the remitted phase, and their previous episode was predominantly depressive (25 of 32). All patients underwent a neuropsychological assessment during the past 2 years to check the absence of cognitive dysfunctions, in particular working memory, episodic memory, executive functions, and attention span.

General non-inclusion criteria were a history of neurological disorders or systemic illnesses with neurological complications other than BD, lack of fluency in French, recent history (6 months) of substance abuse or dependence except for tobacco, brain stimulation procedure, or previous head trauma with unconsciousness. All participants received a compensation of 10€. This study was approved by the French Institutional Ethic Committee (CPP Sud-Est II) and conforms to international standards for testing with human participants (Declaration of Helsinki). The experiment was registered at ClinicalTrials.gov (NCT05025566).

Table VII.1: Socio-demographic and clinical characteristics for healthy control (HC) and bipolar disorder (BD).

Demographics	BD (n = 32)	HC (n = 32)	P-Value
Sex (Female/Male)	18/14	19/13	1.000
Age in years. Mean (SD)	45.09 (13.03)	44.38 (12.02)	0.7513
Years of education. Mean (SD)	3.81 (2.28)	4.03 (2.39)	0.6904
UPPS	100.02 (15.75)	99.94 (11.73)	0.8128

Range	[72-130]	[76-126]	
SHAPS Range	1.18 (1.14) [0-4]	0.71 (0.92) [0-3]	0.13
MADRS Range	3.80 (3.71) [0-14]	-	-
YMRS Range	1.10 (1.64) [0-5]	-	-

Abbreviations: BD = bipolar disorder; HC = healthy controls; MADRS = Montgomery Asberg Depression Rating Scale; SHAPS = Snaith–Hamilton pleasure scale; UPPS = UPPS Impulsive Behavior Scale; YMRS = Young Mania Rating Scale.

b. Behavioral Task

The probabilistic learning task designed for this study was adapted from Chambon et al. (2020), and modified to include mood ratings (Chambon et al., 2020). It consists of a simple choice between two symbols, a "better" one with a higher probability of reward (75%) or a lower probability of punishment (25%), and a "worse" one with a low probability of reward (25%) or a high probability of punishment (75%). The possible outcomes in the reward condition were +2 (37.5%), +1 (37.5%), and 0 (25%) for the better symbol, and +2 (12.5%), +1 (12.5%) and 0 (75%) for the worse symbol. In the punishment condition, they were -2 (37.5%), -1 (37.5%), and 0 (25%) for the worse symbol, and -2 (12.5%), -1 (12.5%), and 0 (75%) for the better symbol.

The inclusion of a 0 in both conditions can lead to lower learning rates and thus longer learning times, but is necessary to separate reward and punishment tasks within the same experiment. The task design included four different learning conditions in a 2x2 structure. Participants were presented with a rewarding pair or a punishing pair and were instructed to either choose their preferred symbol themselves (free), or to match the computer's choice (forced) by pressing the corresponding key. Each of the four learning-conditions ((1) Free-Reward, (2) Forced-Reward, (3) Free-Punishment, and (4) Forced-Punishment) had its own associated pair of symbols (**Figure VII.VII.1a**). Conditions were pseudo-randomly mixed to avoid boring the subjects. Each of the four symbol pairs (conditions) appeared 24 times per session, resulting in 96 trials per session. All participants'

data was included. All except two performed 4 sessions, totaling 384 trials per subject. Participants performed a training session in which the symbols were replaced with letters prior to the experimental sessions (**Figure VII.VII.1d**). This was done to get subjects used to the task without creating associations. Each session used new symbol pairs to avoid association transfer from previous sessions. Subjects received written instructions before the tasks, which were repeated orally if necessary. The task's goal was to maximize points by equally seeking gains and avoiding losses, through the correct choice of the "better" symbol of a pair. Before each task the respective agency was shown (Actor or Observer, 0.5 seconds), after which a fixation cross (0.5 seconds) was followed by a pair of symbols from the agathodaemon alphabet, randomly arranged to the left and right of the fixation cross. After a symbol was confirmed by the participant using the corresponding arrow keys, the resulting feedback was shown (0.5 seconds) (**Figure VII.VII.1b**). The probabilities associated with the symbols were implicit and had to be learned. To evaluate the correct choice rates in the forced condition, "open choices" were enabled during the last two trials of each forced condition (**Figure VII.VII.1c**, "open choices"). The computer-generated choices were matched to the free choices of the participants to match the amounts of right and wrong choices in the forced condition and the free condition. This makes learning rates and mood ratings comparable between the conditions (**Figure VII.VII.1c**).

Mood ratings consisted of a slider with a scrambled starting position, on which participants could rate how they currently feel (**Figure VII.VII.1b**). Mood ratings occurred pseudo-randomly in 25% of trials, but never in two subsequent trials, and equally frequently for each of the 4 conditions.

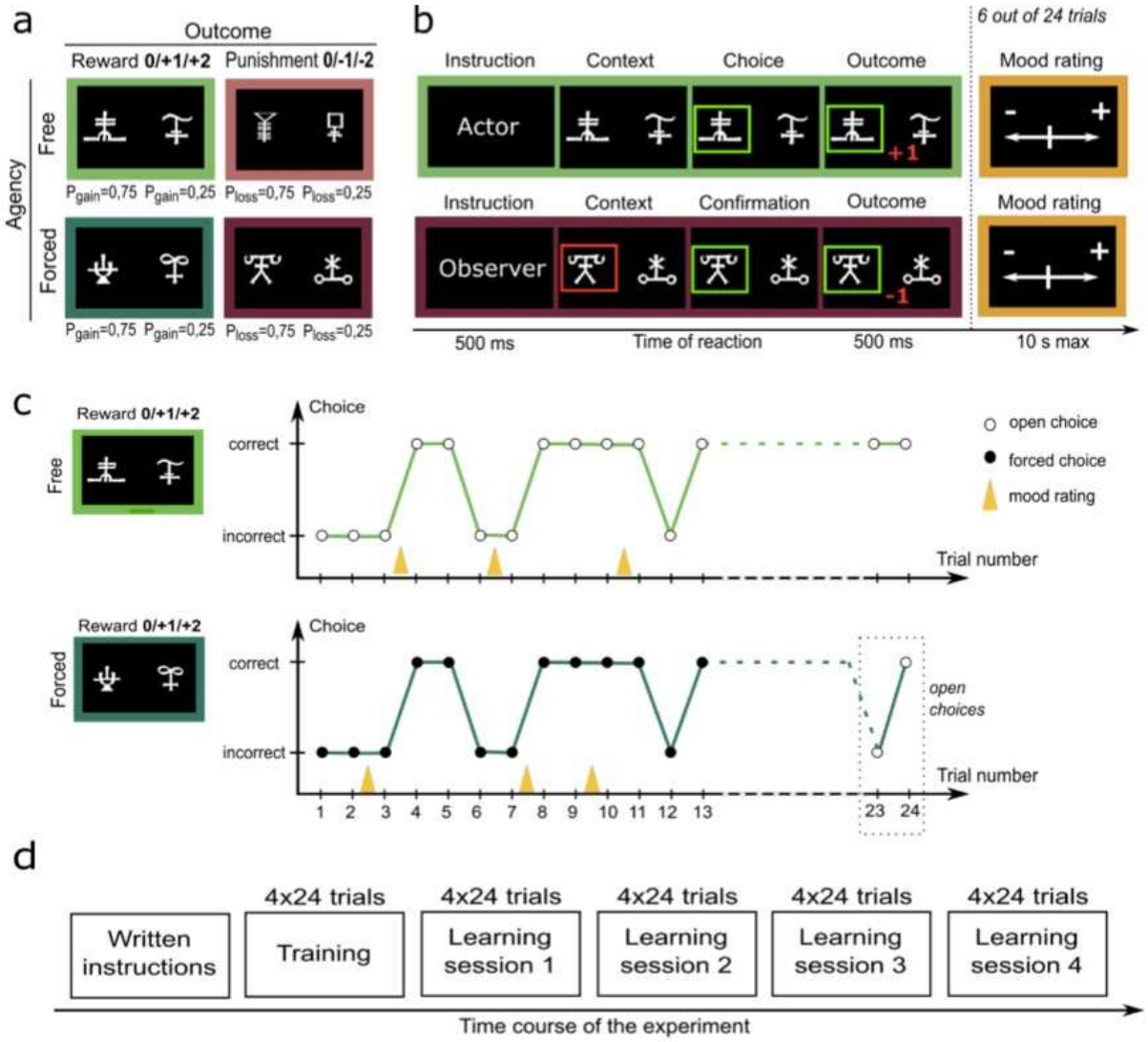

Figure VII.VII.1: Experimental design of the reinforcement-learning tasks. a) Each of the four experimental conditions (Free-Reward, Forced-Reward, Free-Punishment, and Forced-Punishment has its associated pair. Outcomes and probabilities are indicated above and below. **b)** The sequence of screens seen by a subject during a single trial. Examples of free-reward (above) and forced-punishment (below) trials. In the free condition, participants choose a symbol, while in the forced condition they must confirm the computer's predetermined choice by pressing the according key. Mood-ratings occurred in 25% of all trials. Time in milliseconds. **c)** Typical process during a session. In the free condition (above), the subject will choose symbols that will turn out to be either correct or incorrect, and the frequency of correct choices increases in later trials if learning occurs. In the forced condition (below), the computer matches the choices of the respective free condition, which must be confirmed by the subject. This was done to keep the free and forced conditions comparable. The last two trials of each forced condition are open choices, enabling a test of the subjects' learning. **d)** Time course of the experiment. Each subject performed a training session and four subsequent learning sessions in one sitting.

c. Statistical analysis

MATLAB and PsychToolBox were used to program the task and acquire the data. Data and statistical analysis were conducted in MATLAB (R2020b), RStudio (2023.03.0), and GraphPad Prism (9.5.1). Demographic and clinical data for the two groups were analyzed using Mann-Whitney tests for age, education level, SHAPS scores, and UPPS scores. Fisher's exact test was performed for gender. Statistical analysis of this study 2x2x2 design requires repeated-measures three-way ANOVAs with two levels each for Outcome (Reward, Punishment), Pathology (HC, BD), and Agency (Free, Forced). After testing data for normality with QQ plots, ANOVAs were performed for the dependent outcome variables of correct choice (learning) and mood. Correct choice rates for each of the "free" conditions were tested against H0 = 50% using two-tailed one-sample t-tests after testing for normality using the d'Agostino & Pearson test.

3. Preliminary Results

To confirm that subjects understood the task, and to elucidate possible differences in the learning rates, subjects' learning performances during the free-choice conditions were tested against 50%. This represents the null hypothesis of random choices occurring without learning (**Figure VII.VII.2**). Following this, a three-way ANOVA analysis was performed to distinguish the influence of agency on learning performances (**Figure VII.VII.3**). Lastly, a three-way ANOVA analysis was performed on mood ratings, showing that the agency's impact on mood depends on the outcome (**Figure VII.VII.4**). No significant differences between BD-I and BD-II were found in learning performance (Bonferroni's multiple comparisons tests, $p > 0.9999$ for all conditions) and in subjective mood ratings (Dunnett's T3 multiple comparison tests, $p > 0.6062$ for all conditions), allowing both bipolar types to be pooled for all analyses.

a. Bipolar patients show learning deficiency in the free-reward condition.

To show that the participants of both HC and BD groups learned during the task, their respective free choices in the reward and punishment conditions were checked against 50%

(**Figure VII.VII.2a**) and plotted across trials which indicates their learning performance over time (**Figure VII.VII.2b Figure VII.VII.2c**).

Though BD patients seemed to have a slight performance increase over time in the reward condition (**Figure VII.VII.2b**), it was not significant against the 50% expected from random choice ($p = 0.2138$) when all trials were averaged. All other conditions passed the test against 50% (HC-reward: $p = 0.0006$; HC-punishment: $p < 0.0001$; BD-punishment: $p < 0.0001$) (**Figure VII.VII.2b**). This indicates that both BD and HC subjects understood the task since they performed equally well in the punishment condition, but shows that BD subjects learned less efficiently than HC in the reward condition.

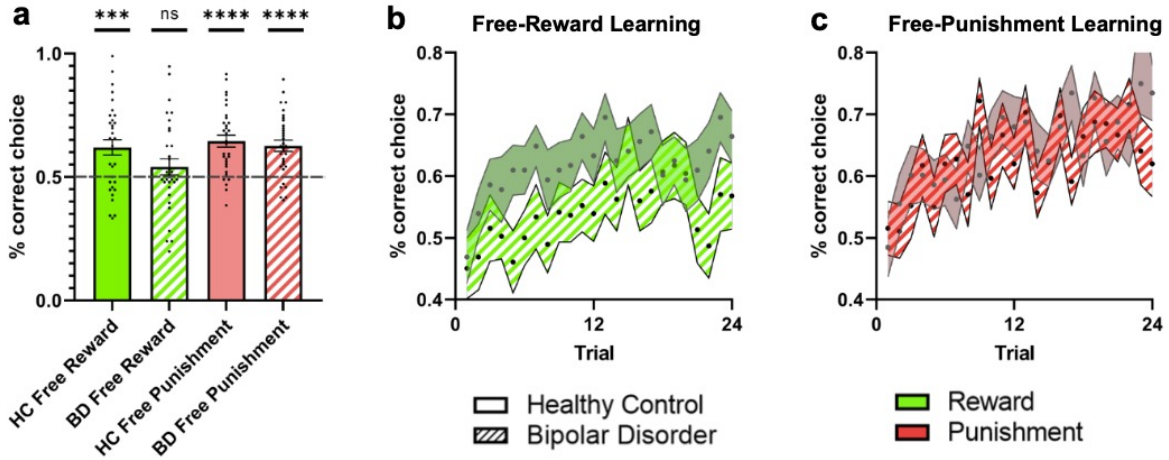

Figure VII.VII.2: Correct choice rates in Free-Reward and Free-Punishment learning conditions. a) Average correct choice rates tested against 50%. Each of the conditions was tested using two-tailed one-sample t-tests, after testing for normality using the d'Agostino & Pearson test ($p > 0.4$ for all conditions). Each dot represents the average score per subject, calculated from 96 trial values. Note how reward learning seems to be less efficient in BD than in HC since all conditions except BD Free-Reward show statistically significant learning. Error bars represent the mean ± S.E.M. **b) and c)** Learning performance over time. A dot represents the average of 128 trials (32 subjects over 4 sessions). Correct choice rate increases in all conditions, and is almost completely overlayed in the Free-Punishment condition. The filled area shows HC, shaded area shows BD. Outlines represent the mean ±S.E.M.

b. Agency influences learning.

To assess potential between-group learning differences and explore the effect of agency, a 2x2x2 repeated-measures three-way ANOVA of the correct choice rates was performed. **Figure VII.VII.3a** visualizes the input data. For this analysis, only the last two trials of 24, the "open choice" trials (see **Figure VII.VII.3c**), were taken for each condition. This reduces statistical power but allows us to compare the free and forced conditions since only the last two trials in the forced conditions tested participants' learning by giving them an open choice. No general difference between punishment and reward learning was found ($F_{(1,62)}$ = 1.780, p = 0.187), and there were no significant interaction effects, which will be discussed later. The main effects of Pathology (BD vs HC: $F_{(1,62)}$ = 4.435, p = 0.0393) and of Agency (Free vs Forced: $F_{(1,62)}$ = 7.456, p = 0.0082, pooling shown in **Figure VII.VII.3b**) were revealed. BD resulted in lower learning performance than healthy controls on average, and forced choices resulted in lower learning performances on average than free choices.

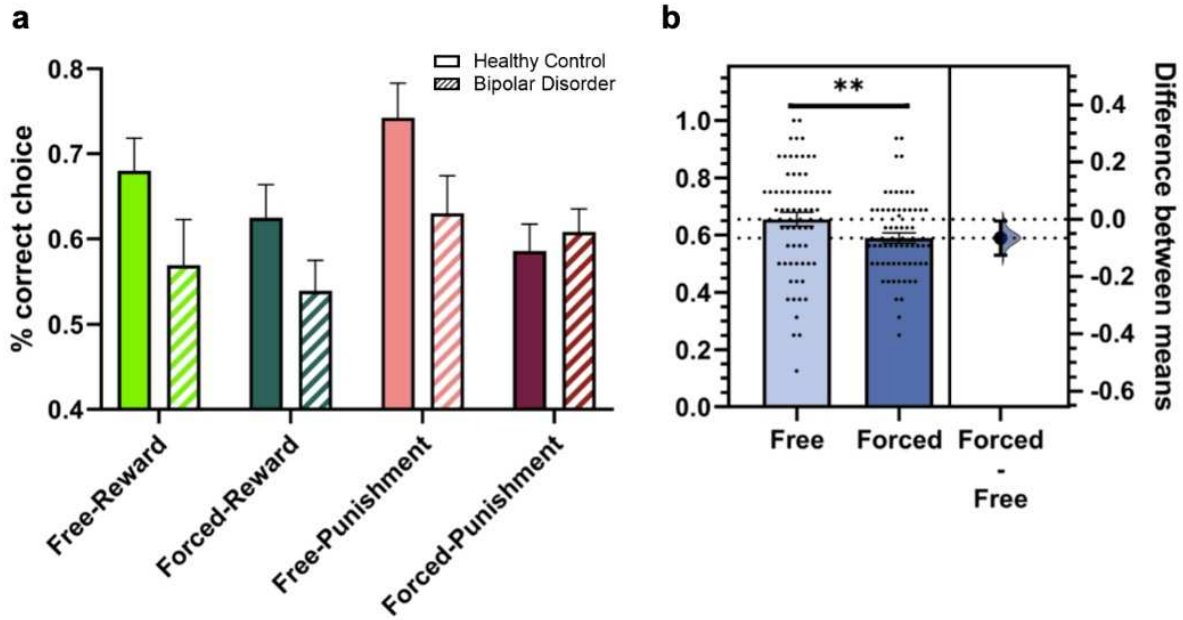

Figure VII.VII.3: Analysis of agency's effect on learning. a) Graph showing the input for the three-way repeated measures ANOVA, depicting learning rates calculated from the last two "open choice" test trials of each condition. Error bars represent the mean ±S.E.M. **b)** Estimation plot showing the main effect of agency on learning rates. The average percentage of correct choices in open trials on the left y-axis. Dots represent individual subjects. Free choices result in higher learning performance than forced choices. The right side shows the quantitative difference between the two factors, the bar representing the 95% CI including distribution information.

c. Agency influences mood differently depending on the received outcome.

Having seen the influence of agency on learning, a repeated-measures three-way ANOVA was performed to assess the possible effects of agency, pathology, and outcome on mood. **Figure VII.VII.4a** show the input data as z-scored mood ratings. Z-scoring allows to display of mood in relation to the average mood of all conditions in terms of standard deviation.

The analysis revealed a main effect of outcome ($F_{(1,62)} = 82.89$, $p < 0.0001$), showing that punishment led to a lower subjective mood rating than reward, and an interaction effect of agency x outcome ($F_{(1,62)} = 20.76$, $p < 0.0001$) (**Figure VII.VII.4b**). There was no significant main effect of pathology ($F_{(1,62)} = 0.178$, $p = 0.6745$), or between free and forced choices ($F_{(1,62)} = 3.034$, $p = 0.0865$). A post-hoc Tukey's multiple comparison-adjusted t-test was performed on the categories that take part in the Agency x Outcome interaction (**Figure VII.VII.4b**), with Forced-Reward resulting in a significantly lower mood than Free-Reward ($p = 0.0003$). Free or forced choices did not result in significantly different moods when the outcome was punishing. This implies that choosing actively impacts mood more in the reward condition than in the punishment condition.

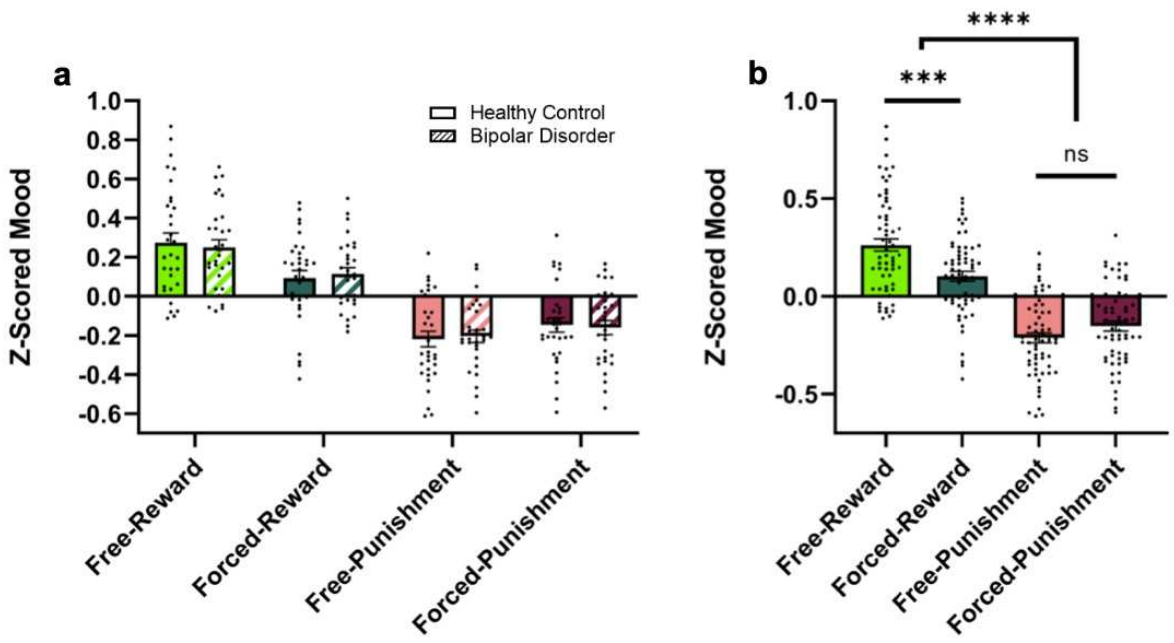

Figure VII.VII.4: Z-scored mood ratings. Mood was subjectively rated on a scrambled slider between -50 and +50. Dots represent individual subjects' deviation from the mean in terms of SD. **a)** Input data for the three-way repeated measures ANOVA. Each dot represents a subject's average of 24 ratings per condition. No difference in mood could be detected between BD and HC subjects. **b)** Interaction effect of Agency x Outcome. Mood was rated higher after reward than after punishment, and Mood was affected differently in reward when the choice was made freely. In punishment, no difference in mood after free and forced choices could be detected. Post-hoc analysis with Tukeys' multiple comparisons test on the categories involved in the interaction. Error bars show mean ± S.E.M.

Generally, the agency is a major influence on mood and learning, since it creates significant differences in both. The outcome (reward/punishment) only has a significant influence on mood, but not on learning, while bipolar disorder only has a significant influence on learning, but not on momentary mood fluctuations. The agency is not only connected to reinforcement learning but also to the changes in mood resulting from received outcomes.

4. Preliminary discussion

In this study, an instrumental reinforcement-learning task was used to assess the influences of interactions between the independent variables of outcome (reward vs. punishment), agency (free vs. forced), and pathology (HC vs. BD) on the dependent variables of learning

and mood. This was done to explore the effects of agency on mood, which has not been quantitatively assessed to date, and to test if the pathophysiological mechanisms underlying reward hyposensitivity in BD patients during learning (Pizzagalli et al., 2008; Pouchon et al., 2023) are also reflected in their mood.

No significant differences between BD-I and BD-II were found in learning performance and subjective mood ratings. We found an altered sensitivity to reward in BD subjects for learning from free-choice trials compared to HC, but no altered sensitivity regarding their mood changes compared to HC. Being an agent in a choice was found to improve learning performance in both reward and punishment conditions, but to only significantly affect subjective well-being in the reward condition.

a. Free-choice learning

BD subjects' impaired reward-learning is consistent with previous studies' findings (Pizzagalli et al., 2008; Pouchon et al., 2023), presenting further evidence for the hyposensitivity hypothesis for euthymic BD patients. Supporting this, a previous systematic review found that 6/15 studies with BD patients indicated hyposensitivity while 9/15 studies found no difference, and none of those studies found reward-hypersensitivity (Miskowiak et al., 2019). It also found attention deficits in BD patients, but this alone cannot explain the lower learning performance in reward trials, since no performance drop could be observed in the forced condition (**Figure VII.VII.2c**). It should be noted that the difference between BD and HC free-reward learning observed in **Figure VII.VII.2a** and **Figure VII.VII.2b** depends on the outcome, suggesting an interaction effect between pathology and outcome. This interaction effect was not significant ($p = 0.2515$) in the ANOVA analysis (**Figure VII.VII.3**), since only using the last two of 24 trials for analysis majorly reduces the statistical power of the test, compared to the comparisons of free choices (**Figure VII.VII.2**) which used data from all 24 free-choice trials.

b. Agency's influence on learning

Assessing the effect of agency on learning performance, it can be seen that higher performances are associated with free choices, which is coherent with previous studies, for example, Murayama, K. et al. (2013) found that the vmPFC drops in activation when the choice is forced, but is activated in free-choice conditions, resulting in higher learning performances (Chambon et al., 2020; Murayama et al., 2015). Chambon et al. (2020) document better learning performances in the reward condition when participants are free to choose, but equal learning performance between reward and punishment when participants are forced to choose (Chambon et al., 2020). We do not see this positivity bias, as we see similar average learning performances between punishment and reward (**Figure VII.VII.3**).

c. Healthy control and bipolar disorder subjects' mood ratings

The finding that BD learning performance was worse than HC in the free-reward condition (**Figure VII.VII.2a**) indicating reward hyposensitivity in this domain, but that BD did not show a significant difference to HC in their mood following free-reward trials (**Figure VII.VII.4a**), could suggest that some of the neural pathophysiological mechanisms responsible for this hyposensitivity might be connected to areas of the brain which are more related to reward-learning than to areas responsible for short-term changes in mood. The finding that mood did not statistically differ between BD and HC might seem surprising since BD is characterized as a mood disorder. However since BD is defined by longer-term mood shifts, this finding is still consistent with the general functioning of BD patients during euthymia (Grande et al., 2016).

d. Agency's asymmetric outcome-dependent influence on mood

Seeing the increased effectiveness of free choices in learning (**Figure VII.VII.3b**), it could be hypothesized that free choices also result in an increased effect on mood. This hypothesis would rely on the fact that both mood and learning are more dependent on the unexpectedness of an outcome than on its absolute value, meaning they are both related to PE (Rutledge et al., 2014). The act of making a free choice would create a stronger

expectation than experiencing a forced choice, which could result in more defined prediction errors and thus an increased influence on mood. We found with an Agency x Outcome interaction that this is indeed the case for reward, but might not be the case for punishment.

e. Strengths and limitations

A strength of this study is the relatively large number of bipolar subjects (n=32), though a potential limitation of this study is that most BD subjects had a previous episode of depressive polarity (25 out of 32 subjects), which limits the predictive power towards BD patients with previous manic or hypomanic episodes. An additional limitation is that BD patients were medicated with different drugs, potentially altering their cognitive processes. Though it is almost impossible to perform a study on unmedicated patients, this nonetheless represents an additional confounding factor in the results.

5. References

Ashok, A.H., Marques, T.R., Jauhar, S., Nour, M.M., Goodwin, G.M., Young, A.H., Howes, O.D., 2017. The dopamine hypothesis of bipolar affective disorder: the state of the art and implications for treatment. Mol Psychiatry 22, 666–679. https://doi.org/10.1038/mp.2017.16

Blain, B., Rutledge, R.B., 2020. Momentary subjective well-being depends on learning and not reward. eLife 9, e57977. https://doi.org/10.7554/eLife.57977

Chambon, V., Théro, H., Vidal, M., Vandendriessche, H., Haggard, P., Palminteri, S., 2020. Information about action outcomes differentially affects learning from self-determined versus imposed choices. Nat Hum Behav 4, 1067–1079. https://doi.org/10.1038/s41562-020-0919-5

Eldar, E., Rutledge, R.B., Dolan, R.J., Niv, Y., 2016. Mood as Representation of Momentum. Trends in Cognitive Sciences 20, 15–24. https://doi.org/10.1016/j.tics.2015.07.010

Grande, I., Berk, M., Birmaher, B., Vieta, E., 2016. Bipolar disorder. The Lancet 387, 1561–1572. https://doi.org/10.1016/S0140-6736(15)00241-X

Huys, Q.J., Pizzagalli, D.A., Bogdan, R., Dayan, P., 2013. Mapping anhedonia onto reinforcement learning: a behavioural meta-analysis. Biol Mood Anxiety Disord 3, 12. https://doi.org/10.1186/2045-5380-3-12

Johnson, S.L., Edge, M.D., Holmes, M.K., Carver, C.S., 2012. The behavioral activation system and mania. Annual review of clinical psychology 8, 243–267.

Long, X., Wang, X., Tian, F., Cao, Y., Xie, H., Jia, Z., 2022. Altered brain activation during reward anticipation in bipolar disorder. Transl Psychiatry 12, 300. https://doi.org/10.1038/s41398-022-02075-w

Maia, T.V., Frank, M.J., 2011. From reinforcement learning models to psychiatric and neurological disorders. Nat Neurosci 14, 154–162. https://doi.org/10.1038/nn.2723

Mason, L., O'Sullivan, N., Montaldi, D., Bentall, R.P., El-Deredy, W., 2014. Decision-making and trait impulsivity in bipolar disorder are associated with reduced prefrontal regulation of striatal reward valuation. Brain 137, 2346–2355.

Miskowiak, K.W., Seeberg, I., Kjaerstad, H.L., Burdick, K.E., Martinez-Aran, A., Bonnin, C., Bowie, C.R., Carvalho, A.F., Gallagher, P., Hasler, G., Lafer, B., López-Jaramillo, C., Sumiyoshi, T., McIntyre, R.S., Schaffer, A., Porter, R.J., Purdon, S., Torres, I.J., Yatham, L.N., Young, A.H., Kessing, L.V., Van Rheenen, T.E., Vieta, E., 2019. Affective cognition in bipolar disorder: A systematic review by the ISBD targeting cognition task force. Bipolar Disord 21, 686–719. https://doi.org/10.1111/bdi.12834

Murayama, K., Matsumoto, M., Izuma, K., Sugiura, A., Ryan, R.M., Deci, E.L., Matsumoto, K., 2015. How self-determined choice facilitates performance: A key role of the ventromedial prefrontal cortex. Cerebral Cortex 25, 1241–1251.

Palminteri, S., Khamassi, M., Joffily, M., Coricelli, G., 2015. Contextual modulation of value signals in reward and punishment learning. Nat Commun 6, 8096. https://doi.org/10.1038/ncomms9096

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R.J., Frith, C.D., 2006. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. Nature 442, 1042–1045. https://doi.org/10.1038/nature05051

Pizzagalli, D.A., Goetz, E., Ostacher, M., Iosifescu, D.V., Perlis, R.H., 2008. Euthymic Patients with Bipolar Disorder Show Decreased Reward Learning in a Probabilistic Reward Task. Biological Psychiatry 64, 162–168. https://doi.org/10.1016/j.biopsych.2007.12.001

Pouchon, A., Vinckier, F., Dondé, C., Gueguen, M.C., Polosan, M., Bastin, J., 2023. Reward and punishment learning deficits among bipolar disorder subtypes. Journal of Affective Disorders 340, 694–702. https://doi.org/10.1016/j.jad.2023.08.075

Rutledge, R.B., Skandali, N., Dayan, P., Dolan, R.J., 2014. A computational and neural model of momentary subjective well-being. Proc. Natl. Acad. Sci. U.S.A. 111, 12252–12257. https://doi.org/10.1073/pnas.1407535111

Snaith, R.P., Hamilton, M., Morley, S., Humayan, A., Hargreaves, D., Trigwell, P., 1995. A Scale for the Assessment of Hedonic Tone the Snaith–Hamilton Pleasure Scale. Br J Psychiatry 167, 99–103. https://doi.org/10.1192/bjp.167.1.99

Vinckier, F., Rigoux, L., Oudiette, D., Pessiglione, M., 2018. Neuro-computational account of how mood fluctuations arise and affect decision making. Nat Commun 9, 1708. https://doi.org/10.1038/s41467-018-03774-z

Zaghloul, K.A., Blanco, J.A., Weidemann, C.T., McGill, K., Jaggi, J.L., Baltuch, G.H., Kahana, M.J., 2009. Human substantia nigra neurons encode unexpected financial rewards. Science 323, 1496–1499.

DISCUSSION GENERALE

VIII

Qu'avons-nous appris ?

1. Il existe une altération sélective de l'apprentissage par récompense durant la rémission

Nous avons vu en introduction que les études portant sur les performances en apprentissage par renforcement dans le trouble bipolaire présentaient des **résultats disparates** (voir le **Tableau III.1**), notamment durant l'euthymie. Dans les cas de résultats hétérogènes lors d'une revue de littérature, et dans la mesure où ces résultats sont quantifiables, la **méta-analyse** prend tout son sens pour essayer de déterminer si un effet existe en augmentant la puissance statistique des études prises de façon isolée (Haidich, 2010; Moher et al., 2009). C'est pour cela que nous avons effectué notre première étude via cette approche méta-analytique des résultats des études comportementales. **Nos résultats préliminaires suggèrent un apprentissage altéré de façon sélective sur la récompense mais pas sur la punition dans le trouble bipolaire par rapport aux sujets sains**. Nous avons également trouvé que ce déficit se maintenait durant **l'euthymie**, pouvant de ce fait constituer un **marqueur-trait de la pathologie**, même si des études spécifiques sur le sujet apparaissent nécessaires (S. de Sá et al., 2016; Srivastava et al., 2019).

2. Qui semble expliquée par une hyposensibilité à la récompense

Nous avons ensuite étudié cela selon une autre approche et de manière **expérimentale** dans notre deuxième étude (qui a fait l'objet d'une publication), mêlant psychologie expérimentale et modélisation computationnelle. Sur les analyses faites en *model-free*, nous avons **confirmé cette altération sélective de l'apprentissage par récompense, sans altération de l'apprentissage par punition, dans le trouble bipolaire en rémission**. Nous n'avons pas répliqué les résultats de Linke et al. (2011), ce déficit ne semblant pas être en lien avec l'effet du dernier épisode, venant soutenir l'hypothèse d'un possible **marqueur-trait** du trouble bipolaire, et non un marqueur-état (Linke et al., 2011). Grâce à **l'approche computationnelle** de notre deuxième étude, nous avons essayé d'incriminer les **mécanismes cognitifs sous-jacents à l'apprentissage par renforcement** impliqués dans ce déficit de performance sélective sur l'apprentissage par récompense dans le trouble bipolaire (voir la **Figure I.6**). Ainsi, selon notre modèle, le déficit d'apprentissage par récompense que nous avons mis en avant à la fois dans la première et dans la deuxième étude (via deux méthodologies distinctes) serait dû à une **sensibilité émoussée à la récompense** par rapport aux sujets sains. Nous avons également observé une possible différence (non significative) entre les BD-I et les BD-II sur l'apprentissage de la punition, possiblement en lien avec **davantage de choix stochastiques chez les BD-II que les BD-I**.

3. L'agentivité biaise sélectivement l'humeur positive lors des choix

Enfin, nous nous sommes demandés si **l'agentivité pouvait contribuer à déstabiliser l'humeur** dans le trouble bipolaire lors d'une troisième étude (expérimentale) mêlant là encore psychologie expérimentale et modélisation computationnelle. Le fait de se sentir auteur, impliqué, ou responsable (ou non) d'une conséquence (bonne ou mauvaise) liée à un choix qu'un sujet a fait pourrait-il influencer l'humeur ? Suite à une accumulation de

prises de décision, cette agentivité pourrait-elle intervenir dans les fluctuations de l'humeur de façon différente que chez les sujets sains, et ainsi constituer un nouveau potentiel biomarqueur candidat pouvant expliquer les rechutes thymiques ? Nous avons mis en avant un résultat original au cours de ce travail de thèse. En effet, il semble que **le poids de l'agentivité sur notre humeur diffère entre la récompense et la punition**. C'est-à-dire que nous sommes **davantage « heureux » lorsque la récompense vient de notre propre fait** plutôt qu'une récompense « imposée » (qui vient sans effort, ce qui pourrait être étudié dans une autre étude), alors que nous ne sommes pas davantage tristes ou frustrés lorsqu'une punition est de notre fait ou non. Ces résultats proviennent d'analyses *model-free*, et sont encore en cours en *model-based*. La **Figure VIII.1** résume les résultats des 3 études de ce travail de thèse.

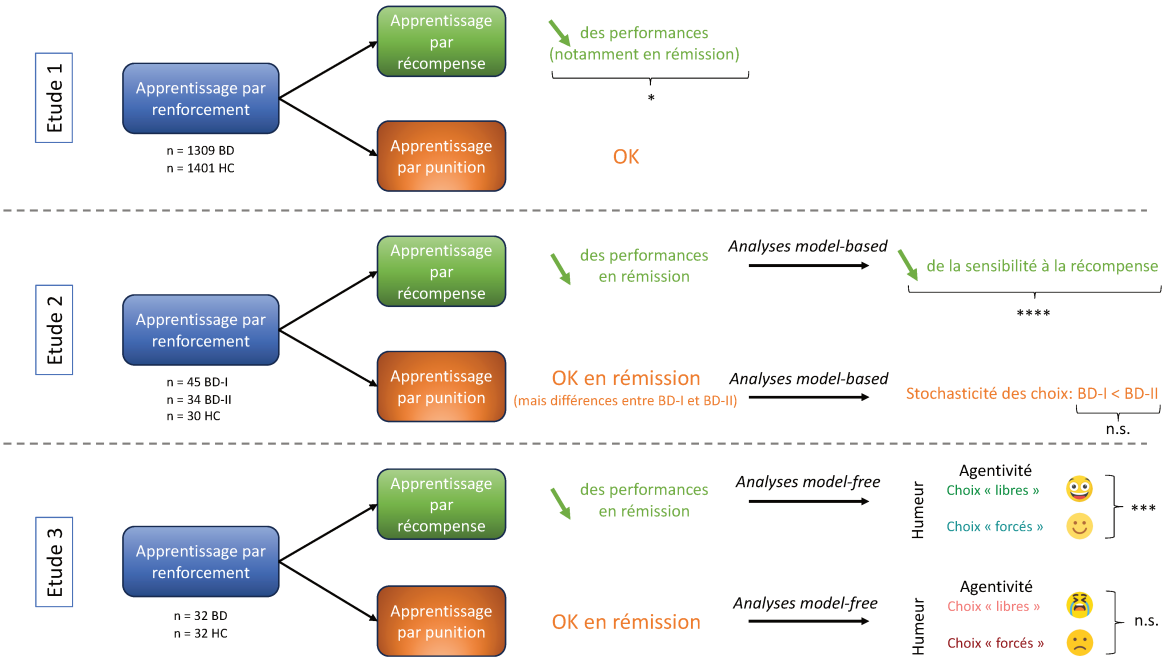

Figure VIII.1 : Résumé des résultats des trois études de ce travail de thèse. Les diagrammes illustrent les principaux résultats de la première (en haut), de la deuxième (au milieu) et de la troisième (en bas) études du travail de doctorat présenté dans ce manuscrit. **Abbreviations**: BD = bipolar disorder; HC = healthy control, n.s. = non significative.

IX

Les modèles des fluctuations de l'humeur dans le trouble bipolaire

1. Le modèle de l'équipe de Robb Rutledge (University College London)

L'équipe de Robb Rutledge a proposé un **modèle des fluctuations pathologiques de l'humeur dans le trouble bipolaire**, partant du principe que **les fluctuations de l'humeur sont fortement liées aux signaux d'erreur de prédiction de la récompense**, qui sont représentés par l'activité dans le striatum ventral (**Figure IX.1**) (Eldar et al., 2016; Mason et al., 2017). Ainsi, les **surprises positives** suscitent une activité striatale et un **état d'esprit positif**. Selon leur modèle, **la perception des récompenses ultérieures s'en trouve alors biaisée**, leur valeur perçue augmentant lorsque l'humeur est élevée (et inversement, leur valeur perçue diminuant lorsque l'humeur est basse). **Les attentes qui guident les décisions futures sont actualisées** sur la base de ces erreurs de prédiction de récompense liées à l'humeur. Un **biais émotionnel** modéré aide les individus à s'adapter rapidement à un environnement qui change, que ce soit pour le meilleur ou pour le pire. Cependant, **si l'humeur biaise fortement la perception de la récompense, comme ils**

le proposent dans le cas du trouble bipolaire, il en résulte une récurrence des erreurs de prédiction de la récompense. Selon leur modèle, le même paramètre de biais émotionnel élevé entraîne une **hypersensibilité aux récompenses lorsque l'humeur est élevée** et une **hyposensibilité aux récompenses lorsque l'humeur est basse**.

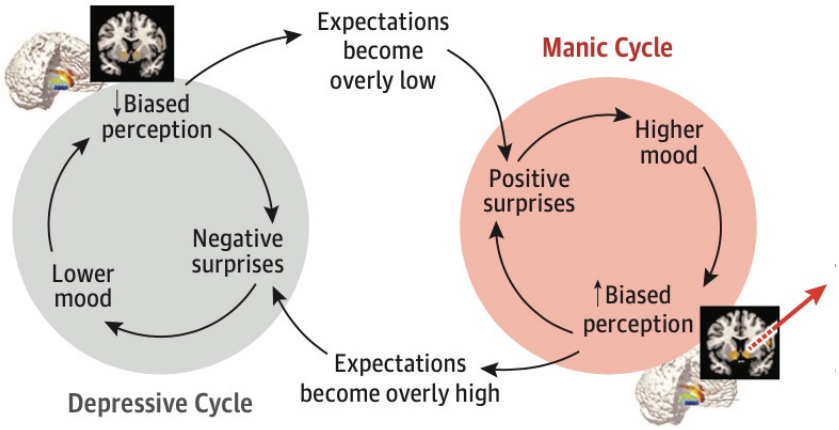

Figure IX.1 : Modèle de l'instabilité de l'humeur proposé par l'équipe de Robb Rutledge. L'humeur augmente à la suite de récompenses inattendues, qui font que les récompenses suivantes sont perçues comme meilleures qu'elles ne le sont, ce qui augmente les attentes futures et accroît encore l'humeur (cycle maniaque). Au fil du temps, les attentes deviennent de plus en plus éloignées des résultats réels, et les surprises négatives qui en résultent déclenchent un cycle dépressif. *Figure issue de (Mason et al., 2017).*

Il n'est en revanche pas pris en compte la **période d'euthymie** et **l'hyposensibilté à la récompense** (ou hypersensibilité selon les théories) basale durant cette période, qui pourrait être rajoutée dans ce modèle et le complexifier. En effet, ce modèle décrit bien les processus impliqués dans le « cycle » de la dépression et celui de l'(hypo)manie, ainsi que le changement d'une humeur pathologique à une autre, **mais peu la période intercritique**. Nous avons vu qu'il n'était pas si simple de parler d'un effet rémanent de la polarité d'un épisode thymique sur les choix ultérieurs, les travaux de Linke et al. (2011) n'ayant pas été répliqués (Linke et al., 2011). Une **hyposensibilité à la récompense** durant l'euthymie, comme nous le supposons, pourrait moduler les attentes, les erreurs de prédiction, l'apprentissage, et les comportements qui en découlent.

2. Le modèle de Robin Nusslock (Northwestern University) et Rolland B. Alloy (Temple University)

L'équipe de Robin Nusslock et Rolland B. Alloy ont également proposé un modèle basé sur la **sensibilité à la récompense**, en faisant le parallèle avec le trouble dépressif majeur (Alloy et al., 2016). Comme nous l'avons expliqué dans la partie introductive de ce manuscrit de thèse, la sensibilité à la récompense dans le trouble bipolaire est actuellement débattue. Selon eux, il y aurait une **distinction entre le trouble bipolaire et le trouble dépressif majeur**. En effet, le premier serait initialement caractérisé par une hypersensibilité à la récompense, l'autre par une hyposensibilité à la récompense. Dans les deux cas, cela conduirait à une **anticipation et une réponse atténuée aux récompenses suite à des évènements** négatifs lors d'un état prémorbide de dépression. Ainsi, les sujets risqueraient de développer ou de maintenir une dépression, ayant à la fois une **capacité réduite à rechercher des récompenses et à y réagir** (la vulnérabilité à la dépression se traduisant par une moindre réactivité et une moindre recherche de récompenses) (**Figure IX.2**). Selon eux, la distinction entre le trouble bipolaire et le trouble dépressif majeur tient au fait que dans le trouble bipolaire, **en réponse à la non-réalisation des récompenses ou des objectifs, cette hypersensibilité conduirait à une désactivation excessive ou à une régulation à la baisse de la motivation d'approche et de l'affect**, ce qui, à son tour, conduit à des symptômes dépressifs, en particulier une faible motivation, un ralentissement psychomoteur, une anhédonie, de la fatigue et du désespoir.

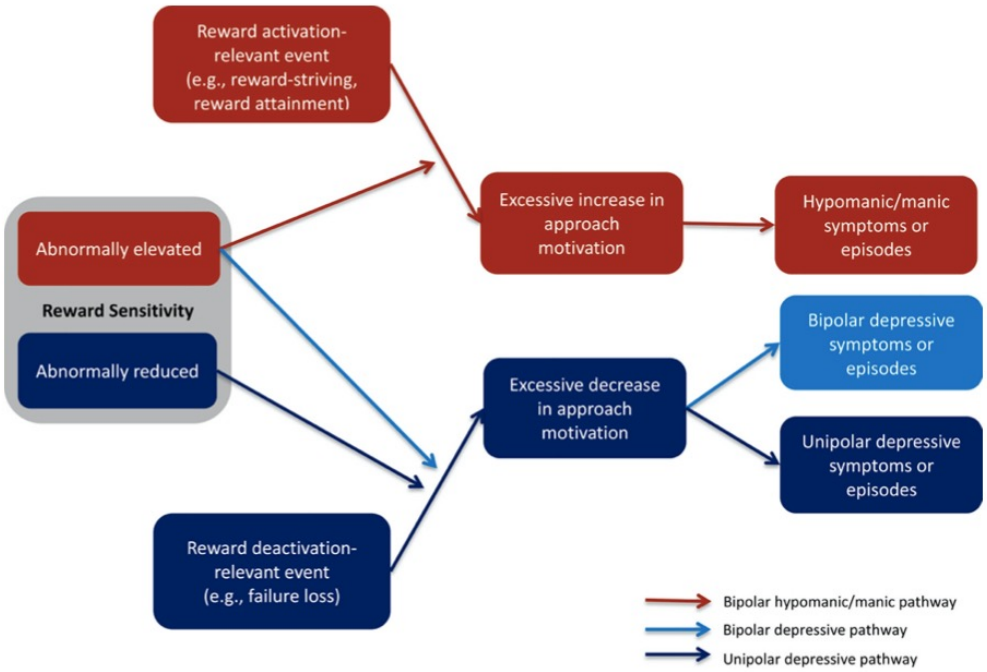

Figure IX.2 : Modèle de cheminement vers un épisode maniaque ou dépressif dans le trouble bipolaire, et d'une dépression dans le trouble dépressif majeur. *Figure issue de (Alloy et al., 2016).*

Ce modèle prend donc en compte la **sensibilité à la récompense** comme facteur pouvant précipiter une rechute thymique, mais se base sur une hypersensibilté à la récompense durant la rémission, et **peine à expliquer clairement comment une hypersensibilté (dans le cas du trouble bipolaire) ou une hyposensibilité (dans le cadre du trouble dépressif majeur) à la récompense conduirait tous les deux à un épisode dépressif**. Nous proposons ainsi un modèle de l'émergence des épisodes thymiques dans le trouble bipolaire en partant de notre théorie d'une hyposensibilté à la récompense lors de la rémission dans le trouble bipolaire.

X

Proposition d'un modèle des fluctuations de l'humeur dans le trouble bipolaire

1. L'hyposensibilité à la récompense comme possible marqueur-trait du trouble bipolaire

Nous pourrions ainsi théoriser les résultats de ces trois études sur une analyse en plusieurs niveaux. Il semblerait que lors de la période de rémission, les patients ayant un trouble bipolaire gardent une altération de l'apprentissage en lien avec la récompense mais pas de celui en lien avec la punition. Sur le plan comportemental, cela se traduirait par le fait que lors des choix qu'ils font dans la vie de tous les jours, **ils apprennent bien des conséquences de leurs actions lorsque ces actions mènent à un évitement de punitions** (de façon égale aux sujets sains), mais ont **des difficultés à apprendre des conséquences de leurs actions lorsque ces actions mènent à un résultat positif**. Cette asymétrie semble s'expliquer par une **hyposensibilité à la récompense**, qui pourrait être en lien avec un hypofonctionnement de certaines structures cérébrales, notamment le **striatum ventral** lors de l'anticipation d'une récompense (Diekhof et al., 2012; Nusslock et al., 2012; Yip et al., 2015). Pour ce travail de thèse, nous nous efforçons d'essayer de

trouver de potentiels candidats pouvant se montrer intéressant à étudier comme biomarqueurs, dans le but de **mieux comprendre l'instabilité thymique dans le trouble bipolaire et les facteurs précipitant les rechutes thymiques**. Ceci pourrait avoir un potentiel d'application clinique dans la prévention des rechutes, par le biais de traitements pharmacologiques et non pharmacologiques (Eldar et al., 2016; Mason et al., 2017).

2. L'erreur de prédiction accumulée comme facteur précipitant des épisodes thymiques

Ainsi, selon notre hypothèse, les patients souffrant de trouble bipolaire **garderaient une hyposensibilité à la récompense durant la période de rémission,** pouvant intervenir dans les rechutes thymiques. De ce fait, par rapport aux sujets sains, ils auraient donc **moins tendance à avoir des comportements orientés vers un objectif gratifiant** durant cette période, et une **plus grande attention ou sensibilité aux informations négatives**, qui peut être à l'origine d'une **perception biaisée des résultats lors des récompenses** et conduire progressivement à des états dépressifs puis des **épisodes dépressifs caractérisés** (Huys et al., 2015; Korn et al., 2014). Du fait d'une altération sélective de l'apprentissage par récompense par rapport à l'apprentissage par évitement de la punition, **ils apprennent moins bien d'un comportement lorsqu'il est suivi de la délivrance d'une récompense**, que lorsqu'il est suivi d'une punition. Comme nous l'avons vu dans la première partie de ce manuscrit de thèse, **ils ont donc moins tendance à répéter ces comportements gratifiants à l'avenir**. Ils seront donc capables d'éviter correctement de reproduire les comportements délivrant un feedback désagréable, **mais pas de reproduire correctement (en tout cas moins) les comportements délivrant un feedback appétitif**. Les affects vont donc se modifier, avec une **diminution du plaisir et l'apparition d'une anhédonie motivationnelle**, conduisant progressivement à une baisse de l'humeur (McLauchlan et al., 2022; Vandendriessche et al., 2023). Nous avons vu qu'il existait un lien avec l'historique des feedbacks et l'humeur (Eldar et al., 2016). Avec le temps et la multiplication des prises de décision, comme les comportements à la recherche de récompense vont se faire plus rare, **l'humeur va progressivement se dégrader** jusqu'à conduire à un **épisode dépressif caractérisé** (Eldar et al., 2016; Huys et al., 2015; Korn et

al., 2014; Mason et al., 2017). En revanche, l'historique des feedbacks positifs sera de moins en moins fréquent et va conduire à une **erreur de prédiction positive de plus en plus importante** dans le cas d'une récompense perçue, d'autant plus s'il s'agit d'un comportement avec un **fort sens de l'agentivité**. Au bout d'un moment, cette « surprise » de la récompense générant une **erreur de prédiction positive très importante pourrait provoquer l'apparition d'un épisode hypomaniaque ou maniaque**. La Figure X.1 propose un modèle de cette théorie.

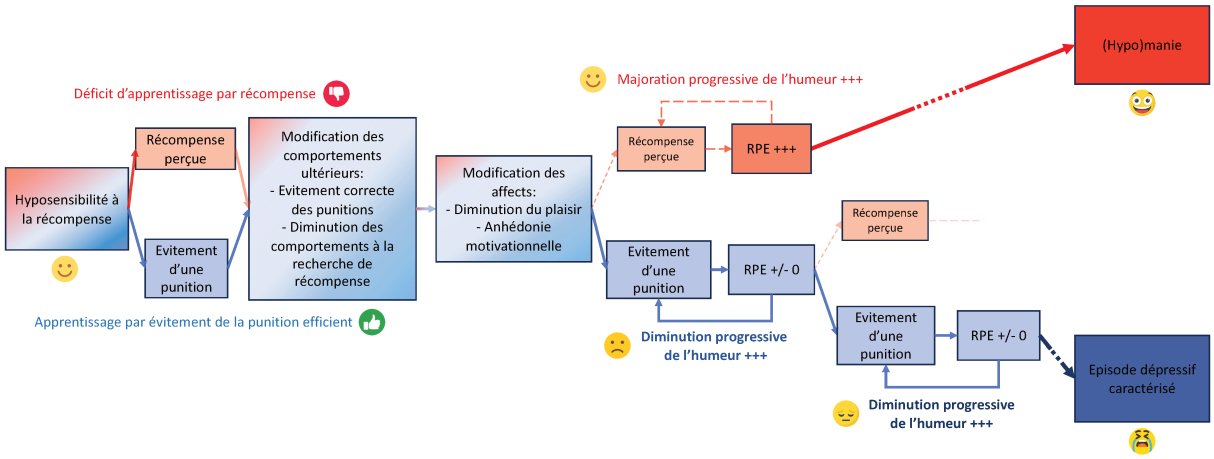

Figure X.1 : Modélisation de l'émergence d'épisode dépressif caractérisé et d'épisode (hypo)maniaque dans le trouble bipolaire. Les diagrammes illustrent les processus intervenant dans la modification progressive de l'humeur jusqu'à des niveaux pathologiques au fur et à mesure des choix faits par le patient atteint de trouble bipolaire, du fait d'un déficit sélectif d'apprentissage par récompense en lien avec une sensibilité émoussée à la récompense. **Abréviations** : RPE = reward prediction error.

3. L'agentivité comme possible modulateur de la sensibilité à la récompense

Grace aux analyses *model-free* de notre troisième étude, nous avons vu que nous sommes **davantage « heureux » lorsque la récompense vient de notre propre fait** plutôt qu'une récompense « imposée » (qui vient sans effort, ce qui pourrait être étudiée dans une autre étude), alors que nous ne sommes pas davantage tristes ou frustrés lorsqu'une punition est de notre fait ou non. Bien que les analyses *model-based* soient encore en cours, nous pouvons

tenter d'intégrer nos résultats dans notre modèle. Lors de la réception d'une récompense, **pour une même erreur de prédiction positive (ou « surprise » positive), l'humeur qui en découle serait plus importante lorsque la récompense survient suite à un comportement libre plutôt qu'un comportement imposé**. Bien qu'il n'y ait pas de différence entre les sujets sains et ceux ayant un trouble bipolaire lors de l'analyse *model-free*, nous pourrions nous demander s'il pourrait exister une interaction entre **l'agentivité et la sensibilité à la récompense ou la valeur attendue**. Les analyses ***model-based*** en cours pourraient nous aider à répondre à cette question. De plus, du fait de la différence entre les BD-I et les BD-II évoquée en introduction, nous pourrions nous demander si **les BD-I ne feraient pas davantage de choix « libres » que de choix « forcés » par rapport aux BD-II**.

4. La stochasticité des choix comme possible marqueur de distinction entre le BD-I et BD-II

En effet, comme évoqué en introduction, certains auteurs émettent l'hypothèse d'un **continuum entre le BD-I et le BD-II** sur la dimension hédonique. Dans ce travail de thèse, nous n'avons pas trouvé de différence comportementale en apprentissage par récompense entre les deux sous-types. Mais bien que cela ne soit pas significatif sur nos analyses, **une tendance se dessine sur une moins bonne performance chez les BD-II par rapport aux BD-I lors de l'apprentissage par punition**. **La stochasticité des choix** semble être la raison de ce déficit d'apprentissage par punition chez les BD-II. Il se pourrait ainsi que **les BD-II font davantage de choix erratiques** dans la vie de tous les jours que les BD-I et les sujets sains, pouvant contribuer à la **moins bonne stabilité clinique entre les épisodes** (Grande et al., 2016). Ceci pourrait expliquer en partie la différence de **trajectoire évolutive** du BD-I (plus stable dans le temps, des épisodes plus importants) et du BD-II (plus de comorbididés, plus d'épisodes mais moins intenses). Ainsi, l'altération en apprentissage par récompense pourrait intervenir dans l'émergence d'un nouvel épisode thymique dans les deux sous-types, mais **la trop grande stochasticité des choix chez les BD-II pourrait également être un mécanisme de moins bonne stabilité lors de la rémission**, voire également générer davantage d'épisodes thymiques.

XI

Perspectives

Comme évoqué dans la partie introductive, ce travail de thèse s'inscrit dans le champ de recherche de la psychiatrie computationnelle. De manière générale, nous avons vu que l'aide de la **modélisation computationnelle** de fonctions cognitives en psychiatrie pourrait nous aider à la fois à **mieux comprendre certains processus cognitifs dysfonctionnels dans l'étiopathogénie ou la physiopathologie d'un trouble, ou encore les bases neurales rattachées à ces dysfonctionnements** (Chen & Takahashi, 2017). Au-delà de la compréhension de la physiopathologie, cela ouvre alors des pistes prometteuses pour la **personnalisation de la thérapeutique**. Dans les troubles de l'humeur et en psychiatrie de manière plus générale, la prise en charge se fait en première ligne par **une intervention pharmacologique et en psychothérapie** (Grande et al., 2016). Les secondes lignes de prise en charge concernent souvent des thérapeutiques telles **que la stimulation magnétique transcrânienne (rTMS), la stimulation transcrânienne par courant continu (tDCS), la Kétamine ou Eskétamine** (Lefaucheur et al., 2017; McClintock et al., 2018; Milev et al., 2016; Yatham et al., 2018). A des stades plus avancés ou résistants de la pathologie, des thérapies plus invasives et souvent encore à l'étude viennent potentiellement compléter l'arsenal thérapeutique, telles que **la stimulation cérébrale profonde (DBS)**

(Figee et al., 2022; Mutz, 2023). C'est dans cet ordre que nous allons voir le potentiel champ d'application de la psychiatrie computationnelle dans la prise en charge des patients.

1. Les sciences computationnelles pour personnaliser la psychothérapie

En effet, partant du principe que certaines des variables décrites par ces modèles semblent refléter l'activité neuronale de régions cérébrales spécifiques, et que ces mêmes paramètres peuvent offrir une **description unique et personnalisée des symptômes d'un patient**, certains auteurs ont par exemple proposé d'utiliser la **modélisation computationnelle pour guider une psychothérapie cognitive et comportementale (TCC)** (Nair et al., 2020). Ils proposent ainsi d'associer une évaluation médicale à une batterie de test comportementaux comprenant par exemple des tâches d'apprentissage par renforcement ou de coût de l'effort pour capturer les paramètres dysfonctionnels propres au patient. Ces paramètres pourront spécifiquement être pris en charge en TCC. L'intérêt de la modélisation computationnelle est d'utiliser la puissance des mathématiques pour **quantifier et suivre les résultats de la thérapie, dans le but de prévenir les rechutes** (**Figure XI.1**).

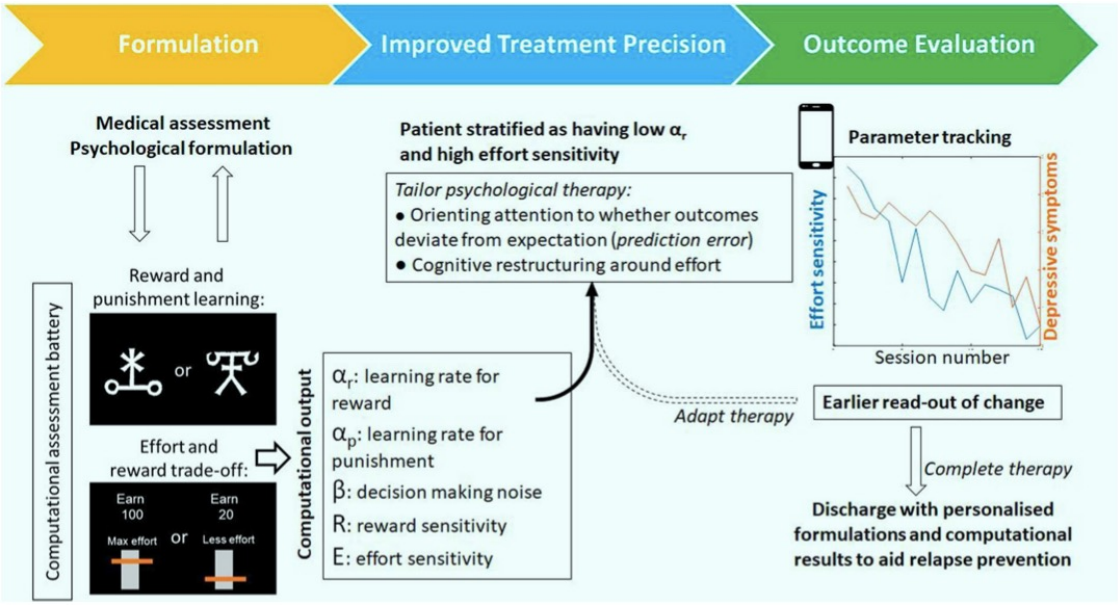

Figure XI.1 : Exemple de parcours d'un patient dans le cadre d'une psychothérapie intégrant des évaluations computationnelles. Parallèlement à l'évaluation médicale, les évaluations basées sur les tâches expérimentales produisent des « paramètres » individuels propres au patient pour une série de processus computationnels, tels que l'apprentissage par récompense ou le coût de l'effort. À partir de cette analyse, un profil computationnel personnalisé pourrait être utilisé pour adapter les composantes de la thérapie (par exemple, l'attention et l'apprentissage pendant l'activation comportementale par rapport à la restructuration cognitive des croyances relatives à l'effort). Ces paramètres computationnels fournissent des marqueurs objectifs pour évaluer le changement et, du fait qu'ils soient plus proches des mécanismes sous-jacents qui génèrent ou maintiennent les symptômes, ils peuvent fournir une lecture plus précoce du changement qui précède l'amélioration des symptômes. Les tâches existent sous des formes adaptées à une utilisation en ligne ou sur des appareils intelligents qui peuvent également améliorer l'accès et l'engagement des patients. *Figure issue de (Nair et al., 2020).*

Nous pourrions ainsi aisément imaginer appliquer nos travaux directement avec ce type de prise en charge dans le futur. En effet, nous pourrions imaginer une telle recherche en **adaptant la tâche utilisée dans notre première étude pour une utilisation sur smartphone**, couplée à d'autres tâches utilisées dans notre équipe ou parmi nos collaborateurs sur le coût de l'effort (Cecchi et al., 2022; Pessiglione, Le Bouc, et al., 2018; Vinckier et al., 2018), dans le but d'avoir une vision plus globale sur les symptômes de la dépression ou de l'(hypo)manie. Nous pourrions effectuer une **analyse initiale en *model-based* de façon personnalisée par patient** afin d'en retirer les paramètres « dysfonctionnels », puis **centrer la psychothérapie TCC sur ces paramètres** (par exemple réduction de la sensibilité à la récompense). Il faudrait cependant imaginer que les

psychothérapeutes soient formés à la prise en charge de ces « paramètres » en TCC. Les patients referaient les mêmes tâches **chaque semaine** (en changeant les symboles), tout en analysant les données avec le même modèle qu'initialement, dans le but de monitorer l'évolution et la corréler avec l'évaluation clinique. Il pourrait même être envisagé de l'associer à des algorithmes de ***machine learning*** comme le proposent certains auteurs, dans le but de faire de la réduction de dimensionnalité (**Figure XI.2**) (Huys et al., 2016). Ainsi, l'algorithme de *machine learning* pourrait apprendre de lui-même à détecter si le patient améliore ou non ses « variables », et tenter de **prédire les rechutes** en les corrélant avec l'évaluation clinique.

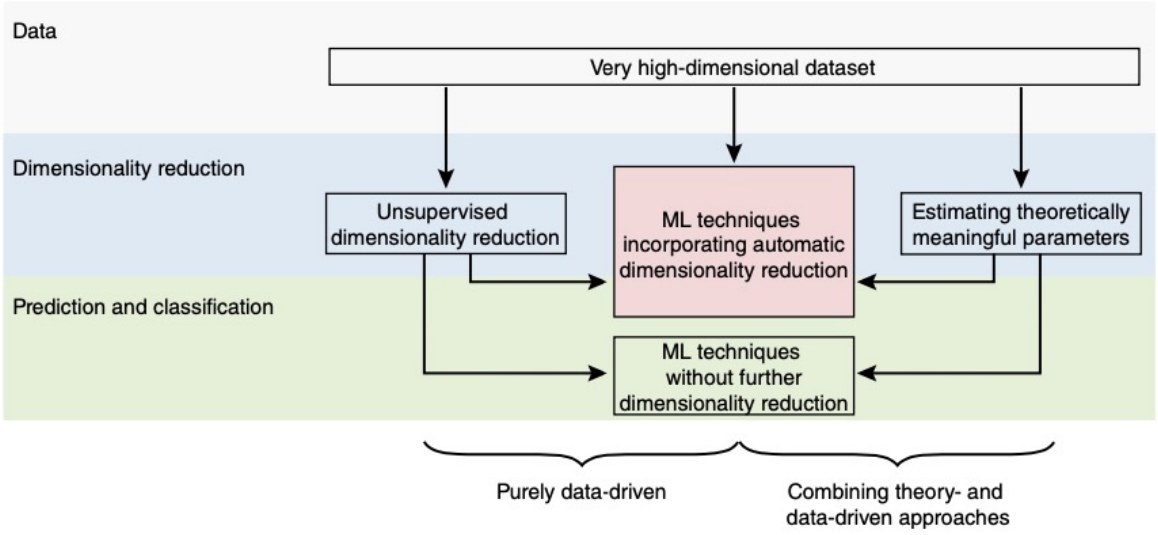

Figure XI.2 : Exploiter et gérer la haute dimensionnalité des ensembles de données psychiatriques. Les approches purement axées sur les données (branches de gauche et du milieu) et les combinaisons d'approches théoriques et axées sur les données (branche de droite) peuvent être utilisées pour analyser de grands ensembles de données afin de parvenir à des applications cliniquement utiles. La réduction de la dimensionnalité est une étape clé pour éviter l'ajustement excessif. Elle peut être réalisée en tant qu'étape de prétraitement à l'aide de méthodes non supervisées avant l'application d'autres techniques de *machine learning* (ML) avec ou sans réduction supplémentaire de la dimensionnalité (branche gauche) ; à l'aide de techniques de ML qui limitent automatiquement le nombre de variables pour la prédiction ; à l'aide de la régularisation ou de la sélection de modèles bayésiens (branche centrale) ; ou à l'aide de modèles guidés par la théorie qui, par essence, projettent les données originales à haute dimension dans un espace à faible dimension de paramètres théoriquement significatifs, qui peuvent ensuite être introduits dans des algorithmes de ML qui peuvent ou non réduire davantage la dimensionnalité (branche de droite). *Figure issue de (Huys et al., 2016).*

2. Les sciences computationnelles pour personnaliser la pharmacologie

Plusieurs auteurs ont proposé de **personnaliser le choix des traitements pharmacologiques** selon le profil pharmacologique du traitement et son affinité pour certains récepteurs, en tentant de chercher des facteurs prédictifs de réponse, du moins théorique, en utilisant les sciences computationnelles, notamment dans les troubles de l'humeur (Rutledge & Adams, 2017). En effet, la modélisation computationnelle pourrait permettre de mieux comprendre chez un patient les mécanismes sous-jacents impliqués dans un symptôme et/ou un processus cognitif. Il est par exemple proposé que pour un même symptôme (ou syndrome), telle que **l'apathie**, la modélisation computationnelle pourrait permettre une compréhension plus fine des mécanismes sous-jacents impliqués (Pessiglione, Vinckier, et al., 2018). Des études ont montré que le compromis entre l'effort et la récompense implique des systèmes corticaux et sous-corticaux spécifiques (**Figure XI.3 A**) (Pessiglione, Vinckier, et al., 2018). **Dans l'idéal, il y aurait une correspondance univoque entre des composants neuronaux spécifiques et des variables du modèle computationnel**. Ainsi, l'adaptation de modèles computationnels au comportement des patients permettrait de déduire le mécanisme dysfonctionnel en termes cognitifs (par exemple, l'hyposensibilité à la récompense) et neuronaux (par exemple, le manque de dopamine au niveau de la connectivité striatale et du vmPFC). Cette approche computationnelle peut donc non seulement donner un aperçu du déficit de motivation, mais aussi aider à personnaliser le traitement (**Figure XI.3 B**) (Pessiglione, Le Bouc, et al., 2018). L'identification de la variable dysfonctionnelle peut informer sur l'intervention thérapeutique, **à condition que les effets computationnels des traitements potentiels soient connus** (ce qui n'est manifestement pas le cas à l'heure actuelle).

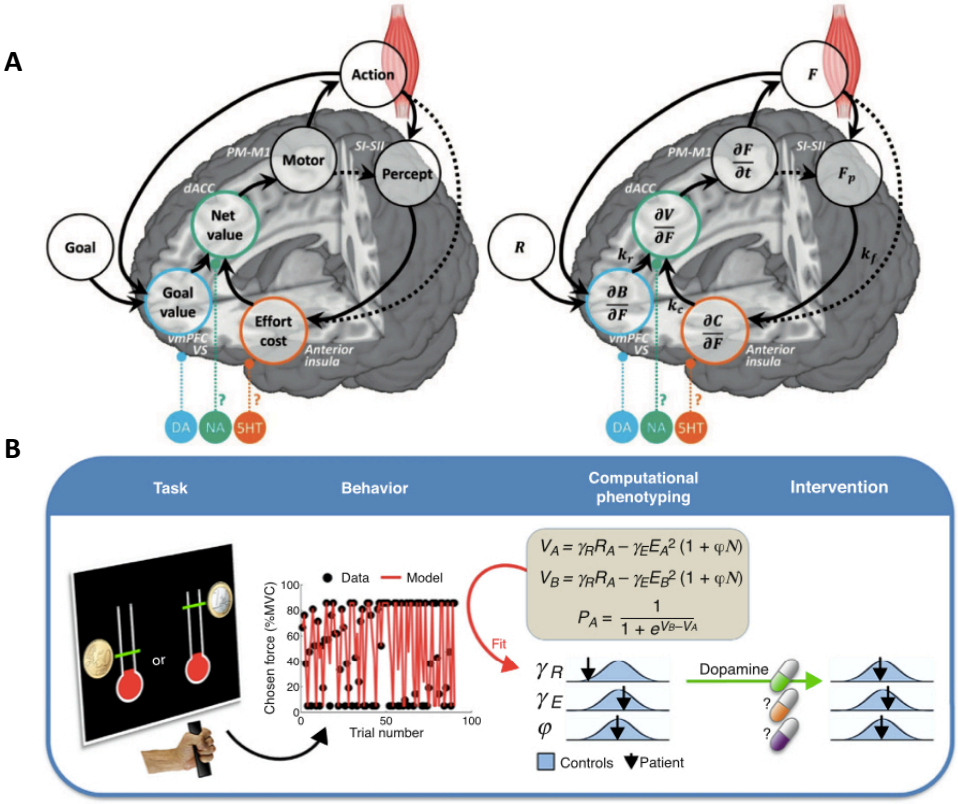

Figure XI.3 : Exemple d'application théorique du phénotypage computationnel dans le traitement de l'apathie. A) La figure illustre les mécanismes que le cerveau pourrait mettre en œuvre pour générer un comportement dans une tâche de motivation. A gauche sont représentés les emplacements anatomiques plausibles pour la représentation des variables de calcul impliquées dans le modèle du coût de l'effort. A droite, il est présenté une dérivation mécaniste du comportement optimal dans un réseau supposé à l'échelle du cerveau. **B) Illustration schématique de l'approche du phénotypage computationnel.** La première étape consiste à recueillir les performances dans une tâche comportementale, ici une tâche d'actualisation de l'effort. Ensuite, le modèle computationnel est ajusté au comportement de choix, en ajustant les paramètres libres de sorte que la probabilité de choix générée par le modèle (ligne) corresponde le plus possible aux choix observés (cercles). L'ensemble des paramètres ajustés définit un phénotype computationnel, qui peut être comparé à la distribution des phénotypes observés dans la population normale. Le paramètre déviant dans cet exemple est la sensibilité à la récompense. Le candidat le plus prometteur serait un traitement à effet pro-dopaminergique, puisqu'il a été démontré qu'il améliorait la sensibilité à la récompense dans un certain nombre de conditions ainsi que chez des sujets sains. La dernière étape consiste à tester à nouveau le patient et à réajuster le modèle, afin de vérifier que le traitement a eu l'effet de calcul escompté, en faisant passer le paramètre déviant dans la plage normale. 5HT = sérotonine ; B = bénéfice attendu ; C = coût attendu ; DA = dopamine ; dACC = cortex cingulaire antérieur dorsal ; F = force produite ; Fp = force perçue ; M1 = cortex moteur primaire ; NA = noradrénaline ; PM = cortex prémoteur ; R = niveau de récompense ; SI-II = cortex somatosensoriel primaire et secondaire ; V = valeur nette ; vmPFC = cortex préfrontal ventromédian ; VS = striatum ventral. *Figure adaptée de (Pessiglione, Vinckier, et al., 2018) et (Pessiglione, Le Bouc, et al., 2018).*

De même que pour le guidage de l'intervention en psychothérapie, le guidage de l'intervention pharmacologique **nécessite encore des travaux de recherche pour que le phénotypage computationnel soit utilisable en pratique clinique**. Concernant une potentielle hyposensibilité à la récompense chez un patient, il pourrait être utilisé des traitements antidépresseurs ayant un **effet pro-dopaminergique** dans le cadre d'une dépression unipolaire, la dopamine étant en lien avec la sensibilité à la récompense (Le Bouc et al., 2016; Zénon et al., 2016), ou encore des traitements par Pramipexole, Aripiprazole ou Cariprazine (non disponible en France) pour la dépression bipolaire, en évitant les traitements ayant un effet anti-dopaminergique (Azorin & Simon, 2019). Ces extrapolations sont en revanche moins claires pour les autres variables (Pessiglione, Le Bouc, et al., 2018).

3. Les sciences computationnelles pour personnaliser la neuromodulation

Dans l'idéal où il y aurait bien une **correspondance univoque entre des composants neuronaux spécifiques et des variables du modèle computationnel**, celui-ci pourrait nous aider à choisir une **cible spécifique** à un patient pour de la neuromodulation non invasive par exemple. En effet, nous savons que les techniques de neuromodulation non invasive telles que la rTMS ou la stimulation électrique transcrânienne (tES), surtout actuellement étudiée et utilisée en stimulation à courant continu (tDCS) pour cette dernière, présentent des résultats bien établis dans la dépression (Lefaucheur et al., 2017; McClintock et al., 2018), nous savons également qu'il existe une **importante variabilité de réponse inter-sujets**, suggérant l'exploration de nouvelles cibles que le cortex préfrontal dorsolatéral (dlPFC) (Downar & Daskalakis, 2013). Ainsi, nous pourrions imaginer un phénotypage computationnel pour les patients bénéficiant de rTMS ou tDCS **venant suggérer une cible anatomo-fonctionnelle à moduler selon les corrélats entre les bases neurales et les paramètres du modèle**. Ainsi, il apparaitrait par exemple intéressant de cibler le **cortex préfrontal ventro-médian (vmPFC) chez les patients déprimés ayant une hyposensibilité à la récompense** (Kearney-Ramos et al., 2018; Winker et al., 2020).

Par ailleurs, nous savons que les techniques invasives telles que l'électroencéphalographie stéréotaxique et la stimulation cérébrale profonde (DBS) permettent de mesurer avec précision et de moduler de manière causale l'activité neurophysiologique du cerveau. Ainsi, certains auteurs ont proposé une approche innovante couplant ces mesures électrophysiologiques invasives et la psychiatrie computationnelle, appelant cela la **psychiatrie computationnelle invasive** (Saez & Gu, 2023). L'intérêt de cette approche est de faire progresser notre compréhension mécaniste des calculs neuronaux des états mentaux en fournissant une **description spatio-temporelle précise de l'activité neuronale**, ce qui est traditionnellement impossible à réaliser à l'aide de techniques non invasives sur des sujets humains. De plus, elle offre un moyen direct et immédiat de **moduler les états cérébraux par la stimulation de régions et de circuits neuronaux définis de manière computationnelle**, fournissant ainsi des informations à la fois causales et thérapeutiques. La plupart des travaux cliniques en cours pour la DBS dans la dépression se concentrent encore sur le ciblage anatomique d'une seule région, qui pourrait bientôt être remplacé par un **ciblage individualisé à partir des résultats obtenus au niveau des circuits** (Saez & Gu, 2023). Des approches complémentaires utilisent des méthodes convergentes de modélisation et de stimulation pour exploiter les interventions de **stéréo-encéphalographie (sEEG)** de l'épilepsie afin de développer des approches de stimulation personnalisées pour le traitement de la dépression (**Figure XI.4**) (Sani et al., 2018; Scangos et al., 2021; Sheth et al., 2022).

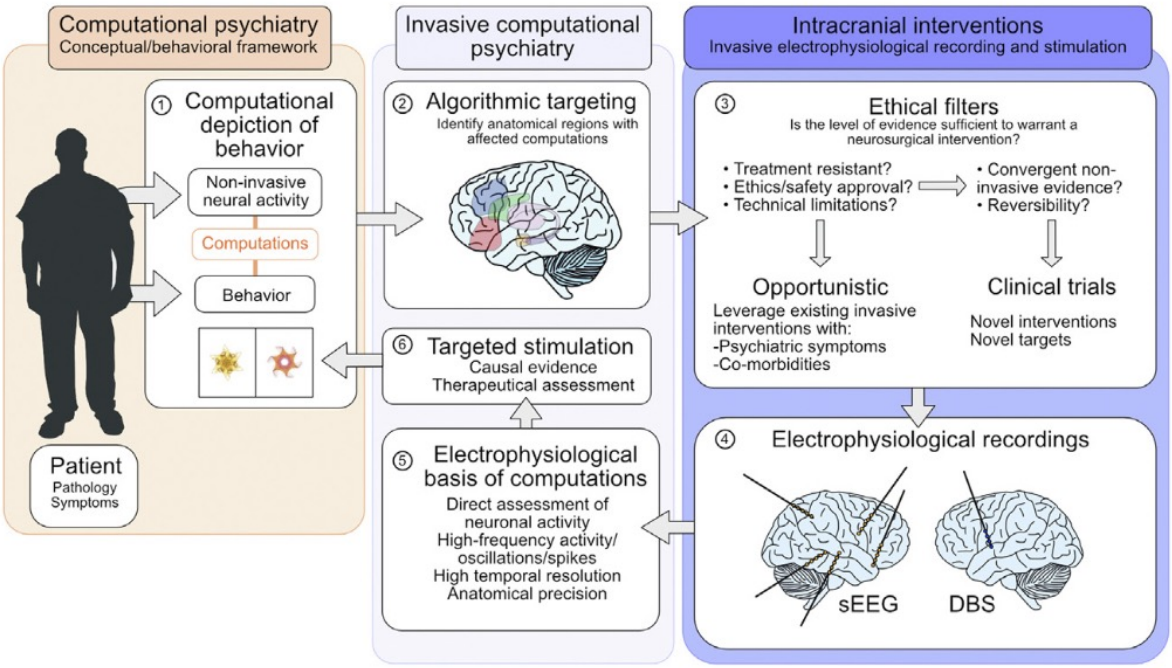

Figure XI.4 : Opportunités pour la psychiatrie computationnelle invasive. La psychiatrie computationnelle fournit des modèles quantitatifs qui décrivent le comportement et spécifient les calculs sous-jacents (1) et permet d'identifier les régions et les calculs affectés par l'état de la maladie (2). La combinaison avec les approches intracrâniennes existantes (3 et 4) ouvre la voie à des descriptions plus détaillées de la base neurophysiologique du comportement (c'est-à-dire l'activité à travers les bandes de fréquence, la résolution temporelle fine) et de ses calculs associés (5) et au développement de paradigmes de neurostimulation ciblée sur l'anatomie (6). Actuellement, cette approche méthodologique peut être mise en œuvre de manière opportuniste en tirant parti de l'existence de comorbidités psychiatriques chez les patients qui ne subissent pas d'interventions neurochirurgicales ou en développant des interventions ad hoc s'il existe des preuves suffisantes pour justifier de nouveaux essais cliniques invasifs. DBS, stimulation cérébrale profonde ; sEEG, électroencéphalographie stéréotaxique. *Figure issue de (Saez & Gu, 2023).*

XII

Conclusion

Nous avons vu tout au long de ce manuscrit de thèse l'intérêt qu'il y avait à s'intéresser à **l'apprentissage par renforcement, le traitement de la récompense et la motivation dans les troubles de l'humeur et notamment le trouble bipolaire**. L'approche par **psychologie expérimentale** est connue depuis plusieurs décennies, souvent couplée à de **l'imagerie fonctionnelle**. Il est toutefois relativement récent de rajouter un niveau d'analyse supplémentaire pour faire le lien entre ces deux niveaux, à savoir la **modélisation computationnelle** basée sur des théories. Cette approche dite de **psychiatrie computationnelle** nous a ainsi aidé à mieux comprendre les altérations de l'apprentissage par renforcement qui pourraient constituer un **marqueur-trait dans le trouble bipolaire**. Au-delà de ce travail de thèse, il serait intéressant de poursuivre ces travaux dans l'objectif d'une application thérapeutique de manière globale, à savoir la combinaison de la psychothérapie et la pharmacologie, et éventuellement de la neuromodulation non-invasive, voire invasive. Ceci nécessiterait toutefois de **poursuivre les travaux en science fondamentale**, ou du moins translationnelle, grâce à l'utilisation concomitante (en plus de la modélisation computationnelle basée sur des théories) de tâches de psychologie expérimentale, imagerie fonctionnelle par IRMf, EEG, sEEG, neuromodulation

pharmacologique, magnétique ou électrique non-invasive, ou encore invasive par DBS ou ultrasons focalisés dans un futur proche. Ceci permettrait notamment de **rattacher des bases neurales et des intervention thérapeutiques spécifiques** (par psychothérapie, pharmacologique, en neuromodulation) aux paramètres libres des différents modèles utilisés. Ceci commence à être le cas pour certains, comme la sensibilité à la récompense mais davantage de travaux sont nécessaires pour appliquer la psychiatrie computationnelle en pratique clinique courante.

Liste des abréviations

5HT : Sérotonine, 86

BAS : *Behavioral Approach System*, 55

BD : *Bipolar disorder*, 71

BD NOS : *BD Not Otherwise Specified*, 32

BD-I : *Bipolar I disorder*, 32, 33, 55, 61, 70, 79

BD-II : *Bipolar II disorder*, 32, 33, 61, 70, 79, 80

CCA : Cortex congulair anterieur, 57

CIM : Classification internationale des maladies, 41

DA : Dopamine, 29, 86

dACC : Cortex congulair anterieur dorsal, 86

DBS: Stimulation cérébrale profonde, 81, 88, 89

dlPFC : Cortex prefrontal dorso-latéral, 87

DSM : Manuel diagnostique et statistique des troubles mentaux, 32, 41

DSM-5 : Manuel diagnostique et statistique des troubles mentaux 5ème édition, 34

EPILT : Explicit Proabilistic Incentive Leanring Task, 54

HC : Sujets sains, 71

IGT : Iowa Gambling Task, 54

ISBD : International society for bipolar disorder, 57

OFC : Cortex orbito-frontal, 57

M1 : Cortex moteur primaire, 86

ML : *Machine learning*, 84

n.s. : Non significatif, 71

NA : Non disponible, 54, 86

PBLT : Probabilistic learning task, 54

PCT : Probabilistic classification task, 54

PD : *Parkinson disease*, 28

PFC : Cortex préfrontal, 57

PM : Cortex prémoteur, 86

PPE : prédiction sur la probabilité de gagner, 48, 49

Liste des Figures

Liste des tables

Références

Abelson, J. L., Khan, S., Liberzon, I., Erickson, T. M., & Young, E. A. (2008). Effects of Perceived Control and Cognitive Coping on Endocrine Stress Responses to Pharmacological Activation. *Biological Psychiatry*, *64*(8), 701-707. https://doi.org/10.1016/j.biopsych.2008.05.007

Abler, B., Greenhouse, I., Ongur, D., Walter, H., & Heckers, S. (2008). Abnormal Reward System Activation in Mania. *Neuropsychopharmacology*, *33*(9), 2217-2227. https://doi.org/10.1038/sj.npp.1301620

Abohamza, E., Weickert, T., Ali, M., & Moustafa, A. A. (2020). Reward and punishment learning in schizophrenia and bipolar disorder. *Behavioural Brain Research*, *381*, 112298. https://doi.org/10.1016/j.bbr.2019.112298

Adida, M., Clark, L., Pomietto, P., Kaladjian, A., Besnier, N., Azorin, J., Jeanningros, R., & Goodwin, G. M. (2008). Lack of insight may predict impaired decision making in manic patients. *Bipolar disorders*, *10*(7), 829-837.

Adida, M., Jollant, F., Clark, L., Besnier, N., Guillaume, S., Kaladjian, A., Mazzola-Pomietto, P., Jeanningros, R., Goodwin, G. M., Azorin, J.-M., & Courtet, P. (2011). Trait-Related Decision-Making Impairment in the Three Phases of Bipolar Disorder. *Biological Psychiatry*, *70*(4), 357-365. https://doi.org/10.1016/j.biopsych.2011.01.018

Adida, M., Jollant, F., Clark, L., Guillaume, S., Goodwin, G. M., Azorin, J.-M., & Courtet, P. (2015). Lithium might be associated with better decision-making performance in euthymic bipolar patients. *European Neuropsychopharmacology*, *25*(6), 788-797. https://doi.org/10.1016/j.euroneuro.2015.03.003

Akbari, V., Rahmatinejad, P., & Mohammadi, S. D. (2019). Comparing Neurocognitive Profile of Patients with Borderline Personality and Bipolar-II Disorders. *Iranian Journal of Psychiatry*, *14*(2), 113-119.

Alloy, L. B., & Abramson, L. Y. (2010). The role of the behavioral approach system (BAS) in bipolar spectrum disorders *Directions in Psychological Scien* 189-194.

Alloy, L. B., Abramson, L. Y., Walshaw, P. D., Cogswell, A., Grandin, L. D., Hughes, M. E., Iacoviello, B. M., Whitehouse, W. G., Urosevic, S., Nusslock, R., & Hogan, M. E. (2008). Behavioral Approach System and Behavioral Inhibition System sensitivities and bipolar spectrum disorders : Prospective prediction of bipolar mood episodes. *Bipolar Disorders*, *10*(2), 310-322. https://doi.org/10.1111/j.1399-5618.2007.00547.x

Alloy, L. B., Bender, R. E., Whitehouse, W. G., Wagner, C. A., Liu, R. T., Grant, D. A., Jager-Hyman, S., Molz, A., Choi, J. Y., & Harmon-Jones, E. (2012). High Behavioral Approach System (BAS) sensitivity, reward responsiveness, and goal-striving predict first onset of bipolar spectrum disorders : A prospective

behavioral high-risk design. *Journal of abnormal psychology*, *121*(2), 339.

Alloy, L. B., & Nusslock, R. (2019). Future Directions for Understanding Adolescent Bipolar Spectrum Disorders : A Reward Hypersensitivity Perspective. *Journal of Clinical Child & Adolescent Psychology*, *48*(4), 669-683. https://doi.org/10.1080/15374416.2019.1567347

Alloy, L. B., Nusslock, R., & Boland, E. M. (2015). The development and course of bipolar spectrum disorders : An integrated reward and circadian rhythm dysregulation model. *Annual review of clinical psychology*, *11*, 213-250.

Alloy, L. B., Olino, T., Freed, R. D., & Nusslock, R. (2016). Role of Reward Sensitivity and Processing in Major Depressive and Bipolar Spectrum Disorders. *Behavior Therapy*, *47*(5), 600-621. https://doi.org/10.1016/j.beth.2016.02.014

APA. (s. d.). *American Psychiatric Association. (2013). Cautionary statement for forensic use of DSM-5. In Diagnostic and statistical manual of mental disorders (5th ed.). Washington, DC: Author. Http://dx.doi.org/10.1176/appi.books.9780890425596 .CautionaryStatement. (N.d.).*

Applegate, E., El-Deredy, W., & Bentall, R. P. (2009). Reward responsiveness in psychosis-prone groups : Hypomania and negative schizotypy. *Personality and Individual Differences*, *47*(5), 452-456.

Ashok, A. H., Marques, T. R., Jauhar, S., Nour, M. M., Goodwin, G. M., Young, A. H., & Howes, O. D. (2017). The dopamine hypothesis of bipolar affective disorder : The state of the art and implications for treatment. *Molecular Psychiatry*, *22*(5), 666-679. https://doi.org/10.1038/mp.2017.16

Azorin, J.-M., & Simon, N. (2019). Dopamine Receptor Partial Agonists for the Treatment of Bipolar Disorder. *Drugs*, *79*(15), 1657-1677. https://doi.org/10.1007/s40265-019-01189-8

Balodis, I. M., & Potenza, M. N. (2015). Anticipatory reward processing in addicted populations : A focus on the monetary incentive delay task. *Biological psychiatry*, *77*(5), 434-444.

Barch, D. M., Carter, C. S., Gold, J. M., Johnson, S. L., Kring, A. M., MacDonald, A. W., Pizzagalli, D. A., Ragland, J. D., Silverstein, S. M., & Strauss, M. E. (2017). Explicit and implicit reinforcement learning across the psychosis spectrum. *Journal of Abnormal Psychology*, *126*(5), 694-711. https://doi.org/10.1037/abn0000259

Bayer, H. M., & Glimcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, *47*(1), 129-141.

Beaumont, S. (2018). *La psychiatrie computationnelle : Vers une nouvelle approche théorique de la psychopathologie.*

Berghorst, L. H., Kumar, P., Greve, D. N., Deckersbach, T., Ongur, D., Dutra, S. J., & Pizzagalli, D. A. (2016). Stress and reward processing in bipolar disorder : A functional magnetic resonance imaging study. *Bipolar disorders*, *18*(7), 602-611.

Bermpohl, F., Kahnt, T., Dalanay, U., Hägele, C., Sajonz, B., Wegner, T., Stoy, M., Adli, M., Krüger, S., Wrase, J., Ströhle, A., Bauer, M., & Heinz, A. (2009). Altered representation of expected value in the orbitofrontal cortex in mania. *Human Brain Mapping*, *31*(7), 958-969. https://doi.org/10.1002/hbm.20909

Berridge, K. C. (2012). From prediction error to incentive salience : Mesolimbic computation of reward motivation. *European Journal of neuroscience*, *35*(7), 1124-1143.

Berridge, K. C., & Kringelbach, M. L. (2015). Pleasure Systems in the Brain. *Neuron*, *86*(3), 646-664. https://doi.org/10.1016/j.neuron.2015.02.018

Berridge, K. C., & Robinson, T. E. (2003). Parsing reward. *Trends in Neurosciences*, *26*(9), 507-513. https://doi.org/10.1016/S0166-2236(03)00233-9

Berridge, K. C., & Robinson, T. E. (2016). Liking, wanting, and the incentive-sensitization theory of addiction. *American Psychologist*, *71*(8), 670-679. https://doi.org/10.1037/amp0000059

Berridge, K. C., Venier, I. L., & Robinson, T. E. (1989). Taste reactivity analysis of 6-hydroxydopamine-induced aphagia : Implications for arousal and anhedonia hypotheses of dopamine function. *Behavioral neuroscience*, *103*(1), 36.

Bjork, J. M., & Hommer, D. W. (2007). Anticipating instrumentally obtained and passively-received rewards : A factorial fMRI investigation. *Behavioural Brain Research*, *177*(1), 165-170. https://doi.org/10.1016/j.bbr.2006.10.034

Blain, B., & Rutledge, R. B. (2020). Momentary subjective well-being depends on learning and not reward. *eLife*, *9*, e57977. https://doi.org/10.7554/eLife.57977

Blaisdell, A. (2008). Cognitive Dimension of Operant Learning. *Learning Theory and Behavior. Learning and Memory: A Comprehensive Reference*, *1*, 173-195. https://doi.org/10.1016/B978-012370509-9.00184-4

Brambilla, P., Perlini, C., Bellani, M., Tomelleri, L., Ferro, A., Cerruti, S., Marinelli, V., Rambaldelli, G., Christodoulou, T., Jogia, J., Dima, D., Tansella, M., Balestrieri, M., & Frangou, S. (2013). Increased salience of gains *versus* decreased associative learning differentiate bipolar disorder from schizophrenia during incentive decision making. *Psychological Medicine*, *43*(3), 571-580. https://doi.org/10.1017/S0033291712001304

Brauer, L. H., & De Wit, H. (1997). High dose pimozide does not block amphetamine-induced euphoria in normal volunteers. *Pharmacology Biochemistry and Behavior*, *56*(2), 265-272.

Bray, S., & O'Doherty, J. (2007). Neural coding of reward-prediction error signals during classical conditioning wit faces. *Journal of neurophys* 3036-3045.

Buehner, M. J., & May, J. (2003). Rethinking Temporal Contiguity and the Judgement of Causality : Effects of Prior Knowledge, Experience, and Reinforcement Procedure. *The Quarterly Journal of Experimental Psychology Section A*, *56*(5), 865-890. https://doi.org/10.1080/02724980244000675

Burdick, K. E., Braga, R. J., Gopin, C. B., & Malhotra, A. K. (2014). Dopaminergic Influences on Emotional Decision Making in Euthymic Bipolar Patients. *Neuropsychopharmacology*, *39*(2), 274-282. https://doi.org/10.1038/npp.2013.177

Caletti, E., Paoli, R. A., Fiorentini, A., Cigliobianco, M., Zugno, E., Serati, M., Orsenigo, G., Grillo, P., Zago, S., Caldiroli, A., Prunas, C., Giusti, F., Consonni, D., & Altamura, A. C. (2013). Neuropsychology, social cognition and global functioning among bipolar, schizophrenic patients and healthy controls : Preliminary data. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00661

Cannon, C. M., & Palmiter, R. D. (2003). Reward without Dopamine. *The Journal of Neuroscience*, *23*(34), 10827-10831. https://doi.org/10.1523/JNEUROSCI.23-34-10827.2003

Caseras, X., Lawrence, N. S., Murphy, K., Wise, R. G., & Phillips, M. L. (2013). Ventral Striatum Activity in Response to Reward : Differences Between Bipolar I and II Disorders. *American Journal of Psychiatry*, *170*(5), 533-541. https://doi.org/10.1176/appi.ajp.2012.12020169

Cecchi, R., Vinckier, F., Hammer, J., Marusic, P., Nica, A., Rheims, S., Trebuchon, A., Barbeau, E. J., Denuelle, M., Maillard, L., Minotti, L., Kahane, P., Pessiglione, M., & Bastin, J. (2022). Intracerebral mechanisms explaining the impact of incidental feedback on mood state and risky choice. *eLife*, *11*, e72440. https://doi.org/10.7554/eLife.72440

Chambon, V., Théro, H., Vidal, M., Vandendriessche, H., Haggard, P., & Palminteri, S. (2020). Information about action outcomes differentially affects learning from self-determined versus imposed choices. *Nature Human Behaviour*, *4*(10), 1067-1079. https://doi.org/10.1038/s41562-020-0919-5

Chase, H. W., Nusslock, R., Almeida, J. R., Forbes, E. E., LaBarbara, E. J., & Phillips, M. L. (2013). Dissociable patterns of abnormal frontal cortical activation during anticipation of an uncertain reward or loss in bipolar versus major depression. *Bipolar Disorders*, *15*(8), 839-854. https://doi.org/10.1111/bdi.12132

Chen, C., & Takahashi, T. (2017). Reward Processing in Depression : The Computational Approach. *Computational Models of Brain and Behavior*, 57-71.

Chou, Y.-H., Wang, S.-J., Lin, C.-L., Mao, W.-C., Lee, S.-M., & Liao, M.-H. (2010). Decreased brain serotonin transporter binding in the euthymic state of bipolar I but not bipolar II disorder : A SPECT study: Brain serotonin transporter in bipolar disorder. *Bipolar Disorders*, *12*(3), 312-318. https://doi.org/10.1111/j.1399-5618.2010.00800.x

Clark, L., Iversen, S. D., & Goodwin, G. M. (2001). A neuropsychological investigation of prefrontal cortex involvement in acute mania. *American Journal of Psychiatry*, *158*(10), 1605-1611.

Collins, A., & Khamassi, M. (2021). *Initiation à la modélisation computationnelle*.

Cullen, B., Ward, J., Graham, N. A., Deary, I. J., Pell, J. P., Smith, D. J., & Evans, J. J. (2016). Prevalence and correlates of cognitive impairment in euthymic adults with bipolar disorder : A systematic review. *Journal of Affective Disorders*, *205*, 165-181. https://doi.org/10.1016/j.jad.2016.06.063

DePasque Swanson, S., & Tricomi, E. (2014). Goals and task difficulty expectations modulate striatal responses to feedback. *Cognitive, Affective, & Behavioral Neuroscience*, *14*, 610-620.

de Wit, S., & Dickinson, A. (2009). Associative theories of goal-directed behaviour : A case for animal–human translational models. *Psychological Research PRPF*, *73*, 463-476.

Dickinson, A. (1980). *Contemporary animal learning theory*. Cambridge University Press. https://books.google.fr/books?id=QnBFzwEACAAJ

Dickinson, A. (1985). Actions and habits : The development of behavioural autonomy. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *308*(1135), 67-78.

Diekhof, E. K., Kaps, L., Falkai, P., & Gruber, O. (2012). The role of the human ventral striatum and the medial orbitofrontal cortex in the representation of reward magnitude–An activation likelihood estimation meta-analysis of neuroimaging studies of passive reward expectancy and outcome processing. *Neuropsychologia*, *50*(7), 1252-1266.

Diflorio, A., & Jones, I. (2010). Is sex important? Gender differences in bipolar disorder. *International Review of Psychiatry*, *22*(5), 437-452. https://doi.org/10.3109/09540261.2010.514601

Downar, J., & Daskalakis, Z. J. (2013). New Targets for rTMS in Depression : A Review of Convergent Evidence. *Brain Stimulation*, *6*(3), 231-240.

https://doi.org/10.1016/j.brs.2012.08.006

Drevets, W. C., Öngür, D., & Price, J. L. (1998). Neuroimaging abnormalities in the subgenual prefrontal cortex : Implications for the pathophysiology of familial mood disorders. *Molecular Psychiatry*, *3*(3), 220-226. https://doi.org/10.1038/sj.mp.4000370

Duek, O., Osher, Y., Belmaker, R. H., Bersudsky, Y., & Kofman, O. (2014). Reward sensitivity and anger in euthymic bipolar disorder. *Psychiatry Research*, *215*(1), 95-100. https://doi.org/10.1016/j.psychres.2013.10.028

Dutra, S. J., Cunningham, W. A., Kober, H., & Gruber, J. (2015). Elevated striatal reactivity across monetary and social rewards in bipolar I disorder. *Journal of Abnormal Psychology*, *124*(4), 890-904. https://doi.org/10.1037/abn0000092

Edge, M. D., Johnson, S. L., Ng, T., & Carver, C. S. (2013). Iowa gambling task performance in euthymic bipolar I disorder : A meta-analysis and empirical study. *Journal of Affective Disorders*, *150*(1), 115-122. https://doi.org/10.1016/j.jad.2012.11.027

Eldar, E., & Niv, Y. (2015). Interaction between emotional state and learning underlies mood instability. *Nature Communications*, *6*(1), 6149. https://doi.org/10.1038/ncomms7149

Eldar, E., Rutledge, R. B., Dolan, R. J., & Niv, Y. (2016). Mood as Representation of Momentum. *Trends in Cognitive Sciences*, *20*(1), 15-24. https://doi.org/10.1016/j.tics.2015.07.010

Etain, B., Mathieu, F., Liquet, S., Raust, A., Cochet, B., Richard, J. R., Gard, S., Zanouy, L., Kahn, J. P., Cohen, R. F., Bougerol, T., Henry, C., Leboyer, M., & Bellivier, F. (2013). Clinical features associated with trait-impulsiveness in euthymic bipolar disorder patients. *Journal of Affective Disorders*, *144*(3), 240-247. https://doi.org/10.1016/j.jad.2012.07.005

Everitt, B. J., Dickinson, A., & Robbins, T. W. (2001). The neuropsychological basis of addictive behaviour. *Brain Research Reviews*, *36*(2-3), 129-138. https://doi.org/10.1016/S0165-0173(01)00088-1

Ferrari, A. J., Baxter, A. J., & Whiteford, H. A. (2011). A systematic review of the global distribution and availability of prevalence data for bipolar disorder. *Journal of Affective Disorders*, *134*(1-3), 1-13. https://doi.org/10.1016/j.jad.2010.11.007

Figee, M., Riva-Posse, P., Choi, K. S., Bederson, L., Mayberg, H. S., & Kopell, B. H. (2022). Deep Brain Stimulation for Depression. *Neurotherapeutics*, *19*(4), 1229-1245. https://doi.org/10.1007/s13311-022-01270-3

Frey, B. N., Andreazza, A. C., Houenou, J., Jamain, S., Goldstein, B. I., Frye, M. A., Leboyer, M., Berk, M., Malhi, G. S., Lopez-Jaramillo, C., Taylor, V. H., Dodd, S., Frangou, S., Hall, G. B., Fernandes, B. S., Kauer-Sant'Anna, M., Yatham, L. N., Kapczinski, F., & Young, L. T. (2013). Biomarkers in bipolar disorder : A positional paper from the International Society for Bipolar Disorders Biomarkers Task Force. *Australian & New Zealand Journal of Psychiatry*, *47*(4), 321-332. https://doi.org/10.1177/0004867413478217

Gallagher, S. (2000). Philosophical conceptions of the self : Implications for cognitive science. *Trends in Cognitive Sciences*, *4*(1), 14-21. https://doi.org/10.1016/S1364-6613(99)01417-5

García-Gutiérrez, M. S., Navarrete, F., Sala, F., Gasparyan, A., Austrich-Olivares, A., & Manzanares, J. (2020). Biomarkers in Psychiatry : Concept, Definition, Types

and Relevance to the Clinical Reality. *Frontiers in Psychiatry*, *11*, 432. https://doi.org/10.3389/fpsyt.2020.00432

Geana, A., Barch, D. M., Gold, J. M., Carter, C. S., MacDonald, A. W., Ragland, J. D., Silverstein, S. M., & Frank, M. J. (2022). Using Computational Modeling to Capture Schizophrenia-Specific Reinforcement Learning Differences and Their Implications on Patient Classification. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *7*(10), 1035-1046. https://doi.org/10.1016/j.bpsc.2021.03.017

Gibbon, J., Baldock, M., Locurto, C., Gold, L., & Terrace, H. (1977). Trial and intertrial durations in autoshaping. *Journal of Experimental psychology: Animal behavior processes*, *3*(3), 264.

Glickman, S. E., & Schiff, B. B. (1967). A biological theory of reinforcement. *Psychological Review*, *74*(2), 81-109. https://doi.org/10.1037/h0024290

Gomide Vasconcelos, A., Sergeant, J., Corrêa, H., Mattos, P., & Malloy-Diniz, L. (2014). When self-report diverges from performance : The usage of BIS-11 along with neuropsychological tests. *Psychiatry Research*, *218*(1-2), 236-243. https://doi.org/10.1016/j.psychres.2014.03.002

Grande, I., Berk, M., Birmaher, B., & Vieta, E. (2016). Bipolar disorder. *The Lancet*, *387*(10027), 1561-1572. https://doi.org/10.1016/S0140-6736(15)00241-X

Greifeneder, R., Scheibehenne, B., & Kleber, N. (2010). Less may be more when choosing is difficult : Choice complexity and too much choice. *Acta Psychologica*, *133*(1), 45-50. https://doi.org/10.1016/j.actpsy.2009.08.005

Grunze, H. (2015). Bipolar Disorder. In *Neurobiology of Brain Disorders* (p. 655-673). Elsevier. https://doi.org/10.1016/B978-0-12-398270-4.00040-9

Gu, Y., Zhou, C., Yang, J., Zhang, Q., Zhu, G., Sun, L., Ge, M., & Wang, Y. (2020). A transdiagnostic comparison of affective decision-making in patients with schizophrenia, major depressive disorder, or bipolar disorder. *PsyCh Journal*, *9*(2), 199-209. https://doi.org/10.1002/pchj.351

Hägele, C., Schlagenhauf, F., Rapp, M., Sterzer, P., Beck, A., Bermpohl, F., Stoy, M., Ströhle, A., Wittchen, H.-U., Dolan, R. J., & Heinz, A. (2015). Dimensional psychiatry : Reward dysfunction and depressive mood across psychiatric disorders. *Psychopharmacology*, *232*(2), 331-341. https://doi.org/10.1007/s00213-014-3662-7

Haidich, A. B. (2010). Meta-analysis in medical research. *Hippokratia*, *14*(Suppl 1), 29-37.

Hamdani, N. (2012). Immuno- inflammatory markers of bipolar disorder a review of evidence. *Frontiers in Bioscience*, *E4*(6), 2170-2182. https://doi.org/10.2741/e534

Hardman, C. A., Herbert, V. M., Brunstrom, J. M., Munafò, M. R., & Rogers, P. J. (2012). Dopamine and food reward : Effects of acute tyrosine/phenylalanine depletion on appetite. *Physiology & behavior*, *105*(5), 1202-1207.

Hassall, C. D., Hajcak, G., & Krigolson, O. E. (2019). The importance of agency in human reward processing. *Cognitive, Affective, & Behavioral Neuroscience*, *19*(6), 1458-1466. https://doi.org/10.3758/s13415-019-00730-2

Henry, C. (2012). Emotional dysfunction as a marker of bipolar disorders. *Frontiers in Bioscience*, *E4*(7), 2622-2630. https://doi.org/10.2741/e578

Hull, C. L. (1943). *Principles of behavior : An introduction to behavior theory*. Appleton-Century Company, Incorporated New York.

Huys, Q. J. M., Daw, N. D., & Dayan, P. (2015). Depression : A Decision-Theoretic Analysis. *Annual Review of Neuroscience*, *38*(1), 1-23. https://doi.org/10.1146/annurev-neuro-071714-033928

Huys, Q. J. M., Maia, T. V., & Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature Neuroscience*, *19*(3), 404-413. https://doi.org/10.1038/nn.4238

Ibanez, A., Cetkovich, M., Petroni, A., Urquina, H., Baez, S., Gonzalez-Gadea, M. L., Kamienkowski, J. E., Torralva, T., Torrente, F., Strejilevich, S., Teitelbaum, J., Hurtado, E., Guex, R., Melloni, M., Lischinsky, A., Sigman, M., & Manes, F. (2012). The Neural Basis of Decision-Making and Reward Processing in Adults with Euthymic Bipolar Disorder or Attention-Deficit/Hyperactivity Disorder (ADHD). *PLoS ONE*, *7*(5), e37306. https://doi.org/10.1371/journal.pone.0037306

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C., & Wang, P. (2010). Research Domain Criteria (RDoC) : Toward a New Classification Framework for Research on Mental Disorders. *American Journal of Psychiatry*, *167*(7), 748-751. https://doi.org/10.1176/appi.ajp.2010.09091379

Jogia, J., Dima, D., Kumari, V., & Frangou, S. (2012). Frontopolar cortical inefficiency may underpin reward and working memory dysfunction in bipolar disorder. *The World Journal of Biological Psychiatry*, *13*(8), 605-615. https://doi.org/10.3109/15622975.2011.585662

Johnson, S. L., Mehta, H., Ketter, T. A., Gotlib, I. H., & Knutson, B. (2019). Neural responses to monetary incentives in bipolar disorder. *NeuroImage: Clinical*, *24*, 102018. https://doi.org/10.1016/j.nicl.2019.102018

Kaiser, J., Buciuman, M., Gigl, S., Gentsch, A., & Schütz-Bosbach, S. (2021). The Interplay Between Affective Processing and Sense of Agency During Action Regulation : A Review. *Frontiers in Psychology*, *12*, 716220. https://doi.org/10.3389/fpsyg.2021.716220

Kaiser, J., & Schütz-Bosbach, S. (2018). Sensory attenuation of self-produced signals does not rely on self-specific motor predictions. *European Journal of Neuroscience*, *47*(11), 1303-1310. https://doi.org/10.1111/ejn.13931

Kaiser, J., & Schütz-Bosbach, S. (2019). Proactive control without midfrontal control signals? The role of midfrontal oscillations in preparatory conflict adjustments. *Biological Psychology*, *148*, 107747. https://doi.org/10.1016/j.biopsycho.2019.107747

Kearney-Ramos, T. E., Dowdle, L. T., Lench, D. H., Mithoefer, O. J., Devries, W. H., George, M. S., Anton, R. F., & Hanlon, C. A. (2018). Transdiagnostic Effects of Ventromedial Prefrontal Cortex Transcranial Magnetic Stimulation on Cue Reactivity. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*(7), 599-609. https://doi.org/10.1016/j.bpsc.2018.03.016

Knutson, B., Fong, G. W., Adams, C. M., Varner, J. L., & Hommer, D. (2001). Dissociation of reward anticipation and outcome with event-related fMRI. *Neuroreport*, *12*(17), 3683-3687.

Kok, P., Rahnev, D., Jehee, J. F. M., Lau, H. C., & de Lange, F. P. (2012). Attention Reverses the Effect of Prediction in Silencing Sensory Signals. *Cerebral Cortex*, *22*(9), 2197-2206. https://doi.org/10.1093/cercor/bhr310

Kollmann, B., Scholz, V., Linke, J., Kirsch, P., & Wessa, M. (2017). Reward anticipation revisited-evidence from an fMRI study in

euthymic bipolar I patients and healthy first-degree relatives. *Journal of affective disorders*, *219*, 178-186.

Korn, C. W., Sharot, T., Walter, H., Heekeren, H. R., & Dolan, R. J. (2014). Depression is related to an absence of optimistically biased belief updating about future life events. *Psychological Medicine*, *44*(3), 579-592. https://doi.org/10.1017/S0033291713001074

Kringelbach, M. L., & Berridge, K. C. (2010). The functional neuroanatomy of pleasure and happiness. *Discovery medicine*, *9*(49), 579.

Kringelbach, M. L., O'Doherty, J., Rolls, E. T., & Andrews, C. (2003). Activation of the human orbitofrontal cortex to a liquid food stimulus is correlated with its subjective pleasantness. *Cerebral cortex*, *13*(10), 1064-1071.

Landauer, T. K. (1969). Reinforcement as consolidation. *Psychological Review*, *76*(1), 82-96. https://doi.org/10.1037/h0026746

Le Bouc, R., Rigoux, L., Schmidt, L., Degos, B., Welter, M.-L., Vidailhet, M., Daunizeau, J., & Pessiglione, M. (2016). Computational Dissection of Dopamine Motor and Motivational Functions in Humans. *The Journal of Neuroscience*, *36*(25), 6623-6633. https://doi.org/10.1523/JNEUROSCI.3078-15.2016

Lee, S.-Y., Chen, S.-L., Chang, Y.-H., Chen, S.-H., Chu, C.-H., Huang, S.-Y., Tzeng, N.-S., Wang, C.-L., Lee, I. H., Yeh, T. L., Yang, Y. K., & Lu, R.-B. (2010). The ALDH2 and DRD2/ANKK1 genes interacted in bipolar II but not bipolar I disorder. *Pharmacogenetics and Genomics*, *20*(8), 500-506. https://doi.org/10.1097/FPC.0b013e32833caa2b

Lee, S.-Y., Chen, S.-L., Chen, S.-H., Huang, S.-Y., Tzeng, N.-S., Chang, Y.-H., Wang, C.-L., Lee, I. H., Yeh, T. L., Yang, Y. K., & Lu, R.-B. (2011). The *COMT* and *DRD3* genes interacted in bipolar I but not bipolar II disorder. *The World Journal of Biological Psychiatry*, *12*(5), 385-391. https://doi.org/10.3109/15622975.2010.505298

Lefaucheur, J.-P., Antal, A., Ayache, S. S., Benninger, D. H., Brunelin, J., Cogiamanian, F., Cotelli, M., De Ridder, D., Ferrucci, R., Langguth, B., Marangolo, P., Mylius, V., Nitsche, M. A., Padberg, F., Palm, U., Poulet, E., Priori, A., Rossi, S., Schecklmann, M., … Paulus, W. (2017). Evidence-based guidelines on the therapeutic use of transcranial direct current stimulation (tDCS). *Clinical Neurophysiology*, *128*(1), 56-92. https://doi.org/10.1016/j.
0.087

Leotti, L. A., & Delgado, M. R. (2014). The Value of Exercising Control Over Monetary Gains and Losses. *Psychological Science*, *25*(2), 596-604. https://doi.org/10.1177/0956797613514589

Leotti, L. A., Iyengar, S. S., & Ochsner, K. N. (2010). Born to choose : The origins and value of the need for control. *Trends in Cognitive Sciences*, *14*(10), 457-463. https://doi.org/10.1016/j.tics.2010.08.001

Lewandowski, K. E., Whitton, A. E., Pizzagalli, D. A., Norris, L. A., Ongur, D., & Hall, M.-H. (2016). Reward Learning, Neurocognition, Social Cognition, and Symptomatology in Psychosis. *Frontiers in Psychiatry*, *7*. https://doi.org/10.3389/fpsyt.2016.00100

Leyton, M., Casey, K. F., Delaney, J. S., Kolivakis, T., & Benkelfat, C. (2005). Cocaine craving, euphoria, and self-administration : A preliminary study of the effect of catecholamine precursor depletion. *Behavioral neuroscience*, *119*(6), 1619.

Li, P., Han, C., Lei, Y., Holroyd, C. B., & Li, H. (2011). Responsibility modulates neural mechanisms of outcome processing : An

ERP study: Modulation of outcome processing by responsibility. *Psychophysiology*, *48*(8), 1129-1133. https://doi.org/10.1111/j.1469-8986.2011.01182.x

Li, T., Zhao, F., & Yu, G. (2021). Who is more utilitarian? Negative affect mediates the relation between control deprivation and moral judgment. *Current Psychology*, *40*(8), 4024-4030. https://doi.org/10.1007/s12144-019-00301-1

Lichtenstein, P., Yip, B. H., Björk, C., Pawitan, Y., Cannon, T. D., Sullivan, P. F., & Hultman, C. M. (2009). Common genetic determinants of schizophrenia and bipolar disorder in Swedish families : A population-based study. *The Lancet*, *373*(9659), 234-239. https://doi.org/10.1016/S0140-6736(09)60072-6

Linke, J., King, A. V., Rietschel, M., Strohmaier, J., Hennerici, M., Gass, A., Meyer-Lindenberg, A., & Wessa, M. (2012). Increased Medial Orbitofrontal and Amygdala Activation : Evidence for a Systems-Level Endophenotype of Bipolar I Disorder. *American Journal of Psychiatry*, *169*(3), 316-325. https://doi.org/10.1176/appi.ajp.2011.11050711

Linke, J., Sönnekes, C., & Wessa, M. (2011). Sensitivity to positive and negative feedback in euthymic patients with bipolar I disorder : The last episode makes the difference. *Bipolar Disorders*, *13*(7-8), 638-650. https://doi.org/10.1111/j.1399-5618.2011.00956.x

Lorenz, R. C., Gleich, T., Kühn, S., Pöhland, L., Pelz, P., Wüstenberg, T., Raufelder, D., Heinz, A., & Beck, A. (2015). Subjective illusion of control modulates striatal reward anticipation in adolescence. *NeuroImage*, *117*, 250-257. https://doi.org/10.1016/j.neuroimage.2015.05.024

Lozano, B. E., & Johnson, S. L. (2001). Can personality traits predict increases in manic and depressive symptoms? *Journal of Affective Disorders*, *63*(1-3), 103-111. https://doi.org/10.1016/S0165-0327(00)00191-9

Lutz, K., & Widmer, M. (2014). What can the monetary incentive delay task tell us about the neural processing of reward and punishment? *Neuroscience and Neuroeconomics*, 33-45.

Ma, G., Wang, C., Jia, Y., Wang, J., Zhang, B., Shen, C., Fan, H., Pan, B., & Wang, W. (2018). Electrocardiographic and Electrooculographic Responses to External Emotions and Their Transitions in Bipolar I and II Disorders. *International Journal of Environmental Research and Public Health*, *15*(5), 884. https://doi.org/10.3390/ijerph15050884

Maletic, V., & Raison, C. (2014). Integrated Neurobiology of Bipolar Disorder. *Frontiers in Psychiatry*, *5*. https://doi.org/10.3389/fpsyt.2014.00098

Malloy-Diniz, L. F., Neves, F. S., De Moraes, P. H. P., De Marco, L. A., Romano-Silva, M. A., Krebs, M.-O., & Corrêa, H. (2011). The 5-HTTLPR polymorphism, impulsivity and suicide behavior in euthymic bipolar patients. *Journal of Affective Disorders*, *133*(1-2), 221-226. https://doi.org/10.1016/j.jad.2011.03.051

Marr, D. (2010). *Vision : A computational investigation into the human representation and processing of visual information*. MIT press.

Martino, D. J., Strejilevich, S. A., Torralva, T., & Manes, F. (2011). Decision making in euthymic bipolar I and bipolar II disorders. *Psychological Medicine*, *41*(6), 1319-1327. https://doi.org/10.1017/S0033291710001832

Mason, L., Eldar, E., & Rutledge, R. B. (2017). Mood Instability and Reward Dysregulation-A Neurocomputational Model of Bipolar Disorder. *JAMA*

Psychiatry, *74*(12), 1275-1276. https://doi.org/10.1001/jamapsychiatry.2017.3163

Mason, L., O'Sullivan, N., Montaldi, D., Bentall, R. P., & El-Deredy, W. (2014). Decision-making and trait impulsivity in bipolar disorder are associated with reduced prefrontal regulation of striatal reward valuation. *Brain*, *137*(8), 2346-2355.

McClintock, S. M., Reti, I. M., Carpenter, L. L., McDonald, W. M., Dubin, M., Taylor, S. F., Cook, I. A., O'Reardon, J., Husain, M. M., Wall, C., Krystal, A. D., Sampson, S. M., Morales, O., Nelson, B. G., Latoussakis, V., George, M. S., Lisanby, S. H., & on behalf of both the National Network of Depression Centers rTMS Task Group and the American Psychiatric Association Council on Research Task Force on Novel Biomarkers and Treatments. (2018). Consensus Recommendations for the Clinical Application of Repetitive Transcranial Magnetic Stimulation (rTMS) in the Treatment of Depression : (Consensus Statement). *The Journal of Clinical Psychiatry*, *79*(1), 35-48. https://doi.org/10.4088/JCP.16cs10905

McLauchlan, D. J., Lancaster, T., Craufurd, D., Linden, D. E. J., & Rosser, A. E. (2022). Different depression : Motivational anhedonia governs antidepressant efficacy in Huntington's disease. *Brain Communications*, *4*(6), fcac278. https://doi.org/10.1093/braincomms/fcac278

Mei, S., Yi, W., Zhou, S., Liu, X., & Zheng, Y. (2018). Contextual valence modulates the effect of choice on incentive processing. *Social Cognitive and Affective Neuroscience*, *13*(12), 1249-1258. https://doi.org/10.1093/scan/nsy098

Merikangas, K. R., Jin, R., He, J.-P., Kessler, R. C., Lee, S., Sampson, N. A., Viana, M. C., Andrade, L. H., Hu, C., Karam, E. G., Ladea, M., Medina-Mora, M. E., Ono, Y., Posada-Villa, J., Sagar, R., Wells, J. E., & Zarkov, Z. (2011). Prevalence and Correlates of Bipolar Spectrum Disorder in the World Mental Health Survey Initiative. *Archives of General Psychiatry*, *68*(3), 241. https://doi.org/10.1001/archgenpsychiatry.2011.12

Meyer, B., Johnson, S. L., & Winters, R. (2001). Responsiveness to threat and incentive in bipolar disorder : Relations of the BIS/BAS scales with symptoms. *Journal of psychopathology and behavioral assessment*, *23*, 133-143.

Milev, R. V., Giacobbe, P., Kennedy, S. H., Blumberger, D. M., Daskalakis, Z. J., Downar, J., Modirrousta, M., Patry, S., Vila-Rodriguez, F., Lam, R. W., MacQueen, G. M., Parikh, S. V., Ravindran, A. V., & the CANMAT Depression Work Group. (2016). Canadian Network for Mood and Anxiety Treatments (CANMAT) 2016 Clinical Guidelines for the Management of Adults with Major Depressive Disorder : Section 4. Neurostimulation Treatments. *The Canadian Journal of Psychiatry*, *61*(9), 561-575. https://doi.org/10.1177/0706743716660033

Mirenowicz, J., & Schultz, W. (1994). Importance of unpredictability for reward responses in primate dopamine neurons. *Journal of neurophysiology*, *72*(2), 1024-1027.

Miskowiak, K. W., Seeberg, I., Kjaerstad, H. L., Burdick, K. E., Martinez-Aran, A., Bonnin, C., Bowie, C. R., Carvalho, A. F., Gallagher, P., Hasler, G., Lafer, B., López-Jaramillo, C., Sumiyoshi, T., McIntyre, R. S., Schaffer, A., Porter, R. J., Purdon, S., Torres, I. J., Yatham, L. N., … Vieta, E. (2019). Affective cognition in bipolar disorder : A systematic review by the ISBD targeting cognition task force. *Bipolar Disorders*, *21*(8), 686-719. https://doi.org/10.1111/bdi.12834

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred Reporting Items for Systematic

Reviews and Meta-Analyses : The PRISMA Statement. *PLoS Medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Mowrer, R. R., & Klein, S. B. (2000). *Handbook of contemporary learning theories*. Psychology Press.

Mühlberger, C., Angus, D. J., Jonas, E., Harmon-Jones, C., & Harmon-Jones, E. (2017). Perceived control increases the reward positivity and stimulus preceding negativity. *Psychophysiology*, *54*(2), 310-322. https://doi.org/10.1111/psyp.12786

Mutz, J. (2023). Brain stimulation treatment for bipolar disorder. *Bipolar Disorders*, *25*(1), 9-24. https://doi.org/10.1111/bdi.13283

Nair, A., Rutledge, R. B., & Mason, L. (2020). Under the Hood : Using Computational Psychiatry to Make Psychological Therapies More Mechanism-Focused. *Frontiers in Psychiatry*, *11*, 140. https://doi.org/10.3389/fpsyt.2020.00140

Nusslock, R., & Alloy, L. B. (2017). Reward processing and mood-related symptoms : An RDoC and translational neuroscience perspective. *Journal of Affective Disorders*, *216*, 3-16. https://doi.org/10.1016/j.jad.2017.02.001

Nusslock, R., Almeida, J. R., Forbes, E. E., Versace, A., Frank, E., LaBarbara, E. J., Klein, C. R., & Phillips, M. L. (2012). Waiting to win : Elevated striatal and orbitofrontal cortical activity during reward anticipation in euthymic bipolar disorder adults: Nusslock et al. *Bipolar Disorders*, *14*(3), 249-260. https://doi.org/10.1111/j.1399-5618.2012.01012.x

O'DOHERTY, J. P., Hampton, A., & Kim, H. (2007). Model-based fMRI and its application to reward learning and decision making. *Annals of the New York Academy of sciences*, *1104*(1), 35-53.

Ono, Y., Kikuchi, M., Hirosawa, T., Hino, S., Nagasawa, T., Hashimoto, T., Munesue, T., & Minabe, Y. (2015). Reduced prefrontal activation during performance of the Iowa Gambling Task in patients with bipolar disorder. *Psychiatry Research: Neuroimaging*, *233*(1), 1-8. https://doi.org/10.1016/j.pscychresns.2015.04.003

O'Sullivan, N., Szczepanowski, R., El-Deredy, W., Mason, L., & Bentall, R. P. (2011). fMRI evidence of a relationship between hypomania and both increased goal-sensitivity and positive outcome-expectancy bias. *Neuropsychologia*, *49*(10), 2825-2835.

Palminteri, S., Justo, D., Jauffret, C., Pavlicek, B., Dauta, A., Delmaire, C., Czernecki, V., Karachi, C., Capelle, L., Durr, A., & Pessiglione, M. (2012). Critical Roles for Anterior Insula and Dorsal Striatum in Punishment-Based Avoidance Learning. *Neuron*, *76*(5), 998-1009. https://doi.org/10.1016/j.neuron.2012.10.017

Palminteri, S., Lebreton, M., Worbe, Y., Grabli, D., Hartmann, A., & Pessiglione, M. (2009). Pharmacological modulation of subliminal learning in Parkinson's and Tourette's syndromes. *Proceedings of the National Academy of Sciences*, *106*(45), 19179-19184.

Pasco, J. A., Jacka, F. N., Williams, L. J., Henry, M. J., Nicholson, G. C., Kotowicz, M. A., & Berk, M. (2010). Clinical Implications of the Cytokine Hypothesis of Depression : The Association between Use of Statins and Aspirin and the Risk of Major Depression. *Psychotherapy and Psychosomatics*, *79*(5), 323-325. https://doi.org/10.1159/000319530

Pavlov, I. P. (1927). *Conditioned reflexes : An investigation of the physiological activity of the cerebral cortex*. (p. xv, 430). Oxford Univ. Press.

Pearlson, G. D. (1995). In Vivo D2 Dopamine Receptor Density in Psychotic and Nonpsychotic Patients With Bipolar

Disorder. *Archives of General Psychiatry*, *52*(6), 471. https://doi.org/10.1001/archpsyc.1995.03950180057008

Perlis, R. H., Ostacher, M. J., Patel, J. K., Marangell, L. B., Zhang, H., Wisniewski, S. R., Ketter, T. A., Miklowitz, D. J., Otto, M. W., Gyulai, L., Reilly-Harrington, N. A., Nierenberg, A. A., Sachs, G. S., & Thase, M. E. (2006). Predictors of Recurrence in Bipolar Disorder : Primary Outcomes From the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD). *American Journal of Psychiatry*, *163*(2), 217-224. https://doi.org/10.1176/appi.ajp.163.2.217

Pessiglione, M., Le Bouc, R., & Vinckier, F. (2018). When decisions talk : Computational phenotyping of motivation disorders. *Current Opinion in Behavioral Sciences*, *22*, 50-58. https://doi.org/10.1016/j.cobeha.2017.12.014

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., & Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*, *442*(7106), 1042-1045. https://doi.org/10.1038/nature05051

Pessiglione, M., Vinckier, F., Bouret, S., Daunizeau, J., & Le Bouc, R. (2018). Why not try harder? Computational approach to motivation deficits in neuro-psychiatric diseases. *Brain*, *141*(3), 629-650. https://doi.org/10.1093/brain/awx278

Pizzagalli, D. A., Goetz, E., Ostacher, M., Iosifescu, D. V., & Perlis, R. H. (2008). Euthymic Patients with Bipolar Disorder Show Decreased Reward Learning in a Probabilistic Reward Task. *Biological Psychiatry*, *64*(2), 162-168. https://doi.org/10.1016/j.biopsych.2007.12.001

Polich, J. (2007). Updating P300 : An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128-2148. https://doi.org/10.1016/j.clinph.2007.04.019

Powers, R. L., Russo, M., Mahon, K., Brand, J., Braga, R. J., Malhotra, A. K., & Burdick, K. E. (2013). Impulsivity in bipolar disorder : Relationships with neurocognitive dysfunction and substance use history. *Bipolar Disorders*, *15*(8), 876-884. https://doi.org/10.1111/bdi.12124

Pratt, D. N., Barch, D. M., Carter, C. S., Gold, J. M., Ragland, J. D., Silverstein, S. M., & MacDonald, A. W. (2021). Reliability and Replicability of Implicit and Explicit Reinforcement Learning Paradigms in People With Psychotic Disorders. *Schizophrenia Bulletin*, *47*(3), 731-739. https://doi.org/10.1093/schbul/sbaa165

Proudfit, G. H. (2015). The reward positivity : From basic research on reward to a biomarker for depression: The reward positivity. *Psychophysiology*, *52*(4), 449-459. https://doi.org/10.1111/psyp.12370

Redgrave, P., Rodriguez, M., Smith, Y., Rodriguez-Oroz, M. C., Lehericy, S., Bergman, H., Agid, Y., DeLong, M. R., & Obeso, J. A. (2010). Goal-directed and habitual control in the basal ganglia : Implications for Parkinso
Nature Reviews Neuroscien
760-772.

Redlich, R., Dohm, K., Grotegerd, D., Opel, N., Zwitserlood, P., Heindel, W., Arolt, V., Kugel, H., & Dannlowski, U. (2015). Reward Processing in Unipolar and Bipolar Depression : A Functional MRI Study. *Neuropsychopharmacology*, *40*(11), 2623-2631. https://doi.org/10.1038/npp.2015.110

Rescorla, R. A. (1967). Pavlovian conditioning and its proper control procedures. *Psychological review*, *74*(1), 71.

Rescorla, R. A. (1972). A theory of Pavlovian conditioning : Variations in the effectiveness of reinforcement and non-

reinforcement. *Classical conditioning, Current research and theory*, *2*, 64-69.

Richards, J. M., Plate, R. C., & Ernst, M. (2013). A systematic review of fMRI reward paradigms used in studies of adolescents vs. Adults : The impact of task design and implications for understanding neurodevelopment. *Neuroscience & Biobehavioral Reviews*, *37*(5), 976-991.

Robinson, S., Sandstrom, S. M., Denenberg, V. H., & Palmiter, R. D. (2005). Distinguishing Whether Dopamine Regulates Liking, Wanting, and/or Learning About Rewards. *Behavioral Neuroscience*, *119*(1), 5-15. https://doi.org/10.1037/0735-7044.119.1.5

Robinson, T. E., & Berridge, K. C. (1993). The neural basis of drug craving : An incentive-sensitization theory of addiction. *Brain research reviews*, *18*(3), 247-291.

Roiser, J. P., Cannon, D. M., Gandhi, S. K., Tavares, J. T., Erickson, K., Wood, S., Klaver, J. M., Clark, L., Zarate Jr, C. A., Sahakian, B. J., & Drevets, W. C. (2009). Hot and cold cognition in unmedicated depressed subjects with bipolar disorder. *Bipolar Disorders*, *11*(2), 178-189. https://doi.org/10.1111/j.1399-5618.2009.00669.x

Romaniuk, L., Sandu, A.-L., Waiter, G. D., McNeil, C. J., Xueyi, S., Harris, M. A., Macfarlane, J. A., Lawrie, S. M., Deary, I. J., Murray, A. D., Delgado, M. R., Steele, J. D., McIntosh, A. M., & Whalley, H. C. (2019). The Neurobiology of Personal Control During Reward Learning and Its Relationship to Mood. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *4*(2), 190-199. https://doi.org/10.1016/j.bpsc.2018.09.015

Rowland, T. A., & Marwaha, S. (2018). Epidemiology and risk factors for bipolar disorder. *Therapeutic Advances in Psychopharmacology*, *8*(9), 251-269. https://doi.org/10.1177/2045125318769235

Rutledge, R. B., & Adams, R. A. (2017). Computational Psychiatry. In A. A. Moustafa (Éd.), *Computational Models of Brain and Behavior* (p. 29-42). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119159193.ch3

Rutledge, R. B., Skandali, N., Dayan, P., & Dolan, R. J. (2014). A computational and neural model of momentary subjective well-being. *Proceedings of the National Academy of Sciences*, *111*(33), 12252-12257. https://doi.org/10.1073/pnas.1407535111

Ryu, V., Ha, R. Y., Lee, S. J., Ha, K., & Cho, H.-S. (2017). Behavioral and Electrophysiological Alterations for Reinforcement Learning in Manic and Euthymic Patients with Bipolar Disorder. *CNS Neuroscience & Therapeutics*, *23*(3), 248-256. https://doi.org/10.1111/cns.12671

S. de Sá, A., Campos, C., B.F. Rocha, N., Yuan, T.-F., Paes, F., Arias-Carrión, O., G. Carta, M., E. Nardi, A., Cheniaux, E., & Machado, S. (2016). Neurobiology of Bipolar Disorder : Abnormalities on Cognitive and Cortical Functioning and Biomarker Levels. *CNS & Neurological Disorders - Drug Targets*, *15*(6), 713-722. https://doi.org/10.2174/1871527315666160321111359

Saez, I., & Gu, X. (2023). Invasive Computational Psychiatry. *Biological Psychiatry*, *93*(8), 661-670. https://doi.org/10.1016/j.biopsych.2022.09.032

Salamone, J. D., & Correa, M. (2012). The mysterious motivational functions of mesolimbic dopamine. *Neuron*, *76*(3), 470-485.

Sani, O. G., Yang, Y., Lee, M. B., Dawes, H. E., Chang, E. F., & Shanechi, M. M. (2018). Mood variations decoded from multi-site intracranial human brain activity. *Nature*

Biotechnology, *36*(10), 954-961. https://doi.org/10.1038/nbt.4200

Satterthwaite, T. D., Kable, J. W., Vandekar, L., Katchmar, N., Bassett, D. S., Baldassano, C. F., Ruparel, K., Elliott, M. A., Sheline, Y. I., Gur, R. C., Gur, R. E., Davatzikos, C., Leibenluft, E., Thase, M. E., & Wolf, D. H. (2015). Common and Dissociable Dysfunction of the Reward System in Bipolar and Unipolar Depression. *Neuropsychopharmacology*, *40*(9), 2258-2268. https://doi.org/10.1038/npp.2015.75

Scangos, K. W., Khambhati, A. N., Daly, P. M., Makhoul, G. S., Sugrue, L. P., Zamanian, H., Liu, T. X., Rao, V. R., Sellers, K. K., Dawes, H. E., Starr, P. A., Krystal, A. D., & Chang, E. F. (2021). Closed-loop neuromodulation in an individual with treatment-resistant depression. *Nature Medicine*, *27*(10), 1696-1700. https://doi.org/10.1038/s41591-021-01480-w

Schreiter, S., Spengler, S., Willert, A., Mohnke, S., Herold, D., Erk, S., Romanczuk-Seiferth, N., Quinlivan, E., Hindi-Attar, C., Banzhaf, C., Wackerhagen, C., Romund, L., Garbusow, M., Stamm, T., Heinz, A., Walter, H., & Bermpohl, F. (2016). Neural alterations of fronto-striatal circuitry during reward anticipation in euthymic bipolar disorder. *Psychological Medicine*, *46*(15), 3187-3198. https://doi.org/10.1017/S0033291716001963

Schultz, W. (2016). Dopamine reward prediction error coding. *Dialogues in Clinical Neuroscience*, *18*(1), 23-32. https://doi.org/10.31887/DCNS.2016.18.1/wschultz

Scott, J., Stanton, B., Garland, A., & Ferrier, I. N. (2000). Cognitive vulnerability in patients with bipolar disorder. *Psychological Medicine*, *30*(2), 467-472. https://doi.org/10.1017/S0033291799008879

Seymour, B., Singer, T., & Dolan, R. (2007). The neurobiology of punishment. *Nature Reviews Neuroscience*, *8*(4), 300-311. https://doi.org/10.1038/nrn2119

Sheth, S. A., Bijanki, K. R., Metzger, B., Allawala, A., Pirtle, V., Adkinson, J. A., Myers, J., Mathura, R. K., Oswalt, D., Tsolaki, E., Xiao, J., Noecker, A., Strutt, A. M., Cohn, J. F., McIntyre, C. C., Mathew, S. J., Borton, D., Goodman, W., & Pouratian, N. (2022). Deep Brain Stimulation for Depression Informed by Intracranial Recordings. *Biological Psychiatry*, *92*(3), 246-251. https://doi.org/10.1016/j.biopsych.2021.11.007

Sienkiewicz-Jarosz, H., Scinska, A., Swiecicki, L., Lipczynska-Lojkowska, W., Kuran, W., Ryglewicz, D., Kolaczkowski, M., Samochowiec, J., & Bienkowski, P. (2013). Sweet liking in patients with Parkinson's disease. *Journal of the neurological sciences*, *329*(1-2), 17-22.

Skinner, B. F. (1938). *The behavior of organisms; an experimental analysis*. D. Appleton-Century Company, Incorporated New York, London.

Small, D. M., Zatorre, R. J., Dagher, A., Evans, A. C., & Jones-Gotman, M. (2001). Changes in brain activity related to eating chocolate : From pleasure to aversion. *Brain*, *124*(9), 1720-1733.

Song, J., Kuja-Halkola, R., Sjölander, A., Bergen, S. E., Larsson, H., Landén, M., & Lichtenstein, P. (2018). Specificity in Etiology of Subtypes of Bipolar Disorder : Evidence From a Swedish Population-Based Family Study. *Biological Psychiatry*, *84*(11), 810-816. https://doi.org/10.1016/j.biopsych.2017.11.014

Srivastava, C., Bhardwaj, A., Sharma, M., & Kumar, S. (2019). Cognitive Deficits in Euthymic Patients With Bipolar Disorder : State or Trait Marker? *Journal of Nervous & Mental Disease*, *207*(2), 100-105. https://doi.org/10.1097/NMD.0000000000000920

Staddon, J. E. R. (2016). *Adaptive Behavior and Learning*. Cambridge University Press. https://books.google.fr/books?id=p2OKCwAAQBAJ

Stahl, S. M. (2013). *Stahl's essential psychopharmacology : Neuroscientific basis and practical applications, 4th ed.* (p. xv, 608). Cambridge University Press.

Stolz, D. S., Müller-Pinzler, L., Krach, S., & Paulus, F. M. (2020). Internal control beliefs shape positive affect and associated neural dynamics during outcome valuation. *Nature Communications*, *11*(1), 1230. https://doi.org/10.1038/s41467-020-14800-4

Strauss, G. P., Thaler, N. S., Matveeva, T. M., Vogel, S. J., Sutton, G. P., Lee, B. G., & Allen, D. N. (2015). Predicting psychosis across diagnostic boundaries : Behavioral and computational modeling evidence for impaired reinforcement learning in schizophrenia and bipolar disorder with a history of psychosis. *Journal of Abnormal Psychology*, *124*(3), 697-708. https://doi.org/10.1037/abn0000039

Subedi, K. (2022). *General Psychology : Conceptual and Methodological Frameworks* (p. 1-31).

Suri, R. E., & Schultz, W. (1999). A neural network model with dopamine-like reinforcement signal that learns a spatial delayed response task. *Neuroscience*, *91*(3), 871-890.

Teixeira, A. L., Salem, H., Frey, B. N., Barbosa, I. G., & Machado-Vieira, R. (2016). Update on bipolar disorder biomarker candidates. *Expert Review of Molecular Diagnostics*, *16*(11), 1209-1220. https://doi.org/10.1080/14737159.2016.1248413

Thorndike, E. L. (1911). *Animal intelligence : Experimental studies.* (p. viii, 297). Macmillan Press. https://doi.org/10.5962/bhl.title.55072

Thorndike, E. L. (1933). A theory of the action of the after-effects of a connection upon it. *Psychological Review*, *40*(5), 434.

Thuillard, S., & Dan-Glauser, E. S. (2017). The regulatory effect of choice in Situation Selection reduces experiential, exocrine and respiratory arousal for negative emotional stimulations. *Scientific Reports*, *7*(1), 12626. https://doi.org/10.1038/s41598-017-12626-7

Thuillard, S., & Dan-Glauser, E. S. (2021). Efficiency of Illusory Choice Used as a Variant of Situation Selection for Regulating Emotions : Reduction of Positive Experience But Preservation of Physiological Downregulation. *Applied Psychophysiology and Biofeedback*, *46*(1), 115-132. https://doi.org/10.1007/s10484-020-09484-x

Tricomi, E. M., Delgado, M. R., & Fiez, J. A. (2004). Modulation of Caudate Activity by Action Contingency. *Neuron*, *41*(2), 281-292. https://doi.org/10.1016/S0896-6273(03)00848-1

Trost, S., Diekhof, E. K., Zvonik, K., Lewandowski, M., Usher, J., Keil, M., Zilles, D., Falkai, P., Dechent, P., & Gruber, O. (2014). Disturbed Anterior Prefrontal Control of the Mesolimbic Reward System and Increased Impulsivity in Bipolar Disorder. *Neuropsychopharmacology*, *39*(8), 1914-1923. https://doi.org/10.1038/npp.2014.39

Urošević, S., Abramson, L. Y., Harmon-Jones, E., & Alloy, L. B. (2008). Dysregulation of the behavioral approach system (BAS) in bipolar spectrum disorders : Review of theory and evidence. *Clinical psychology review*, *28*(7), 1188-1205.

Van Enkhuizen, J., Henry, B. L., Minassian, A., Perry, W., Milienne-Petiot, M., Higa, K. K., Geyer, M. A., & Young, J. W. (2014). Reduced Dopamine Transporter Functioning Induces High-Reward Risk-Preference Consistent with Bipolar Disorder. *Neuropsychopharmacology*, *39*(13),

3112-3122. https://doi.org/10.1038/npp.2014.170

Vandendriessche, H., Demmou, A., Bavard, S., Yadak, J., Lemogne, C., Mauras, T., & Palminteri, S. (2023). Contextual influence of reinforcement learning performance of depression : Evidence for a negativity bias? *Psychological Medicine*, *53*(10), 4696-4706. https://doi.org/10.1017/S0033291722001593

Vinckier, F., Rigoux, L., Oudiette, D., & Pessiglione, M. (2018). Neuro-computational account of how mood fluctuations arise and affect decision making. *Nature Communications*, *9*(1), 1708. https://doi.org/10.1038/s41467-018-03774-z

Wang, K. S., & Delgado, M. R. (2019). Corticostriatal Circuits Encode the Subjective Value of Perceived Control. *Cerebral Cortex*, *29*(12), 5049-5060. https://doi.org/10.1093/cercor/bhz045

Warr, P. B., Sánchez-Cardona, I., Taneva, S. K., Vera, M., Bindl, U. K., & Cifre, E. (2021). Reinforcement Sensitivity Theory, approach-affect and avoidance-affect. *Cognition and Emotion*, *35*(4), 619-635. https://doi.org/10.1080/02699931.2020.1855119

Watkins, C. J., & Dayan, P. (1992). Q-learning. *Machine learning*, *8*, 279-292.

Wessa, M., Kanske, P., & Linke, J. (2014). Bipolar disorder : A neural network perspective on a disorder of emotion and motivation. *Restorative Neurology and Neuroscience*, *32*(1), 51-62. https://doi.org/10.3233/RNN-139007

White, N. M. (1996). Addictive drugs as reinforcers : Multiple partial actions on memory systems. *Addiction (Abingdon, England)*, *91*(7), 921-949; discussion 951-965.

Whitton, A. E., Lewandowski, K. E., & Hall, M.-H. (2021). Smoking as a Common Modulator of Sensory Gating and Reward Learning in Individuals with Psychotic Disorders. *Brain Sciences*, *11*(12), 1581. https://doi.org/10.3390/brainsci11121581

Whitton, A. E., Treadway, M. T., & Pizzagalli, D. A. (2015). Reward processing dysfunction in major depression, bipolar disorder and schizophrenia. *Current Opinion in Psychiatry*, *28*(1), 7-12. https://doi.org/10.1097/YCO.0000000000000122

Winker, C., Rehbein, M. A., Sabatinelli, D., & Junghofer, M. (2020). Repeated noninvasive stimulation of the ventromedial prefrontal cortex reveals cumulative amplification of pleasant compared to unpleasant scene processing : A single subject pilot study. *PLOS ONE*, *15*(1), e0222057. https://doi.org/10.1371/journal.pone.0222057

Yatham, L. N., Kennedy, S. H., Parikh, S. V., Schaffer, A., Bond, D. J., Frey, B. N., Sharma, V., Goldstein, B. I., Rej, S., Beaulieu, S., Alda, M., MacQueen, G., Milev, R. V., Ravindran, A., O'Donovan, C., McIntosh, D., Lam, R. W., Vazquez, G., Kapczinski, F., … Berk, M. (2018). Canadian Network for Mood and Anxiety Treatments (CANMAT) and International Society for Bipolar Disorders (ISBD) 2018 guidelines for the management of patients with bipolar disorder. *Bipolar Disorders*, *20*(2), 97-170. https://doi.org/10.1111/bdi.12609

Yechiam, E., Hayden, E. P., Bodkins, M., O'Donnell, B. F., & Hetrick, W. P. (2008). Decision making in bipolar disorder : A cognitive modeling approach. *Psychiatry Research*, *161*(2), 142-152. https://doi.org/10.1016/j.psychres.2007.07.001

Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nature Reviews Neuroscience*, *7*(6), 464-476.

Yip, S. W., Worhunsky, P. D., Rogers, R. D., & Goodwin, G. M. (2015). Hypoactivation

of the Ventral and Dorsal Striatum During Reward and Loss Anticipation in Antipsychotic and Mood Stabilizer-Naive Bipolar Disorder. *Neuropsychopharmacology*, *40*(3), 658-666. https://doi.org/10.1038/npp.2014.215

Zénon, A., Devesse, S., & Olivier, E. (2016). Dopamine Manipulation Affects Response Vigor Independently of Opportunity Cost. *The Journal of Neuroscience*, *36*(37), 9516-9525. https://doi.org/10.1523/JNEUROSCI.4467-15.2016

Résumé

Les personnes atteintes de trouble bipolaire (TB) présentent une augmentation pathologique des comportements orientés vers un but à la recherche de plaisir durant les épisodes maniaques, et une diminution lors des épisodes dépressifs. La période de rémission peut comprendre des altérations cognitives, dont l'étude permettrait d'identifier des marqueurs spécifiques de la maladie et contribuer à développer des traitements personnalisés. Pour ce travail de thèse, nous nous sommes demandé si les personnes ayant un TB présenteraient des altérations en apprentissage par renforcement durant la rémission, et si ces altérations seraient différentes selon le type de TB. En effet, la principale différence entre le type I (TB-I) et le type II (TB-II) est la présence d'épisode maniaque dans le TB-I et non dans le TB-II, et dont l'intensité des comportements à visée hédonique est supérieure aux épisodes hypomaniaques. Enfin, compte-tenu de la part de l'environnement dans l'émergence des épisodes thymiques, nous nous intéressons à l'impact de l'agentivité sur l'humeur dans le TB en rémission. Nous avons effectué pour ce travail de thèse une 1ère étude méta-analytique des études comportementales en apprentissage par renforcement dans le TB, mettant en évidence une altération sélective de l'apprentissage par récompense par rapport à la punition, notamment durant la rémission. Nous avons ensuite une 2ème étude, expérimentale, sur 79 sujets ayant un TB en rémission et 30 sujets sains contrôles (SC), en leur faisant passer une tâche d'apprentissage probabiliste afin de mieux caractériser ce déficit selon le sous-type de TB. Nous avons montré une altération de l'apprentissage par récompense chez les personnes ayant un TB, et que les TB-I apprennent mieux de leurs échecs que de leurs succès par rapport aux SC. L'apport de la modélisation computationnelle nous a permis de mettre en avant que cette altération était due à une hyposensibilité à la récompense. Concernant l'implication de l'environnement dans ce déficit d'apprentissage, nous savons qu'il existe un lien bidirectionnel entre les choix et l'humeur. Nous savons également que les choix en apprentissage peuvent être altérés de façon différente selon le niveau d'agentivité. Pour ce travail de thèse, nous nous demandons

si cela pourrait impacter l'humeur des personnes ayant un TB. Nous avons ainsi mené une 3ème et nouvelle étude expérimentale chez 32 personnes ayant un TB contre 32 SC, effectuant une variante de notre première tâche en manipulant l'agentivité et en observant l'impact sur l'humeur. Les résultats préliminaires ne retrouvent pas de différence entre les sujets sains et le groupe TB, mais une humeur significativement meilleure après une récompense obtenue de choix libres plutôt que de choix forcés, et une humeur moins bonne après les punitions, sans distinction entre la condition libre et forcée. Ces travaux mêlant une approche méta-analytique (comportementale) et une approche de psychiatrie computationnelle associant paradigmes expérimentaux et modélisation computationnelle, nous permettent de proposer une nouvelle modélisation des fluctuations pathologiques de l'humeur dans le TB. Nos résultats pourraient à terme permettre de personnaliser la prise en charge des personnes ayant un TB, qui présentent des comportements davantage orientés vers l'évitement de la punition que vers la recherche de la satisfaction, contribuant au maintien de symptômes résiduels de la sphère motivationnelle durant la rémission.

Abstract

People with bipolar disorder (BD) show a pathological increase in goal-directed, pleasure-seeking behaviors during manic episodes, and a decrease during depressive episodes. The remission period may include cognitive alterations, the study of which could identify disease-specific markers and contribute to the development of personalized treatments. For this thesis work, we wondered whether people with TB would show alterations in reinforcement learning during remission, and whether these alterations would differ according to the type of BD. Indeed, the main difference between type I (BD-I) and type II (BD-II) is the presence of manic episodes in BD-I and not in BD-II, whose intensity of hedonic behaviors is higher than hypomanic episodes. Finally, given the role of the environment in the emergence of thymic episodes, we are interested in the impact of agency on mood in BD in remission. For this thesis work, we carried out a 1st meta-analytical study of behavioral studies of reinforcement learning in BD, highlighting a selective alteration in reward versus punishment learning, particularly during remission. We then carried out a 2nd experimental study on 79 subjects with BD in remission and 30 healthy controls (HC), using a probabilistic learning task to better characterize this deficit according to BD subtype. We showed that reward-based learning was impaired in BD patients, and that BD-I patients learned better from their failures than from their successes, compared with HC patients. The contribution of computational modeling enabled us to put forward that this alteration was due to a hyposensitivity to reward. With regard to the involvement of the environment in this learning deficit, we know that there is a bidirectional link between choices and mood. We also know that learning choices can be altered differently depending on the level of agency. For this thesis work, we wondered whether this might impact the mood of people with BD. We therefore conducted a 3rd and new experimental study in 32 BD versus 32 HC subjects, performing a variant of our first task by manipulating agency and observing the impact on mood. Preliminary results showed no difference between

healthy subjects and the BD group, but a significantly better mood after a reward obtained from free rather than forced choices, and a worse mood after punishments, with no distinction between the free and forced condition. This work, combining a meta-analytical (behavioral) approach and a computational psychiatry approach combining experimental paradigms and computational modeling, enables us to propose a new model of pathological mood fluctuations in BD. Our findings could ultimately help personalize the management of people with BD, who exhibit behaviors more oriented towards punishment avoidance than satisfaction seeking, contributing to the maintenance of residual motivational sphere symptoms during remission.