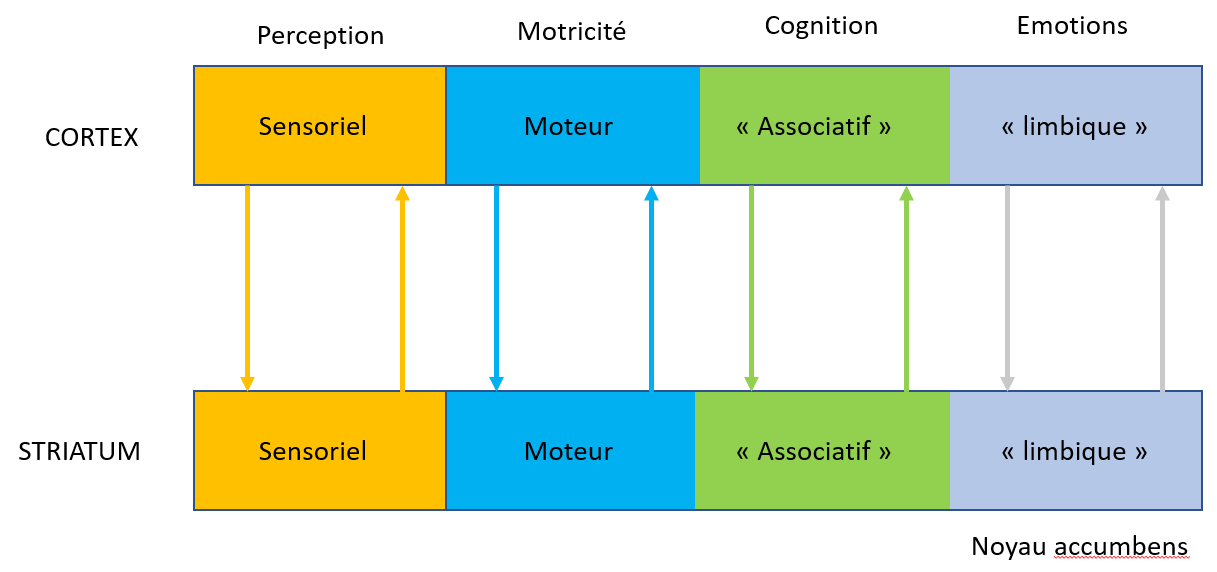
Cerveau et Récompense : Quelles structure cérébrales permettent aux récompenses d’affecter nos comportements ?

* **Qu’est qu’une** **récompense** ? = renforçateur = stimulus appétitif
  + Exemple nourriture :
    - Son apparition est renforçante
    - Sa disparition est punitive
    - Elle déclenche des comportements d’approche
  + Apparition = Renforçateur positif
  + Disparition = Punition négative
  + Provoque des comportements d’approche
  + L’inverse pour un punisseur
* La **dopamine** est-elle vraiment le NM du plaisir ? == plan part 1
* EXP **James Olds** & **Peter Milner** : (1954)
  + Prouver qu’être focus = renforçateur
  + **Paradigme d’autostimulation** : Permettre à l’animal de stimuler son cerveau par l’appuis sur un levier
    - Rat, appuis sur le levier -> stimulation du cerveau par une électrode dans le cerveau
  + N’a pas marché sauf pour un rat qui appuyé frénétiquement jusqu’à ne plus manger ect -> électrode au mauvais endroit
  + La stimulation directe du cerveau peut agir comme une récompense mais uniquement si certaines régions sont stimulées
  + **Comportement d’autostimulation** : Appuis frénétique sur le levier lors d’un paradigme d’autostimulation
* Quelle région provoque un comportement d’autostimulation ?
  + En majorité des régions du système dopaminergique
  + Mais aussi d’autre !!
  + Effet tunnel dans la recherche car du coup la majorité des études porte sur la dopamine et créé la hype dans les médias autour de la dopamine
  + Tronc cérébrale
    - Mésencéphale = partie supérieure du tronc cérébrale : deux petits noyaux dopaminergiques cote a cote
      * Aire tegmentale ventrale
      * Substance noire (car gorgé de dopamine et quand coloré au microscope devient tout noir)
      * Projection massive dans le cerveau
  + Projection :
    - **Aire ventro-tegmentale** :
      * =====> Cortex préfrontal
        + + -> Hippocampe
        + + -> Amygdale
      * ==> **Noyau accumbens**
    - **Substance noire** 
      * -> **Noyau caudé** (maladie de Parkinson ici connexion détruire)
      * -> **Putamen**
  + Schéma dans le diapo
  + Chez le rat : noyau caudé/accumbens + putamen == **striatum** (n’existe pas chez le rat) -> Endroit confus en neurscience
    - On ne parle pas du striatum pour des humains c’est un abus de langage
    - **Noyau accumbens = striatum ventral**
    - Noyau caudé + putamen = **striatum dorsolatéral**
    - Certain article utilise les deux termes dans le même paragraphe
* Quelles sont les projections de l’aire ventro-tegmentale ?
* Quelles sont les projections de la substance noire ?
* Fonction de ces structures ?
  + Hippocampe :
    - Mémoire, consolidation systémique
    - Cognition spatiale (exemple oiseau, taxi)
    - Apprentissage de séquence
    - Contexte (des comportements, apprentissage qui nécessite une configuration de stimulus)
  + Amygdale :
    - Emotions : aspect attentionnel
    - Ensemble de neurone associé à une émotion, associé à la récompense
    - Connection dense avec tout le cerveau
  + Cortex préfrontal : (=inhibition)
    - Prise de décision (=inhibition)
    - Planification (=inhibition)
    - Contrôle cognitif (=inhibition)
  + Striatum (Noyaux accumbens, noyaux caudé, putamen) :
    - Toujours en dessous du cortex
    - Boucle cortico-cortical : Cortex -> **Striatum** -> **Noyaux gris centraux = ganglions de la base** -> **Pallidum** -> thalamus -> Cortex
    - + Striatum régule l’activité du cortex (amplifier ou diminuer)
      * Exemple : Face à un ours : fuir ou attaquer mais pas les deux en même temps
        + Les deux décisions sont équivalentes, le striatum va amplifier la bonne décision
      * Exemple sensoriel :
        + Deux stimulus saillant -> amplification de l’un par le striatum
      * Explique les troubles cognitifs chez les parkinsons
    - Diapo un peu avec les bandes et les couleurs
    - **Noyau accumbens <=> émotion** + boucle avec le cortex limbique (émotion)
    - Une modification de ces régions implique une forte modification du comportement

## Cours 2 : Cerveau et récompense 2

* **Wolfram Schultz** : neurophysionlogiste, exp SingleCellRecording
* Graph rappel : x=temps, y=essai, point = PA
* Neurone dopaminergique :
  + **Réponse tonique :** 
    - PA de temps en temps même sans aucune stimulation
    - Sport augmente le niveau de réponse tonique, la quantité de dopamine tonique
    - (Parkinson = destruction du système dopaminergique = aucun problème de conditionnement)
  + **Réponse phasique :**
    - (lié à la Récompense)
    - Après certain stimulus
    - Latence faible 100ms
    - Durée faible : 200ms
    - -> augmentation très faible de la quantité de dopamine
    - -> Pas lié au plaisir car stimulation trop courte
* **Optogénétique** :
  + Rétrovirus : ADN d’algues luminescentes + vise les neurones dopaminergiques
  + Activation du neurone ⬄ émission de lumière
  + Diode lumineuse -> Activation du neurone
  + == autostimulation mais beaucoup plus précis
* EXP kal desseroth
  + Optogénétique
  + Place preference learning:
    - Un lieu de la cage == récompense = stimulation aire ventro-tegmentale
    - Le rat reste dans cet endroit
  + Stimulation :
    - Brêve : ms -> Simulation d’une réponse phasique
    - Continue : 10s -> Simulation d’eune réponse tonique
  + Résultat :
    - Stimulation phasique -> le rat reste dans la partie récompense de la cage
    - Stimulation tonique -> Aucun effet
* EXP 2 :
  + Same optogénétique + place préférence learning
  + Stimulation : Inhibition de l’aire tegmentale ventrale
    - Inhibition courte = désactivation phasique -> Aversive
    - Inhibition longue = désactivation tonique -> Aucun effet
* -> Cerveau plus complexe qu’un lieu une action : exemple des neurones dopaminergique
* Potentialisation à long terme :
  + Stimulation présynaptique puis post synaptique (+ bon timing)
  + -> renforce la connexion synaptique
  + NM excitateur = glutamate
* Dépression à long terme :
  + Stimulation post synaptique puis présynaptique
  + -> dépression à long terme
* EXP 2008:
  + Un big neurone épinal
    - Dendrite reliée à neurone du cortex + neurone dopaminergique
  + Potentialisation à long terme
    - Avec inhibition des récepteur dopaminergique -> Pas de potentialisation
    - Sans inhibition (de base il y a activation de ces récepteurs) -> Potentialisation
  + A picture containing graphical user interface

    Description automatically generated-> Dopamine joue un rôle de neuromodulation (preuve directe ici)
* Dopamine :
  + NM modulateur
  + Nécessaire pour obtenir une potentialisation à long terme + dépression à long terme
* CCL :
  + Système dopaminergique -> cible les structures importante (boucle de régulation)
  + Dopamine -> modifie les connexions synaptiques
    - Libération phasique -> renforce le comportement
    - Inhibition phasique -> punit le comportement
  + Deux HP **indépendante !!** :
    - L’effet des récompenses sur notre comportement est médiatisé par le système dopaminergique (système dopaminergique = système de récompense du cerveau) -> Vrais
      * Pour que les récompense est un effet sur le comportement il faut que ça passe par le SD
      * Si on stimules le
    - L’effet addictif des drogues est dû à leur action sur le système dopaminergique -> Vrais
      * Une condition necéssaire pour que les drogue soit addictive est quelle agissent sur le SD
    - La libération phasique de dopamine provoque une sensation de plaisir -> FAUX
      * A cause de Schult
* L’effet des récompenses sur notre comportement est médiatisé par le système dopaminergique
  + Argument :
    - Activer artificiellement le système dopaminergique renforce le comportement
    - Le système dopaminergique est ancien d’un point de vue phylogénétique
    - Le système dopaminergique se projette de manière diffuse vers des cibles importantes
    - La dopamine joue un rôle critique dans la plasticité des connections le cortex et le striatum
    - Interférer avec le système dopaminergique perturbe le renforcement d’un comportement avec des récompenses naturels.
      * -> 2 EXPs
  + EXP 1 :
    - Phase 1 : appui sur le levier -> Nourriture
    - Phase 2 : appui sur le levier -> Nourriture
    - Condition :
      * Injection d’une solution saline = contrôle
      * Injection d’un antagoniste de la dopamine = désactivation du système dopaminergique
    - Résultat :
      * Contrôle : pas de changement
    - Injection : extinction +- rapide en fonction du dosage
  + EXP2 :
    - Souris : facile pour modifier son code génétique
      * -> Réponse phasique des neurones dopaminergique modifié
    - Labyrinthe simple en T avec récompense à droite
    - Groupe :
      * Souris contrôle
      * Souris modifié -> apprenne moins vite dans toute les taches qui implique des récompenses
  + EXP3 : note mentale on ne sais pas plein de chose
    - Lumière -> nourriture
    - Deux rats possibles : origine génétique
      * Sign-tracker : Certain rat s’approche de la lumière
      * Goal-tracker : Certain rat s’approche de la mangeoire
    - Si on perturbe le Système dopaminergique -> perturbation uniquement chez les sign tracker
    - Normalement on devrait avoir une perturbation dans les deux groupes
    - -> Deux régions du cerveau différente, quel système est impliqué chez les goal tracker
    - -> Il y a plein de chose qu’on ne sait pas (interaction, ect)
    - -> Le système dopaminergique n’est pas LE SR, il y a d’autre chose !!
* L’effet addictif des drogues est dû à leur action sur le système dopaminergique
  + EXP rigolote :
    - Joana Fowler : exp neuroimagerie humain
    - PET scan : émission de positron
      * Même chose que l’IRM sauf invasif
    - Injection d’un traceur
      * Cocaïne
      * Méthamphétamine
    - Résultat :
      * Les deux substances vont dans tout le cerveau
      * Au bout d’un certain temps, la substance va en particulier rester dans le noyau acumens
      * Concentration dans le noyau acumens <=corrélé=> impression subjective de l’effet de la drogue
  + -> Drogue = agis sur le système dopaminergique (chaque drogue d’une manière particulière)

## Cours 3 : La dopamine n'est pas l'hormone du plaisir

* Hypothèse 3 : FAUSSE :
  + Argument pour :
    - Cela expliquerait les résultats des expériences d’autostimulation
    - Cela expliquerait pourquoi l’action des drogues sur le système dopaminergique les rends addictives
    - == explique aussi les anciennes HP
    - Des sujets humains reporteraient des sensations zde plaisir intense lorsque des zones liées au renforcement cérébral sont stimulés **(==un mythe comme Phileas Cage)**
      * Etude récente chez les patients Parkinsonien (implantation d’électrode) dans le cadre de l’implantation où on testait l’implant
      * Rapport clinique de Robert Heath, essaye de traiter des comportement « déviant » (homosexuelle). En renforçant/stimulant les neurones dopaminergique
    - -> Méta-Analyse des rapports clinique par Ken Berridge : constante surinterprétation des paroles du patient : « oui il dit ça parce que .. »
  + Que faire des deux premiers arguments : Autre explication
    - **Wolfram Schultz** EXP macaque
      * Problématique : Qu’est ce qui provoque une réponse phasique ?
      * Electrode Aire tegmentale ventrale + conditionnement opérant (image -> appui bouton -> Jus de fruit -> intervalle inter-essai aléatoire)
      * Résultat :
        + Début de l’apprentissage :

Image : pendant la réponse tonique

Jus de fruit : déclenche réponse phasique

* + - * + Une fois l’apprentissage acquis :

Image : réponse phasique

Jus de fruit : réponse tonique, rien

* + - * + Pourtant le singe appuis sur le bouton donc le jus de fruit est toujours une récompense
      * Réfute : dopamine = récompense = plaisir = gros bouton plaisir dans le cerveau
      * Résultat : omission du jus de fruit exceptionnel
        + Image : réponse phasique
        + Jus de fruit : diminution de la concentration de dopamine !
      * CCL :
        + Début de l’apprentissage : Neurone ne réponse pas au stimulus mais à la récompense elle-même
        + Apprentissage acquis : Réponse à la récompense disparait
        + Omission de la récompense : inhibition des neurones au moment où la récompense aurait dû être donnée
    - Chez l’humain : Parkinson
      * Stade avancé de Parkinson -> implantation d’électrodes au niveau de la substance noire pour augmenter le niveau tonique de dopamine au niveau du noyau caudé
      * Tache : Deux cartes rouges/bleu : rouge -> argent, bleu -> rien
      * Résultat : Choix carte rouge -> récompense
        + Début de l’apprentissage : réponse phasique 200ms après la récompense
        + Fin de l’apprentissage : réponse phasique précédente presque disparu
      * Résultat omission de la récompense :
        + Début de l’apprentissage : rien
        + Fin de l’apprentissage : diminution de la réponse dopaminergique
    - EXP humain normal IRM: John O’Doherty
      * Même EXP que chez le singe : image -> bouton -> jus de fruit
      * IRM réponse hémodynamique au niveau du noyau accumbens
        + Car on ne peut pas mesurer dans le mésencéphale (mauvaise vascularisation/orientation)
        + Activation du noyau accumbens ⬄ reflète l’activité du système dopaminergique
      * HP : modélisation mathématique de la réponse hémodynamique pour prévoir le paterne d’IRM à partir de celui des neurones dopaminergiques chez le singe
        + Début de l’apprentissage : réponse hémodynamique
      * Résultat :
        + Début de l’apprentissage : Suis les paterne prédit par le model !
        + Fin de l’apprentissage : same !
    - -> Réponse phasique n’a rien à voir avec le plaisir
* C’est quoi alors la dopamine ? Elle est pourtant liée au renforcement et au drogue
  + Modèle de Rescorla-Wagner
    - Le sujet cherche à prédire l’intensité du SI
    - Prédiction mise à jour lorsque le sujet est surpris
    - Valeur des stimulus : , Valeur globale :
    - Intensité réelle du stimulus :
    - Surprise :
    - Surprise importante -> augmentation proportionnelle de la valeur des stimulus
    - Apprentissage de correction d’erreur

## Cours 4 : Schultz, Rescorla Wagner, même combat

* Pourquoi lien entre dopamine et Rescorla-Wagner ?
  + Début de l’apprentissage
    - I(tensité) = 1
    - V(Stimulus)= 0
    - Surprise : I-V = 1- 0 = 1
    - -> réponse dopaminergique = erreur de prédiction positive
  + Apprentissage acquis
    - I = 1
    - V = 1
    - Surprise = I – V = 0
    - -> Pas de réponse dopaminergique
  + //
    - I=0 -> omission de récompense
    - V = 1
    - Surprise = -1
    - -> Inhibition de la réponse dopaminergique
* Ancienne HP : ~~Le système dopaminergique est le centre de plaisir du cerveau~~
* // Nouvelle HP : Le système dopaminergique implémente l’erreur de prédiction formalisée par le modèle de Rescorla-Wagner
* 4 EXP : avec cette HP et des choses plus compliqué
* EXP 1 : Schultz
  + Réponse de blocage :
    - Le conditionnement de A provoqué le blocage du conditionnement de X
    - Avec choc électrique
    - Phase 1 :
      * G1 : Rien
      * G2 : A+
    - Phase 2 :
      * G1 : AX+
      * G2 : AX+
    - Test :
      * X -> Réponse de peur
      * X -> Rien
    - Rescorla :
      * G1 : -> Modif sur Vx
      * G2 : -> Pas de modif sur Vx
  + Phase 1 : A -> Jus de fruit, B -> Rien
  + Phase 2 : AX -> Jus de fruit, BY -> Jus de fruit (contrôle )
  + Test : X -> Rien, Y -> réponse
  + Résultat : Les neurones dopaminergique modélise parfaitement la surprise de RW
  + Diagram

    Description automatically generatedVoir diapo pour être sûr, c’est des exp importantes, y’a des questions dessus
* EXP2 :
  + Chart, box and whisker chart

    Description automatically generatedSame chez l’humain, avec IRM focus sur le Striatum Ventral (notamment le noyau accubens)
  + **Changement dans la réponse hémodinamique par rapport au niveau base**
* EXP3 :
  + La valeur d’un stimulus converge vers/oscille autour la proba de renforcement
  + La réponse phasique est directement fonction de la probabilité de renforcement
  + Fiorillo, Tobler, & Schultz (2003). Nature, 299, 1898
* Diagram

  Description automatically generatedEXP4 :
  + Même principe chez l’humain
  + Jouer pour gagner de l’argent
  + Un cue indique la probabilité de gagner de l’argent
  + VD : IRM au moment de la récompense (dans le striatum ventral pour pouvoir inféré l’activité des ND)
  + Résultat :
    - Deux courbe : essais renforcé et essais non renforcé
    - Y = changement de la réponse hémodynamique par rapport au niveau de base
    - Positif -> surprise positive // négatif -> surprise négative
    - Essais renforcé et proba faible -> surprise positive -> y positive
    - Essais renforcé et proba forte -> surprise faible -> y proche de zero
    - Essas renforcé proba = 1 -> surprise nul -> y=0
    - Essais non renforcé et proba = 0 -> surprise nul -> y=0
    - Essais non renforcé et proba faible -> surprise faible -> y négatif A picture containing text, device

      Description automatically generated
* EXP5 : Variation de l’intensité du Stimulus
  + X = jus de fruit : quantité : small, medium, large
  + Convergence vers la valeur du stimulus medium
  + L’attente fixé vers la moyenne
  + Small -> erreur de prédiction négative (S= small – medium <0) -> inhibition des neurones
  + Medium -> Erreur nulle = 0 -> réponse dopaminergique normale
  + Large -> surprise positive -> réponse dopaminergique
  + Autre truc : X : small ou medium -> small < V(X) < medium -> I > 0

New Cours 9/02/22:

* **Sutton** et **Barto** : voulait créer **l’apprentissage par renforcement** =
  + **Discipline de recherche d’algorithme d’Apprentissage par essais erreur.**
  + Conditionnement opérant rien compris
  + -> conditionnement pavlovien = RW oui on aime bien ça ressemble à un trial based model
  + Travail à la marge car les réseaux de neurone n’était pas à la mode
  + === **Time Derivative learning**
  + Plus proche du pavlovien
* Trial based model : (ne prends pas en compte la durée du stimulus attention), met à jour les poids qu’à la fin de l’essai
* « on veut un truc qui prend en compte le temps et modifie en direct les poids » -> modifie RW ->
* -> Real-time model : Rend compte du temps pour mettre à jour directement
* RW ne rend pas compte du conditionnement de second ordre -> Or en apprentissage par renforcement, tous les essais n’ont pas forcément de conséquence -> Modifie RW
  + -> === Time Derivative learning
* **Chris Watkins** :
  + Omg Sutton et Barto, votre algo qui s’inspire de RW résous les problèmes de MDP
  + ==== **Q learning**
  + Plus proche du conditionnement opérant -> essaye de prédire les actions pour savoir la meilleurs à prendre
* Deep mind == ancien étudiant de sutton ou barto
* **TD learning** :
  + Explication simple mais inexacte :
    - Stimulus aversif // Stimulus appétitif
    - Ici on vas faire un continuim = valence d’un stimulus
    - Renforçateur primaire / secondaire : en RL on parle de **valeur primaire et secondaire** -> dans RW Intensité du SI = valeur primaire
    - Pour intégrer le fait qu’il n’y a pas de Valeur primaire à chaque essais, on introduit la valeur secondaire dans l’equation de RW
    - Or qu’est ce que la Valeur secondaire ? C’est la prédiction des récompenses futurs à n+1
    - Et on se retrouve avec l’équation du TD-learning
    - avec l’intensité des valeurs primaires, prédition ou valeur actuelle
    - Temps réel mais discrétisé
* Modèle au nivaeu de l’essai : comme RW, est discère au niveau de l’essai, ne prends pas en compte la durée (du stimulus)
* Modèle temps réel : Cicle de prédition très court, pour émuler un temps continue

New Cours 2/03 : Q learning version hardcore

* Problème de décicsion markovien :
  + Agent vs environnement, l’agent interagie avec l’environnement
  + Ensemble d’action possible, ensemble d’état possible
  + Une politique : ensemble de règle qui dit qu’elle action vas être émise dans tel état
  + A chaque action (qui donc modifie l’environnement), l’agent reçois un reward , F fonction de retour (dépend de la state et de l’action émise)
  + Tout dépend uniquement de la state t, pas celle d’avant t-1
  + Probabilité de transition :
  + Valeur d’un état : 
    - Valeur d’un état = somme des récompenses dans le futur en suivant la politique pi
    - Mais ca tend vers plus l’infiny donc pas ouf
    - Il faut valoriser les récompenses proche dans le temps vs celle loin (100€ maintenant ou dans 3 jours, comme en psychologie)
    - On rajoute une sorte de Learning rate pour réduire la valeur avec le temps
    - Fenêtre courte = conséquence immédiate // inverse
    - d
    - Model free vs Model Based :
      * Model Based : (DYNA) cherche à avoir une representation de l’environnement
      * Model free : Q&TD learning
      * Psycho : Modélise la différence entre comportement par habitude ou par dirigé vers un but
* TD learning :
  + Il cherche à calculter la valeur des états pour chaque état
  + En l’estimant
  + Différence entre une meilleur prédiction à t+1 et l’ancienne
  + Lien avec RW :
    - Cas particulier de l’équation sans regarder à t + 1
    - Pas de somme des valeurs
      * Le tout converge vers les valeurs des états si on passe suffisamment sur tout les états
      * En cas de beaucoup beaucoup d’état, on ne peut pas passer partout, on utilise des méthodes d’approximation de fonction (on attribue une valeur à s’ à partir de s.). Ca se fait notamment avec des réseaux de neurone