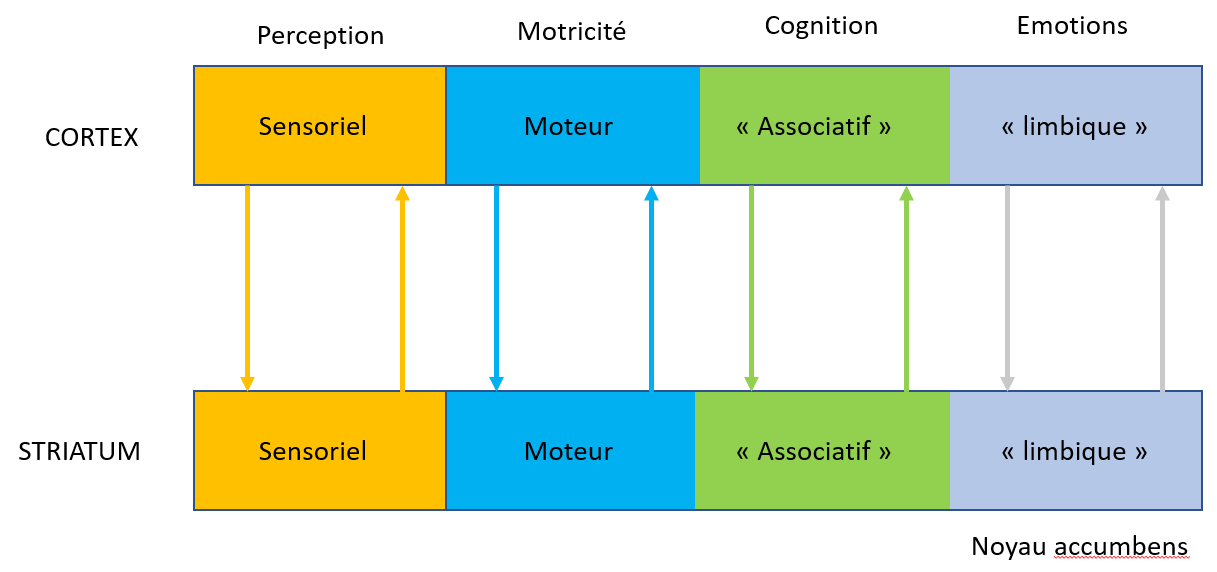
Cerveau et Récompense : Quelles structure cérébrales permettent aux récompenses d’affecter nos comportements ?

* **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
  + **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 tous 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.). Ça se fait notamment avec des réseaux de neurone

New Cours 09/03/2022

* + Avec gamma = 0 on retrouve RW
  + Si pas de ren
* Retour sur les neurone dopaminergique
* Le pique juste après le CS vient de la différence entre La prédiction avant le CS et la prédiction après le CS -> **elle doit être proportionnel à la valeur associative du stimulus alors**
  + Retour exp probabilité de renforcement
    - Cvg de la valeur associative vers la proba
    - Le pique après le CS est en effet proportionnel à la valeur associative du stimulus
  + Retour exp probabilité de renforcement chez l’humain
    - Fonctionne !
    - LA différence par rapport à la ligne de vase
    - Avant le CS = intervalle inter-essai = IIE
    - Type d’essai probabilité équiprobable, y’a a peu près 50/50 que ça soit un bonne essai ou un mauvaise essai
    - Voir figure
  + Temps est aussi un indice
    - Pas de changement dans la prédiction lorsque c’est 50% de chance d’être renforcé
    - IIE = aléatoire = [3,15] min chez l’humain, [3,40] min chez le singe wtf
    - -> Important car sinon le temps devient un indice du début de l’essai
    - EXP schultz :
      * Loupé note de jess
      * A vos marques, prêt ? Partez !
      * Lever la main
      * Condition 1) L’arrivée de « Partez » est anticipé grâce au « A vos marques »
      * Condition 2) Mais des fois, le « partez » arrive plus tôt ou plus tard que prévu
      * Condition 1) Pic dopaminergique lors du « a vos marques », mais rien au « partez »
      * Condition 2) Réponse dopaminergique lors du « partez »
      * Et puis demander au prof mdr
    - A vos marques, prêt ? Partez !
* Retour exp proba de renforcement chez le singe
  + On va regarder les autres neurones :
    - La majorité des neurones se comporte comme on a vu, répond au signal d’erreur
    - Mais une minorité fait :
  + Entre le CS et la récompense : on a une croissance des PA plus on se rapproche de la récompense, c’est activité croissante est maximal à p=0.5 == **Ramping activity = pattern de réponse type croissance, PA de plus en plus vite fréquemment**
  + -> ces neurone semble coder l’incertitude !
  + Niveau modélisation :
    - La learning rate vient influencer la vitesse d’apprentissage
    - Il y a des situations ou il vaut mieux apprendre lentement ou rapidement (environnement instable et stable resp.)
  + -> ces neurones moduleraient la vitesse d’apprentissage comme la learning rate (attention personne n’est sur)
* EXP : théorie du chaos
  + Stimulus un peu particulier avec des point orienté vers une direction, orienté de manière aléatoire. Il peut aller tous dans la même direction diapo
  + Ici le nombre de point/cercle synchronisé indique la probabilité de renforcement
  + -> nécessite un traitement visuel plus complexe
  + Résultat :
    - Dsfq
  + Interprétation : pourquoi on n’avait pas pareil chez Schulz
    - Chez schulz : Pas beaucoup de temps pour détecter si c’est un stimulus codant quelque chose ou non
    - Ici : on crée un décalage entre la détection d’un stimulus, et la reconnaissance de ce qu’il prédit
    - -> Ainsi il y a deux piques
* -> On a donc deux composantes post CS pré récompense
  + Composante détection : influencé par
    - Force physique, intensité
    - Surprise/nouveauté
    - Reward généralisation :
      * Déclenché par des stimulus qui ont un lien avec une récompense
      * Phase 1 : son -> Récompense et image -> choc
      * Phase 2 : son -> récompense et son2 -> choc et son3 -> rien, image -> rien
      * Son2, son3 déclenche la composante détection, pas les autre
      * Seul cas ou les ND vont répondre à un stimulus avec une situation aversive
    - Reward context
      * Composante plus forte dans une situation/contexte où on est beaucoup renforcé
      * // inverse
  + -> utilité : focaliser l’attention sur un stimulus nouvellement vu
  + -> cerveau compliqué que ce qu’on pense
  + Et la deuxième composante enculé tu l’as décrit pas :
    - Si j’ai bien compris c’est la composante de prédiction, celle dont on parlait tout à l’heure, en lien avec le TD learning
    - C’est le quelque chose qui fait « omg y’a quelque chose est-ce que je vais en avoir une récompense » lien avec l’orientation de l’attention
* Autres neurones qui n’ont rien à voir avec la récompense :
  + Exemple 1 : neurone qui s’active lorsque on engage un mouvement lié ou non à la récompense
  + Exemple 2 : Fin du mouvement
  + Exemple 3 :
  + Exemple 4 : s’active entre début l’essai et la fin de l’essai sans modulation avec la probabilité
* -> il reste des mystères, il y a ce qu’on sait et ce que l’on ne sait pas
* RW :
  + La majorité des expériences faite pour RW était en conditionnement de la peur
  + On ne distingue pas la différence entre les situation de renforcement et de peur (sauf peur c’est plus rapide).
  + Mais le système dopaminergique ne répond pas au conditionnement aversive
* -> donc il doit exister une autre régions qui implémente la surprise de RW
  + EXP Lapin paupière
    - On connait bien les bases neurophysio de ce conditionnent
    - Tout dans le cervelet car il est comme un petit cerveau
  + Plein d’autre régions (plus ou moins inconnu) qui semblent implémenter la surprise de RW
  + Parfois juste une réponse à une récompense inattendu, ou juste ç quelque chose qui est inattendu
  + Etude très récente, mais on ne sait pas comment elle interagisse entre elle, qu’elles sont leur fonction
* EXP : **Paul Glimcher**
  + Invente la **neuro-économie** : étude des processus neuronaux lors de choix économique entre différent type de récompense.
  + Choix entre pizza et salade : deux récompense qui joue le rôle du SC
  + On calcule la valeur de ces deux stimulus : quantité de valeur primaire si on consomme ces stimulus maintenant et plus tard (si on ne veut pas grossir par exemple) == processus de valuation == cortex préfrontale/striatum
  + Comparaison des valeur des stimulus -> décision -> consommation de la pizza
  + Evaluation de la quantité de plaisir reçu (qu’importe la source de plaisir) grâce à la pizza
  + -> changement des prédictions
  + **Internal reward currency** = les valeurs primaires seraient exprimées dans cette unité
    - Permet de comparer plusieurs type de plaisir : aller au restaurant ou aller au musée
  + Utilité = Valeur d’un stimulus

New Cours 16/03/2022

* John O’Doherty : EXP identification neuro-économie
  + Tache : Combien êtes-vous prêt à payer au plus pour ça ? (ici « pseudo »récompense = nourriture
  + Contrôle : 1 seule prix possible
  + Groupe expérimentale : 4 prix possible à choisir -> induit un choix économique
  + Résultat : Cortex préfrontal ventro-médial (VMPMC
    - Contrôle : faible activité
    - Expé : L’activité du cortex préfrontal ventro-médial semble refléter le prix de la chose mis en jeux
* Méta-analyse comme O’Deherty :
  + C’est bien le cortex ventro-médial ! qui s’active à chaque fois
  + Même région dans chaque étude :
    - VMPMC principalement
    - Mais aussi le Striatum ventral !
  + Activation indépendamment du type de récompense : Etude récompense social, Etude récompense monétaire
* EXP Todd Hare :
  + Est-ce que ça fonctionne pour les récompenses envers quelqu’un d’autre ?
  + Même expérience mais avec don pour des associations
  + + évaluation de la cause de l’association
  + Résultat :
    - Corrélation évaluation de la cause et don moyen
    - Toujours activation cortex ventro-médial
  + -> C’est bien la même régions quand on fait des cadeaux
* Que préférez vous ?
  + 100€ maintenant VS 100€ dans dix jours -> V(100, 0) > V(100, 10)
  + 10€ now VS 100€ 10 jours -> V(10, 0) > V(100, 10)
  + 90€ now VS 100€ 10 jours -> V(90, 0) = V(100,10)
  + Par dicotomie on peut retrouver la valeur exact
  + Pareil en augmentant le temps d’attente
  + -> Into **courbe de dévaluation** :
    - X = délais pour obtenir cette récompense
    - Y = valeur par rapport à un niveau de base
    - Hyperbole avec un paramètre de d’évaluation
    - Age, fatigue, fumeur/non fumeur
    - Au bout d’un moment les courbe se croise
* Plein d’exemple :
  + Pizza après-demain -> non merci -> 2 jours plus tard, devant la pizza -> ALLER LAAA
  + Soirée dans un mois -> non j’ai exam -> 1 mois plus tard, l’amis nous redemander -> ALLER LAAAA
  + -> Il faut avoir des stratégie pour ne plus pouvoir changer d’avis
* EXP Paul Glimer :
  + Quelle régions du cerveau encode cette évaluation temps/récompense ?
  + Même principe qu’au dessus
  + Résultat :
    - Striatum ventral + VMPMC
    - Superposition de la courbe de dé-évaluation du sujet + de l’activité hémodynamique -> même courbe !!
* EXP :
  + Peut-on prédire sur la base de l’activité cérébrale
  + Phase 1 :
    - IRM : cercle représentant une loterie, une proba de gagner l’argent
  + -> permet de localiser le VS et le VMPMC chez le sujet
  + Phase 2 :
    - IRM : regarder des stimulus, bien de consommation très varié, sans choix, évaluation relativement automatique
    - En dehors du scanner : Choix d’objet « vous allez repartir avec ça ou ça » : création d’une hiérarchie de préférence individuel
    - Prédiction sur ce qui vas se passer en dehors du scanner sur la base des données d’avant
  + Résultat :
    - Oui on peut prédire mais pas parfaitement
    - Intrasujet :
      * X = échelle de préférence (combien de fois ils ont gagné les duel)
      * Y = activation
      * R= corrélation -> Elles ne sont pas giga ouf
      * Corrélation entre la valeur déterminé avec le comportement
    - Groupe :
      * Comparaison entre l’item qui nous intéresse le plus
      * X = item numéros i
      * Y = distance avec l’activation de l’item qui nous intéresse le plus == prédiction
      * Attention niveau de la chance, il faut des distances élevés pour avoir un bon score
  + -> Activité de ces deux régions reflète nos choix
* Neuro marketing
  + Dérive des ces expériences
  + Exemple : choix de la couleur d’une boite
    - Faire une expérience comme ça couterai chère + beaucoup de temps -> pour une information pas giga fiable
    - On peut obtenir des meilleurs résultat avec des fixations oculaire ou bien en demandant

New Cours 23/03/2022

* Chez l’animal on étudie le Cortex orbito-frontal pour avoir la valeur du stimulus
* EXP :
  + Diagram

    Description automatically generatedPoint de fixation -> carrée (couleur, nombre) à gauche et à droite
  + Couleur représente un type de jus de fruit et le nombre la quantité
  + Chart, scatter chart

    Description automatically generatedSaccade oculaire pour choisir le jus
  + Résultats comportementaux :
    - Dans l’image B = carrée bleu
    - Pourcentage de choix vers le B
    - 1 :1 -> il préfère le B
  + Est-ce que on peut trouver des neurones qui code la valeur A et B ?
  + On maintient la valeur du jus A constante (on donne toujours une goutte, un carrée)
  + Record au moment de la décision aka présentation des carrée
  + Diagram

    Description automatically generatedCertain neurone code pour la valeur en elle-même, certain pour la différence entre les deux
* EXP Todd Hare
  + Sur les autre partie du cerveau qui ressorte dans la méta analyse
  + Sujet qui font un régime pour perdre du poids
  + Présentation d’un aliments 3 questions :
    - Pizza, salade
    - Est-ce bon pour la santé ? Est-ce que vous aimez le plat ? == Question contrôle
    - Souhaiteriez-vous en manger si on vous en donne maintenant ?
      * -> Sujet self-contrôle
      * Chart, bar chart

        Description automatically generated-> Sujet inverse
      * En fonction du plat healthy ou non
  + Résultat 1 :
    - SC = self contrôle, NCS = no self controle
    - J’suis mort les NSC qui prenne la bouffe qui aime pas, qu’elle soit healthy ou non
  + Résultat cerveau :
    - L’activité dans le cortex préfrontal ventro-médial est corrélé avec les choix du sujet
    - L’activité au niveau du cortex préfrontal dorsolatéral est plus élevé chez les SC que chez les sujets NSC
    - Corrélation négative entre l’activité du cortex préfrontal ventro-médial et celle du cortex préfrontal dorso-latéral
  + -> Le DLPF inhiberait l’activité du PFCVM pour réduire la valeur d’un stimulus apportant une forte satisfaction sur le court-terme mais ayant des conséquences moins souhaitable sur le long-terme
  + Interprétation :
    - SC ⬄ Gamma élevé (prise en compte des conséquences long terme) -> ce gamma serait reflété par le cortex préfrontal dorsolatéral.
* Autre truc sur les autre partie du cerveau qui ressorte dans la méta analyse
  + Amygdale =
  + Insula = lobe caché = pas très clair mais s’active pour l’anticipation de la récompense

Retour sur le big cadre général : il nous reste à parler du famoso « pleasure center »

* Comment le trouver ? on a qu’a suivre l’information qui arrive au comparateur (système dopaminergique)
  + Big graph, c’est pas trop possible en faite
* On a des expression facial quand on mange quelque chose d’acid ou de sucré, c’est pareil chez les mammifère
* Hotspot / Coldspot
  + On peut reproduire ces expressions en stimulant (chimiquement) des endroit très très précis (hotspot)
  + Effet renforçant / punitif (place renforcement learning)
  + Ils sont relié en chaine, si on en désactive un, les autres des autre zones sont désactivé également
  + Est-ce qu’il existe chez l’humain ? trop petit pour être détécté à l’irm
  + Est-ce que c’est que pour le gout ?
* Chez l’humain
  + Y’a des zones mais qui ne corresponde pas vraiment à ce qu’il y a chez le rat
  + Mais si lésion dans ces zones pas de symptome d’ahédonie, plutôt des problèmes de prise de décision ect

New Cours :

* On doit pas apprendre ce cours on doit y sparkle réfléchir sparkle
* Utilité
  + Utilité prédite : Valeur du stimulus = V(s) =
  + Utilité vécus :
* On peut considérer = P-N
  + P = Stimulus appétitif = plaisir = hot spot
  + N = Stimulus aversif = déplaisir = cold spot
  + Pour tout stimulus on peut le séparer en deux partie comme ça, plaisir // déplaisir
* On peut penser que c’est les dernières découverte mais en faite c’est des idées qu’on retrouve chez les philosophes depuis très longtemps
  + Si on trace l’idée d’utilité on la retrouve
    - En psychologie de l’apprentissage : stimulus appétitif / aversif, tout est blanc ou noir, soit on s’en approche ou on s’en éloigne
    - En économie : les consommateurs sont censés maximiser leur utilité
  + Le point commun : Jeremy Bentham : fondateur du **courant philosophique appelé utilitarisme**
    - Plutôt en avance, il voulait le droit de vote pour tout le monde (femme) plus d’un siècle avant que ça soit implémenter. Pour l’éducation pour tout le monde. Il voulait réformer la prison avec le concept de panoptique
    - Quelqu’un qui voulait changer la société.
    - Son squelette est exposé avec modelé en cire, car son pote médecin l’a momifié
    - Pour lui chaque être humain cherche à minimiser la quantité de douleur et maximiser le plaisir -> Utilité = différence entre plaisir et douleur
    - Meilleurs société = celle qui maximise l’utilité
    - Il disait aussi qu’il n’y avait pas de plaisir plus noble qu’un autre, le plaisir de regarder un tableau est le même que celui de manger
    - Limite
      * Est-ce que maximiser le plaisir c’est maximiser le bonheur
      * Est-ce que faire juste la moyenne pour une société c’est suffisant ? unmec avec plaisir + inf et les autre à 0
      * Il faut que le plaisir et la douleur soit dans la même unité et comparable
        + Douleur = on ne sais pas forcement sa cause, deux cause peuvent donner la même sensation, on peut se tromper sur la cause
        + Donc il envisage le plaisir comme une sensation avec les même paramètre que la douleur :

J’ouvre le frigo pour un sandwish, pas ce qu’il faut, du coup pour avoir le plaisir du sandwish je vais écouter de la musique

On connait l’origine : J’écoute un concerto de mozard dans un giga fauteuil en mangeant un sandwish. On ne saurait pas ce qu’il nous procure le plaisir parmi les trois.

Je pensais que c’était le sandwish mais en faite c’était la musique

* + - * + -> Tout cela n’a pas de sens -> le plaisir ne peut pas être considéré comme une sensation.
        + Le plaisir en écoutant de la musique c’est le plaisir qu’on a en écoutant de la musique, pas le même que celui de manger -> le plaisir est toujours attribuer avec quelque chose, c’est pas interchangeable, c’est pas la même chose. Donc les comparer n’a pas vraiment de sens
      * -> le plaisir serait qualitatif en faite
      * Mais Bentham a ancrer le concept que c’était quantitatif
  + Donc l’idée de centre de plaisir n’est pas cohérente
    - Il faudrait une infinité de bouton plaisir pour chaque stimulus
    - On n’y les problèmes qualitatifs du plaisir et les réduit a des problème quantitatif
  + -> Mais l’idée revient doucement en neuroscience, par le fait que les ND réponde différemment en fonction de la qualité du stimulus
* EXP
  + SC visuel (symbole) -> SI olfactif
  + SI olfactif : valence élevé//moins élevé / sucré // salé
  + Phase d’apprentissage pavlovien
  + Puis IRMf avec seulement les SC visuel
  + Quel région du cerveau code la valence du SI ?
  + On train un NN avec les exemple salé pour qu’il prédise ensuite si c’est salé et sucrée
  + Puis on regarde quel morceau du cerveau arrive le mieux à prédire la valence
  + -> Le cortex ventro médial !! Consistent avec ce qu’on a vu avant
  + Same avec : Quel région du cerveau code l’identité du SI (sucré ou salé)
  + -> Le cortex orbito frontale
    - C’est la même région qu’on étudie chez le singe
    - Peut de recherche sur ce cortex chez l’humain
    - Prendrait en compte la qualité du stimulus (sucrée salée ou sandwish musique)