

ACTIVITÉ 2



L'Arène

Apprentissage par renforcement

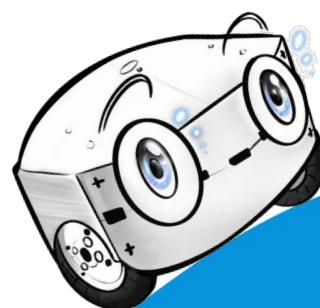
Manuel du professeur	p1
Manuel de l'élève	p11
Présentation de classe	p21



Auteur
Joël RIVET

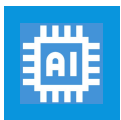


Graphiste
David COHEN



ACTIVITÉ 2

L'Arène



**Apprentissage
par renforcement**



Difficulté
Moyenne



Temps estimé
60 min



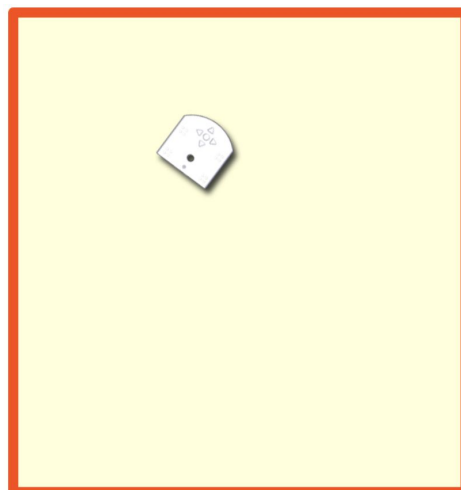
Prérequis

- **Connaître Thymio.**
- Avoir fait l'activité ThymioAI **Premier apprentissage.**



Préparation du matériel

- Disposer d'une enceinte rectangulaire faite de murs solidaires ou d'un espace délimité par des objets suffisamment lourds pour que Thymio ne puisse pas les déplacer.
- Dimensions minimales : 50cm x 50 cm.
- Disposer le robot Thymio à l'intérieur de l'arène.



ACTIVITÉ 2

L'Arène



Mission de Thymio et le principe de son apprentissage

- Thymio se trouve à l'intérieur d'une arène. Son but est d'explorer cette arène sans toucher les murs. Comme d'habitude, au début, Thymio ne sait rien faire.
- Nous décidons de lui apprendre à effectuer sa mission avec une nouvelle méthode appelée **apprentissage par renforcement**. C'est une méthode qui fait aussi partie du domaine de l'intelligence artificielle.
- Nous utiliserons à nouveau un réseau de neurones.
- Dans cette méthode, on ne dit pas à Thymio quelle action il doit faire, parmi les diverses actions possibles. C'est à lui d'en choisir une. Mais comment le robot peut-il savoir quelle action choisir ? Grâce à des récompenses qu'on lui donnera, ou non^[1].



L'apprentissage par renforcement

- Notre nouvelle manière d'entraîner Thymio consistera à lui indiquer si l'action qu'il a choisie, pour une situation des capteurs donnée, est un bon ou un mauvais choix en lui donnant pour cette action une récompense ou une pénalité.
- Le réseau de neurones prend donc connaissance de ces récompenses pour chaque action effectuée. Comment le réseau va-t-il réagir ? Comme le ferait un être humain ou un animal, il va chercher à obtenir le plus de récompenses possible et à fuir les pénalités.
- Le réseau utilise les mathématiques pour progresser. Si on doutait de l'utilité des mathématiques, en voilà un bon exemple d'application.
- Petit à petit, Thymio va effectuer de plus en plus souvent les "bonnes" actions (celles qui rapportent des récompenses), et de moins en moins souvent les "mauvaises".
- Pour découvrir les bonnes ou mauvaises actions, le robot doit toutes les explorer.
- Donc, de temps en temps, le robot tentera de nouvelles actions, sans tenir compte des récompenses.



Youpi !

Action: Avancer
Capteur: rien devant

Récompense accordée!



Ouille !

Action: Avancer
Capteur: mur devant

Pénalité infligée!

^[1] Cela évoque le jeu de devinette, "t'es froid, t'es chaud, tu brûles ! ..."

ACTIVITÉ 2



Réglages alphi

Rappel pour connecter Thymio

1. **Connecter Thymio** avec une des 2 méthodes suivantes :
 - soit relier le robot à l'ordinateur à l'aide du câble USB
 - soit brancher la clé USB et allumer le robot
2. **Lancer le logiciel Thymio Suite** et attendre quelques instants.
3. **Lancer le logiciel Alphi.**
4. Dans **Alphi**, **sélectionnez** votre **robot** qui apparaît sur l'écran bleu pour **établir la connexion**.

Réglages

- Les réglages sont présents dans un fichier que nous allons charger.
- Dans le menu **paramètres**, sélectionner l'option charger des **paramètres d'exemples ...**
- Dans le cadre qui apparaît, double cliquer sur l'option **évitement d'obstacles** (apprentissage par renforcement).
 - Au dessous du réseau on remarque deux barres noires de progression.
 - La barre de gauche affiche les récompenses ou pénalités sous forme de nombre. Une récompense est représentée par un nombre positif, une valeur négative correspond à une pénalité. Les valeurs possibles pour les récompenses sont déjà fixées par Alphi.
 - La barre droite indique le niveau, c'est-à-dire la moyenne des récompenses précédentes. Il correspond à l'état de l'apprentissage de Thymio.
- Par ailleurs, on laissera les boutons **apprentissage** et **exploration** activés.

ACTIVITÉ 2

L'Arène



Un premier apprentissage

- **Observer Thymio dans l'arène**

Lancer le mode **autonome**. Thymio se met en mouvement. Son premier déplacement est choisi au hasard. Observer ses mouvements et ses changements de couleur. Cliquer à nouveau sur le bouton **autonome** au bout de 20 secondes environ pour arrêter l'apprentissage.

Question 1: établir des liens entre ce que fait le robot et les couleurs qu'il prend. Noter la réponse.

Quand Thymio se déplace en avant dans l'arène, il est vert. Quand il s'approche trop près ou qu'il touche un mur, il se colore en rouge.

Vert veut dire : Thymio reçoit une récompense, son action va dans le sens de sa mission.

Rouge veut dire : Thymio reçoit une pénalité, son action est contraire à sa mission.

Poursuivre l'apprentissage pendant 5 à 10 minutes.

Question 2: constatez-vous une évolution dans le comportement de Thymio ? Si oui, la décrire. Noter la réponse.

On observe qu'au début, Thymio heurte assez souvent les murs et peine à s'en dégager.

Puis il réussit à les éviter de plus en plus souvent. On peut même noter plusieurs étapes dans l'apprentissage :

- Au début, Thymio a très vite découvert qu'il ne fallait pas aller en arrière.
- Ensuite il a adopté très vite un des deux comportements suivants (les élèves observeront l'un ou l'autre avec leur Thymio) : soit tourner en rond (c'est le plus courant), soit faire des lignes droites et se retourner lorsqu'il cogne un mur.
- Puis il apprend petit à petit à alterner entre tout droit et virage, choisissant de plus en plus fréquemment tout droit lorsqu'il n'y a pas d'obstacle devant, virage à gauche s'il y a un mur à droite, et virage à droite s'il y a un mur à gauche.

ACTIVITÉ 2

L'Arène



Le rôle de l'exploration

- Cliquer sur le bouton **exploration** pour le désactiver.

Question 3: constatez-vous une évolution dans le comportement de Thymio ? Si oui, la décrire. Noter la réponse.

Thymio fait moins d'erreur, il n'interrompt plus ses lignes droites avec des mouvements inopinés.

Réinitialiser l'IA avec le bouton **réinitialiser l'IA**. Cela fait « oublier » à Thymio tout ce qu'il a appris, il recommence un apprentissage de zéro (garder **autonome** activé, mais **exploration** désactivé).

Question 4: constatez-vous des différences entre ce nouvel apprentissage et le précédent ? Si oui, les décrire. Si vous ne constatez pas de différence, recommencer encore un nouvel apprentissage en appuyant sur **réinitialiser l'IA**. Noter la réponse.

Thymio reste bloqué dans le comportement « tourner en rond » sans découvrir la ligne droite. (Cela n'est pas systématique et il lui arrive d'apprendre correctement même lorsque l'exploration est désactivée).

Conclusion

L'exploration est indispensable à l'apprentissage.

- l'IA essaie de temps en temps des actions autres que celle qu'elle « pense » être la meilleure (lorsque cela arrive l'icône d'action à droite de l'écran s'allume en bleu plutôt qu'en noir). Cela évite de rester bloqué dans un comportement médiocre.
- En revanche, une fois l'apprentissage terminé, l'exploration n'est plus utile, il est intéressant de la désactiver pour obtenir le comportement le plus parfait possible.

ACTIVITÉ 2



Le comportement des réseaux de neurones

Observer le comportement du réseau de neurones

Nous allons observer et noter le comportement en détail sur quelques étapes, au début de l'apprentissage.

1. Réinitialiser l'IA avec le bouton **réinitialiser l'IA**. Placer Thymio au milieu de l'arène. On rappelle que son premier déplacement est choisi au hasard. Pour vous en convaincre vous pouvez cliquer plusieurs fois de suite sur **réinitialiser l'IA** et **autonome**.
2. Compléter la première ligne du tableau. Le petit tiret signifie que les capteurs avant de Thymio ne captent rien car il n'y a rien.
3. Bien regarder Thymio et cliquer sur le bouton **pas à pas**.

Etape pas à pas	Capteurs avant	Reculer à gauche	Avancer à gauche	Avancer	Reculer à droite	Avancer à droite	Valeur récompense	Valeur niveau

ACTIVITÉ 2

L'Arène

Question 5: quel mouvement Thymio a-t-il fait ? Dans la ligne du tableau que vous venez de remplir, relevez la valeur correspondante à ce mouvement. Comparer ce nombre aux valeurs des autres actions. Que constate-t-on ? Noter la réponse.

Cette valeur est la plus grande.

Donc, Thymio effectue le mouvement correspondant à la valeur la plus grande.

Question 6: on observe aussi que le robot a reçu une récompense. Cette récompense semble-t-elle en accord avec le but de la mission ? Expliquer.

Réponse possible : Thymio a tourné à gauche, et il a reçu une récompense de +55. C'est normal, car il n'y a rien devant lui et il peut donc tourner.

Une fois la première récompense décernée, les valeurs de sortie sont re-calculées par le réseau de neurones.

Compléter la 2ème ligne du tableau et devinez alors quel sera le prochain mouvement de Thymio.

Faire quelques clics supplémentaires sur le bouton **pas à pas** en regardant évoluer les récompenses et le niveau.

ACTIVITÉ 2



Le rôle des récompenses

Observer le comportement du réseau de neurones

Nous allons observer et noter le comportement en détail sur quelques étapes, au début de l'apprentissage.

1. Réinitialiser l'IA avec le bouton **réinitialiser l'IA**. Placer Thymio au milieu de l'arène. On rappelle que son premier déplacement est choisi au hasard. Pour vous en convaincre vous pouvez cliquer plusieurs fois de suite sur **réinitialiser l'IA** et **autonome**.
2. Compléter la première ligne du tableau. Le petit tiret signifie que les capteurs avant de Thymio ne captent rien car il n'y a rien.
3. Bien regarder Thymio et cliquer sur le bouton **pas à pas**.

Etape pas à pas	Capteurs avant	Reculer à gauche	Avancer à gauche	Avancer	Reculer à droite	Avancer à droite	Valeur récompense	Valeur niveau

ACTIVITÉ 2

L'Arène

Question 7: comment évolue le niveau quand le robot reçoit une récompense ou au contraire une pénalité ? Que représente le niveau ?

Noter la réponse.

- Si la récompense est positive, le niveau augmente.
- Si la récompense est négative, le niveau diminue.

Le niveau représente la capacité de Thymio à obtenir des récompenses positives. Plus précisément, il est calculé comme la moyenne des récompenses reçues pendant la dernière minute écoulée.

À présent, appuyer sur autonome pour laisser Thymio continuer son apprentissage.

Question 8 : comment évolue le niveau au cours de l'apprentissage ?

Pourquoi ? Expliquer :

Le niveau augmente au cours de l'apprentissage. En effet, Thymio reçoit de plus en plus de récompenses élevées (en particulier lorsqu'il va tout droit) et de moins en moins de punitions (puisqu'il se cogne de moins en moins). En fait, le but de l'apprentissage est précisément de faire augmenter le niveau de Thymio.

ACTIVITÉ 2

L'Arène



Influence de la pénalité sur le comportement appris

Si on résume les différentes valeurs apparues dans la barre de progression, on a relevé :

- **100** : Quand Thymio avance tout droit sans obstacle devant, c'est la valeur la plus élevée.
- **55** : Quand Thymio tourne sans obstacle devant.
- **-50** : Thymio effectue une des multiples "mauvaises" actions comme avancer contre un mur, reculer quand il n'y a rien devant ...

Nous pouvons modifier la valeur de la pénalité.

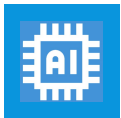
- Ouvrir l'onglet récompense et mettre la pénalité à une petite valeur, par exemple 0 : réinitialiser l'IA et relancer l'apprentissage pendant quelques minutes. Observer le comportement de Thymio, est-il plus audacieux, plus prudent ? Tape-t-il plus ou moins souvent les murs ?
- Mettre une plus grande pénalité, 1,5 par exemple. Réinitialiser à nouveau l'IA et relancer l'apprentissage pendant quelques minutes. Même question que précédemment.

Question 9: résumer comment le comportement de Thymio évolue quand on modifie la valeur de la pénalité. Notez la réponse.

- **Si la pénalité est faible**, Thymio tape souvent les murs mais devient plus audacieux et parcourt pleinement les lieux.
- **Si la pénalité est forte**, Thymio tape moins souvent les murs mais devient plus prudent et reste dans une zone restreinte.

ACTIVITÉ 2

L'Arène



Apprentissage
par renforcement



Difficulté
Moyenne

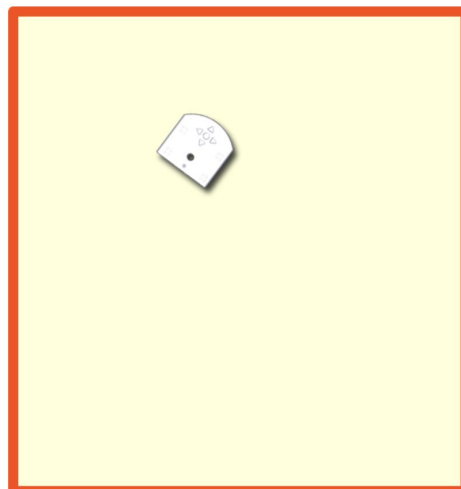


Temps estimé
60 min



Mission de Thymio et le principe de son apprentissage

- Thymio se trouve à l'intérieur d'une arène. Son but est d'explorer cette arène sans toucher les murs. Comme d'habitude, au début, Thymio ne sait rien faire.
- Nous décidons de lui apprendre à effectuer sa mission avec une nouvelle méthode appelée apprentissage par renforcement. C'est une méthode qui fait aussi partie du domaine de l'intelligence artificielle.
- Nous utiliserons à nouveau un réseau de neurones.
- Dans cette méthode, on ne dit pas à Thymio quelle action il doit faire, parmi les diverses actions possibles. C'est à lui d'en choisir une. Mais comment le robot peut-il savoir quelle action choisir ? Grâce à des récompenses qu'on lui donnera, ou non^[1].



^[1] Cela évoque le jeu de devinette, "t'es froid, t'es chaud, tu brûles ! ..."

ACTIVITÉ 2



L'apprentissage par renforcement

- Notre nouvelle manière d'entraîner Thymio consistera à lui indiquer si l'action qu'il a choisi, pour une situation des capteurs donnée, est un bon ou un mauvais choix en lui donnant pour cette action une récompense ou une pénalité.
- Le réseau de neurones prend donc connaissance de ces récompenses pour chaque action effectuée. Comment le réseau va-t-il réagir ? Comme le ferait un être humain ou un animal, il va chercher à obtenir le plus de récompenses possible et à fuir les pénalités.
- Le réseau utilise les mathématiques pour progresser. Si on doutait de l'utilité des mathématiques, en voilà un bon exemple d'application.
- Petit à petit, Thymio va effectuer de plus en plus souvent les "bonnes" actions (celles qui rapportent des récompenses), et de moins en moins souvent les "mauvaises".
- Pour découvrir les bonnes ou mauvaises actions, le robot doit toutes les explorer.
- Donc, de temps en temps, le robot tentera des actions nouvelles, sans tenir compte des récompenses.



Youpi !

Action: Avancer
 Capteur: rien devant
Récompense accordée!



Ouille !

Action: Avancer
 Capteur: mur devant
Pénalité infligée!

ACTIVITÉ 2



Réglages alphi

Rappel pour connecter Thymio

1. **Connecter Thymio** avec une des 2 méthodes suivantes :
 - soit relier le robot à l'ordinateur à l'aide du câble USB
 - soit brancher la clé USB et allumer le robot
2. **Lancer le logiciel Thymio Suite** et attendre quelques instants.
3. **Lancer le logiciel Alphi.**
4. Dans **Alphi**, **sélectionnez** votre **robot** qui apparaît sur l'écran bleu pour **établir la connexion**.

Réglages

- Les réglages sont présents dans un fichier que nous allons charger.
- Dans le menu **paramètres**, sélectionner l'option charger des **paramètres d'exemples ...**
- Dans le cadre qui apparaît, double cliquer sur l'option **évitement d'obstacles** (apprentissage par renforcement).
 - Au dessous du réseau on remarque deux barres noires de progression.
 - La barre de gauche affiche les récompenses ou pénalités sous forme de nombre. Une récompense est représentée par un nombre positif, une valeur négative correspond à une pénalité. Les valeurs possibles pour les récompenses sont déjà fixées par Alphi.
 - La barre droite indique le niveau, c'est-à-dire la moyenne des récompenses précédentes. Il correspond à l'état de l'apprentissage de Thymio.
- Par ailleurs, on laissera les boutons "**apprentissage**" et "**exploration**" activés.

ACTIVITÉ 2

L'Arène



Un premier apprentissage

- Observer Thymio dans l'arène

Lancer le mode **autonome**. Thymio se met en mouvement. Son premier déplacement est choisi au hasard. Observer ses mouvements et ses changements de couleur. Cliquer à nouveau sur le bouton **autonome** au bout de 20 secondes environ pour arrêter l'apprentissage.

Question 1 : établir des liens entre ce que fait le robot et les couleurs qu'il prend. Noter la réponse .

Poursuivre l'apprentissage pendant 5 à 10 minutes.

Question 2 : constatez-vous une évolution dans le comportement de Thymio ? Si oui, la décrire. Noter la réponse.



Le rôle de l'exploration

- Cliquer sur le bouton **exploration** pour le désactiver.

Question 3 : constatez-vous une évolution dans le comportement de Thymio ? Si oui, la décrire. Noter la réponse .

Réinitialiser l'IA avec le bouton réinitialiser l'IA. Cela fait « oublier » à Thymio tout ce qu'il a appris, il recommence un apprentissage de zéro (garder autonome activé, mais exploration désactivé).

Question 4 : constatez-vous des différences entre ce nouvel apprentissage et le précédent ? Si oui, les décrire. Si vous ne constatez pas de différence, recommencer encore un nouvel apprentissage en appuyant sur **réinitialiser l'IA**. Noter la réponse.

ACTIVITÉ 2



Le comportement des réseaux de neurones

Observer le comportement du réseau de neurones

Nous allons observer et noter le comportement en détail sur quelques étapes, au début de l'apprentissage.

1. Réinitialiser l'IA avec le bouton **réinitialiser l'IA**. Placer Thymio au milieu de l'arène. On rappelle que son premier déplacement est choisi au hasard. Pour vous en convaincre vous pouvez cliquer plusieurs fois de suite sur **réinitialiser l'IA** et **autonome**.
2. Compléter la première ligne du tableau. Le petit tiret signifie que les capteurs avant de Thymio ne captent rien car il n'y a rien.
3. Bien regarder Thymio et cliquer sur le bouton **pas à pas**.

Etape pas à pas	Capteurs avant	Reculer à gauche	Avancer à gauche	Avancer	Reculer à droite	Avancer à droite	Valeur récompense	Valeur niveau

ACTIVITÉ 2

L'Arène

Question 5 : quel mouvement Thymio a-t-il fait ? Dans la ligne du tableau que vous venez de remplir, relevez la valeur correspondante à ce mouvement. Comparer ce nombre aux valeurs des autres actions. Que constate-t-on ? Noter la réponse.

Question 6 : on observe aussi que le robot a reçu une récompense. Cette récompense semble-t-elle en accord avec le but de la mission ? Expliquer.

Une fois la première récompense décernée, les valeurs de sortie sont re-calculées par le réseau de neurones.

Compléter la 2ème ligne du tableau et devinez alors quel sera le prochain mouvement de Thymio.

Faire quelques clics supplémentaires sur le bouton **pas à pas** en regardant évoluer les récompenses et le niveau.

Question 7 : comment évolue le niveau quand le robot reçoit une récompense ou au contraire une pénalité ? Que représente le niveau ? Noter la réponse.

À présent, appuyer sur autonome pour laisser Thymio continuer son apprentissage.

Question 8 : comment évolue le niveau au cours de l'apprentissage ? Pourquoi ? Expliquer.

ACTIVITÉ 2

L'Arène

**Influence de la pénalité sur le comportement appris**

Si on résume les différentes valeurs apparues dans la barre de progression, on a relevé :

- **100** : quand Thymio avance tout droit sans obstacle devant, c'est la valeur la plus élevée.
- **55** : quand Thymio tourne sans obstacle devant.
- **-50** : Thymio effectue une des multiples "mauvaises" actions comme avancer contre un mur, reculer quand il n'y a rien devant ...

Nous pouvons modifier la valeur de la pénalité.

- Ouvrir l'onglet récompense et mettre la pénalité à une petite valeur, par exemple 0 : réinitialiser l'IA et relancer l'apprentissage pendant quelques minutes. Observer le comportement de Thymio, est-il plus audacieux, plus prudent ? Tape-t-il plus ou moins souvent les murs ?
- Mettre une plus grande pénalité, 1,5 par exemple. Réinitialiser à nouveau l'IA et relancer l'apprentissage pendant quelques minutes. Même question que précédemment.

Question 9 : résumer comment le comportement de Thymio évolue quand on modifie la valeur de la pénalité. Notez la réponse

ACTIVITÉ 2

L'Arène



Réponses

Nom:

Classe:

Question 1 : établir des liens entre ce que fait le robot et les couleurs qu'il prend.

[]

Question 2 : constatez-vous une évolution dans le comportement de Thymio ? Si oui, la décrire.

[]

Question 3 : constatez-vous une évolution dans le comportement de Thymio ? Si oui, la décrire.

[]

ACTIVITÉ 2

L'Arène



Réponses

Nom:

Classe:

Question 4 : constatez-vous des différences entre ce nouvel apprentissage et le précédent ?

[]

Question 5 : quel mouvement Thymio a-t-il fait ? Dans la ligne du tableau que vous venez de remplir, relevez la valeur correspondante à ce mouvement. Comparer ce nombre aux valeurs des autres actions. Que constate-t-on ?

[]

Question 6 : on observe aussi que le robot a reçu une récompense. Cette récompense semble-t-elle en accord avec le but de la mission ? Compléter la 2ème ligne du tableau et devinez alors quel sera le prochain mouvement de Thymio.

[]

ACTIVITÉ 2

L'Arène



Réponses

Nom:

Classe:

Question 7 : comment évolue le niveau quand le robot reçoit une récompense ou au contraire une pénalité ? Que représente le niveau ?

[]

Question 8 : comment évolue le niveau au cours de l'apprentissage ? Pourquoi ?

[]

Question 9: résumer comment le comportement de Thymio évolue quand on modifie la valeur de la pénalité.

[]

L'Arène



La mission

- Thymio se trouve à l'intérieur d'une arène fermée.
- Son but est **d'explorer** cette arène sans toucher les murs.



Apprentissage renforcé

- Chaque fois que **Thymio exécute une action**, on lui **indique** si l'action qu'il a fait, pour une situation des **capteurs donnée**, est un **bon** ou un **mauvais choix** grâce à des **récompenses** ou des **pénalités**.
- Thymio choisira ses **actions** de sorte à **obtenir un maximum de récompenses**.



Youpi !



Action: Avancer

Capteur: rien devant

Récompense accordée!



Ouille !



Action: Avancer

Capteur: mur devant

Pénalité infligée!



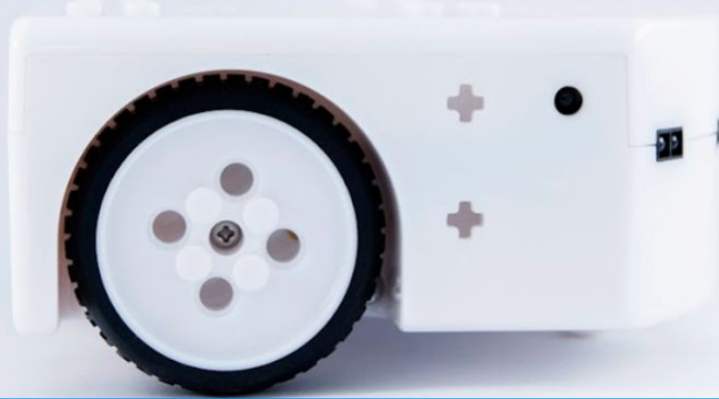
Connexion - Réglages

1. **Connecter Thymio** avec une des 2 méthodes suivantes :
 - soit relier le robot à l'ordinateur à l'aide du câble USB
 - soit brancher la clé USB et allumer le robot
2. **Lancer le logiciel Thymio Suite** et attendre quelques instants.
3. **Lancer le logiciel Alphai.**
4. Dans **Alphai**, sélectionnez votre **robot** qui apparaît sur l'écran bleu pour **établir la connexion.**



Tableau pas à pas

Etape pas à pas	Capteurs avant	Reculer à gauche		Avancer à gauche		Avancer		Reculer à droite		Avancer à droite		Valeur récompense	Valeur niveau



www.thymio.org

thymio
by MÖBSYA

Chemin du Closel 3, 1020 Renens - Switzerland
info@thymio.org