

UN MODÈLE ÉTHIQUE POUR LES MDP MULTI-AGENTS

Mihail Stojanovski¹

Encadrement : Nadjat Bourdache¹, Grégory Bonnet¹, Abdel-Allah Mouaddib¹

¹Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, Caen, FRANCE

{prénom.nom}@unicaen.fr



1. Introduction
2. Modélisation d'une théorie morale
3. Jugement éthique dans une modèle de décision multi-agent
4. Expérimentation

Branches de l'éthique

- ▶ Métaéthique → réflexions philosophiques dévolue à l'être humain
- ▶ Éthique appliquée → réflexions et choix dévolus à l'être humain
- ▶ **Éthique normative** → décrit des procédures fondées sur des théories morales

Théorie morale [Timmons, 2012]

L'objectif pratique d'une théorie morale est de fournir une procédure de décision dont l'utilisation par des agents convenablement informés les conduira de manière fiable à des jugements moraux corrects dans des contextes de délibération et de décision.

Modèle de prise de décision éthique

- ▶ Modéliser des valeurs morales
- ▶ Evaluer un comportement au vu d'une valeur morale
- ▶ Agreger des évaluations pour arriver à une décision

	Décision séquentielle sous l'incertitude	Valeurs et théories morales	Exemples
Modèles fondés sur la logique et l'argumentation	-	+	[Bench-Capon, 2003] [Cointe et al., 2016] [Berreby et al., 2017] [Grandi et al., 2022]
Modèles fondés sur la théorie de la décision	+	-	[Abel et al., 2016] [Svegliato et al., 2021] [De Moura et al., 2020] [Rodriguez-Soto et al., 2020]

Contribution

Concevoir un modèle de décision séquentielle expressif intégrant une représentation riche des valeurs et des théories morales et prenant en compte des **notions d'incertitude** et la **subjectivité de différents agents**.

1. Introduction

2. Modélisation d'une théorie morale

3. Jugement éthique dans une modèle de décision multi-agent

4. Expérimentation

Valeurs morales

- ▶ Valeur morale : honnêteté, loyauté, générosité, imprudence
- ▶ Polarité d'une valeur
- ▶ Poids d'une valeur

Alignement d'un comportement

- ▶ Être aligné sur une valeur (promotion)
- ▶ Être opposé sur une valeur (trahison)
- ▶ Degré d'un alignement

Système des valeurs morales au plusieurs niveaux d'importance

- ▶ Les valeurs d'un même niveau peuvent être de polarité différente.
- ▶ Les valeurs de même niveau peuvent se compenser.

Évaluation morale

- ▶ Comment juger un comportement?

Éléments de l'exemple

Anna découvre que Bob ment à Claire. Elle hésite à le dire à Claire.

- Les valeurs morales en jeu :
 1. Vérité, Loyauté,
 2. Respect de la vie privée
- Niveaux d'importance possibles :
 - ▶ 1er niveau : Vérité et Respect de la vie privée
 - ▶ 2e niveau : Loyauté

Anna examine la possibilité de dire la vérité

- ▶ Bob et Claire n'agissent pas mais en subissent les conséquences.
- ▶ Bob considère que la vérité est moins importante que le respect de la vie privée.
- ▶ Si Anna prend en compte les valeurs de Bob, elle ne dira pas à Claire.

1. Introduction

2. Modélisation d'une théorie morale

3. Jugement éthique dans une modèle de décision multi-agent

4. Expérimentation

Jeux stochastiques [Shapley, 1953]

Un jeu stochastique est une tuple $\langle S, Ag, \{A_i\}_{i \in Ag}, \mathcal{T}, \{R^i\}_{i \in Ag} \rangle$ où:

- ▶ S est l'ensemble des états
- ▶ Ag est l'ensemble des agents
- ▶ A_i est l'ensemble des actions de l'agent i
- ▶ $\mathcal{T} : S \times A_1 \times \dots \times A_{|Ag|} \times S \rightarrow [0, 1]$ est la fonction de transition
- ▶ $R^i : S \times A_1 \times \dots \times A_{|Ag|} \times S \rightarrow \mathbb{R}$ est la fonction de récompense pour l'agent i

Comment intégrer de l'éthique normative dans les jeux stochastiques ?

- ▶ Définir un contexte éthique relatif à chaque agent
- ▶ Modéliser des fonctions de récompenses distinctes fondées sur les valeurs
- ▶ Introduire une notion d'*empathie* entre les agents
- ▶ Concevoir un algorithme de résolution

Contexte éthique

- ▶ $\mathcal{V} = v_1, \dots, v_k$ un ensemble de valeurs morales
- ▶ $\Delta^i = \Delta_1^i \dots \Delta_l^i$ un ensemble de niveaux d'importance pour l'agent i
- ▶ $\mathcal{C}^i = [\mathcal{C}_{h,j}^i \in \mathbb{R}]$ est le *contexte éthique* qui donne le poids et la polarité de v_j au niveau h pour i
- ▶ $\mathcal{C}_{h,j}^i > 0$: v_j dans le niveau Δ_h^i est une valeur positive pour l'agent i
- ▶ $\mathcal{C} = \bigcup_{i=1}^{|Ag|} \mathcal{C}^i = \{\mathcal{C}^1, \dots, \mathcal{C}^{|Ag|}\}$ est le *contexte éthique collectif*

α^i	Vérité	Respect de la vie privée	Loyauté
Δ_1^i	+2	+1	0
Δ_2^i	0	0	+1

Alignement moral de la transition

$\mathcal{E}(s, a_1, \dots, a_{|Ag|}, s')_j^i \in \mathbb{R}$ est l'expression d'un degré de promotion (> 0) ou de trahison (< 0) de la valeur v_j pour l'agent i . Si $\mathcal{E}(s, a_1, \dots, a_{|Ag|}, s')_j^i = 0$, la transition est moralement neutre vis-à-vis de la valeur v_j .

Fonctions de récompense

- ▶ Déterminer la récompense d'une action jointe pour un agent
- ▶ Déterminer une agrégation des récompenses d'un groupe d'agents
- ▶ Agréger les deux notions précédents

Fonction de récompense individuelle

Un tuple qui désigne le bien (∇) et le mal ($\bar{\nabla}$) causés pour le niveau h et l'agent i :

$$\nabla_h^i = \sum_{x_{h,j}^i \in \Delta_h^i} G_{h,j}^i \text{ t.q. } G_{h,j}^i = \begin{cases} x_{h,j}^i \cdot \mathcal{E}(s, a_1, \dots, a_{|Ag|}, s')_j^i & \text{si } x_{h,j}^i \cdot \mathcal{E}(s, a_1, \dots, a_{|Ag|}, s')_j^i > 0, \\ 0 & \text{sinon.} \end{cases}$$

$$\bar{\nabla}_h^i = \sum_{x_{h,j}^i \in \Delta_h^i} |B_{h,j}^i| \text{ t.q. } B_{h,j}^i = \begin{cases} x_{h,j}^i \cdot \mathcal{E}(s, a_1, \dots, a_{|Ag|}, s')_j^i & \text{si } x_{h,j}^i \cdot \mathcal{E}(s, a_1, \dots, a_{|Ag|}, s')_j^i < 0, \\ 0 & \text{sinon.} \end{cases}$$

Fonction de récompense éthique collective : $R^J = \langle R_{\nabla}^J, R_{\bar{\nabla}}^J \rangle$

Pour un sous-ensemble $J \in Ag$ et un niveau h :

$$R_{h,\nabla}^J = \prod_{j \in J} 1 + R_{h,\nabla}^j \quad \text{et} \quad R_{h,\bar{\nabla}}^J = \prod_{j \in J} 1 + R_{h,\bar{\nabla}}^j$$

Récompense éthique emphatique individuelle

Pour l'agent i et le niveau h , \mathfrak{R}_h^i , est définie pour $\star \in \{\nabla_h, \bar{\nabla}_h\}$ telle que :

$$\mathfrak{R}_h^i = R_{\star}^i + \sum_{J \subseteq Ag \setminus \{i\}} (1 - \lambda_J^i) R_{\star}^i + \lambda_J^i R_{\star}^J$$

Avec $\lambda_J^i \in [0, 1]$ l'empathie de l'agent i envers le groupe J .

Fonction de valeur avec h le niveau d'importance

$$V_h^{i,\pi}(s) = \max_{\{a_1, \dots, a_{|Ag|}\} \in A} \sum_{s' \in S} p(s'|s, a_1, \dots, a_{|Ag|}) \left(\mathfrak{R}_h^i(s, a_1, \dots, a_{|Ag|}, s') + \gamma V_h^{i,\pi}(s') \right)$$

Politiques optimales

Soit Π l'espace politique, $\text{Eval} : \Pi \rightarrow \mathbb{R}$ la fonction qui évalue le comportement des agents dépendant de leur alignement sur des valeurs dans la politique π .

$$\pi_{\nabla_h^i}^* = \operatorname{argmax}_{\pi'} V_{\nabla_h^i}^{\pi'}(s) - V_{\nabla_h^i}^{\pi}(s) \quad \text{avec} \quad \pi' \in \pi_{\nabla_{h-1}^i}$$

Résoudre un jeu stochastique éthique

Adaptation de l'algorithme JESP [Nair et al., 2003] produisant des équilibres de Nash

Algorithm 1: Ethical JESP

Data: E-SG Γ , Contexte \mathcal{C}

Result: Politique optimale π

$\pi \leftarrow$ politique jointe aléatoire pour Γ ; $\text{conv} \leftarrow 0$

while $\text{conv} \neq n - 1$ **do**

for $i \leftarrow 1$ à n **do**

$\pi^* \leftarrow \text{bestResponse}(i, \pi_{-i})$

$\text{amelioration} \leftarrow \text{false}$

for $h \leftarrow 1$ à $|\Delta|$ **do**

if $(V^{\Delta h}(\pi^*, i) > V^{\Delta h}(\pi, i))$ **then**

$\text{amelioration} \leftarrow \text{true}$

else if $(V^{\Delta h}(\pi^*, i) < V^{\Delta h}(\pi, i))$ **then**

break

if $\text{amelioration} = \text{true}$ **then**

$\pi \leftarrow \pi^*$; $\text{conv} \leftarrow 0$

else

$\text{conv} \leftarrow \text{conv} + 1$

if $\text{conv} = n - 1$ **then**

break

return π

Signification de *bestPolicy*

bestResponse renvoie les politiques optimales pour l'agent i selon un algorithme VI lexicographique et la fonction de récompense éthique emphatique individuelle.

Préférences sur des politiques

La politique π est préférée à la politique π' si $V^{\Delta h}(\pi) > V^{\Delta h}(\pi')$ ou si $V^{\Delta h}(\pi) = V^{\Delta h}(\pi')$ nous vérifions le même pour le niveau suivant $h + 1$.

Y-a-t-il un équilibre de Nash ?

Oui sous certaines conditions : les empathies doivent être uniformes sur tous les groupes d'agents.

1. Introduction

2. Modélisation d'une théorie morale

3. Jugement éthique dans une modèle de décision multi-agent

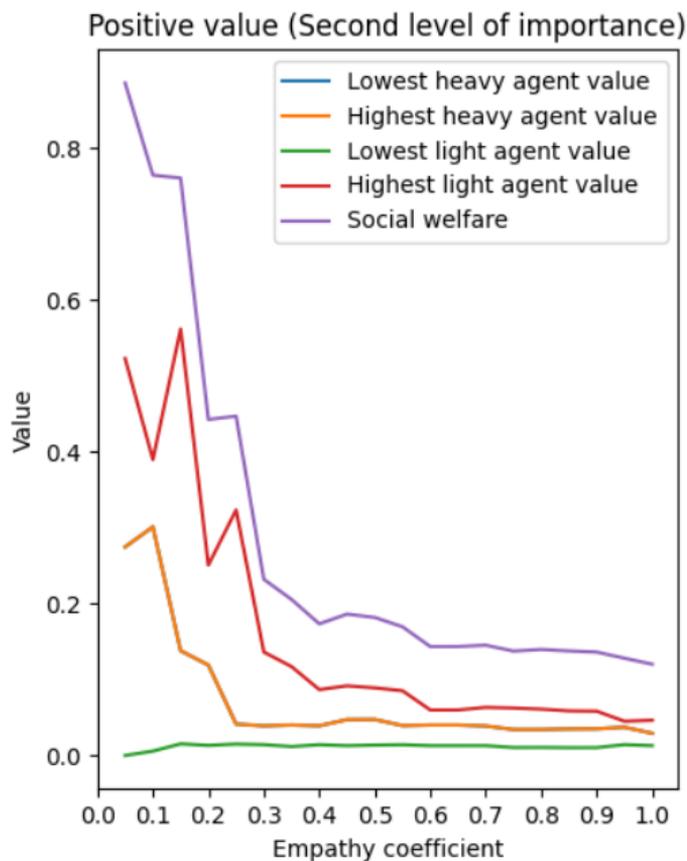
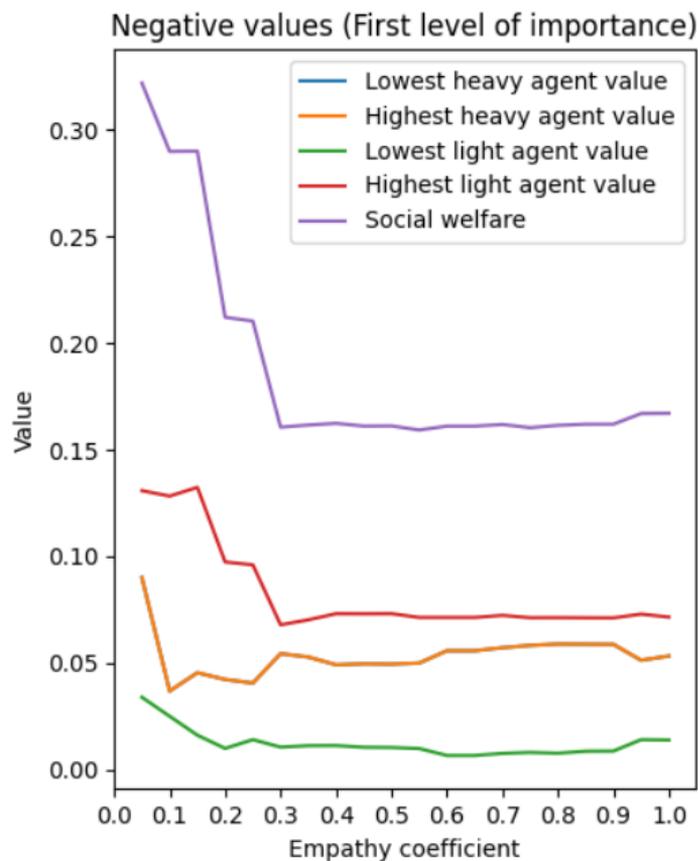
4. Expérimentation

Exemple : Problème d'allocation séquentielle décentralisée

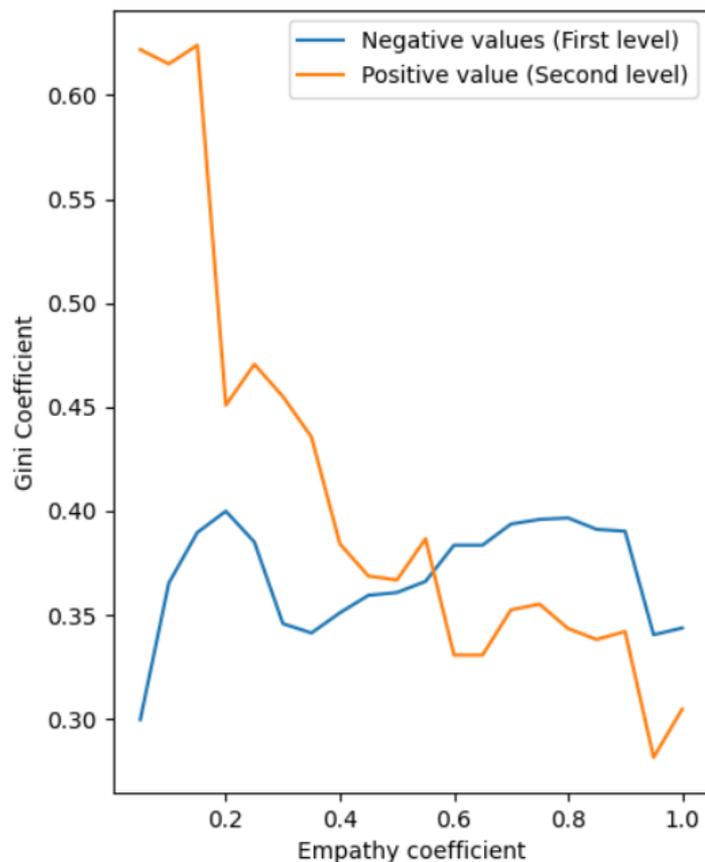
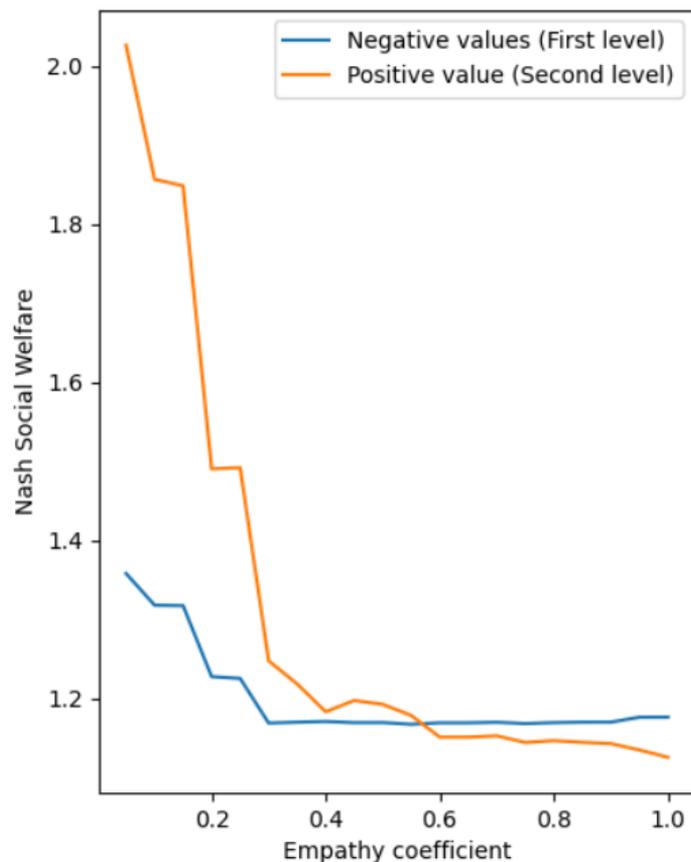
- ▶ Agents lourds et légers
- ▶ Produits de nécessité et de luxe
- ▶ Un état est défini par l'inventaire des produits disponibles et l'inventaire des agents
- ▶ Les agents ont des demandes minimales et maximales pour les produits de nécessité

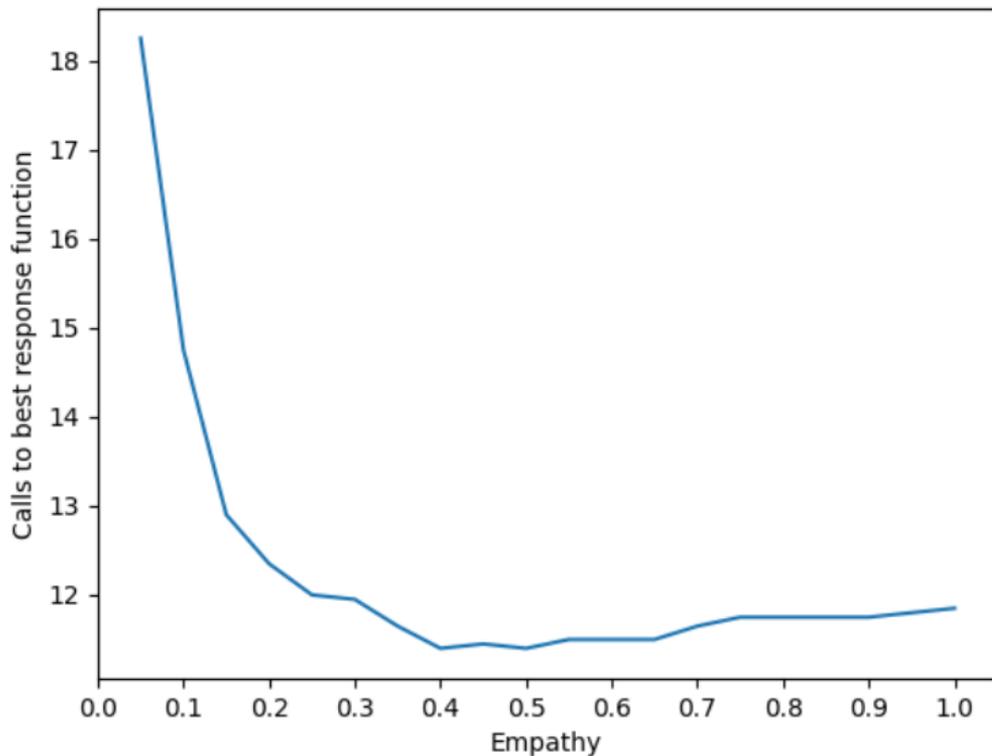
Valeurs morales

1. **Famine** : valeur négative (ne pas prendre des produits de nécessité)
2. **Gaspillage** : valeur négative (détruire un produit)
3. **Satisfaction de luxe** : valeur positive (prendre des produits de luxe)



Résultats : Trois agents légers et un agent lourd





Contributions

- ▶ Un modèle de jeu stochastique qui intègre des valeurs explicites avec des hiérarchies strictes entre les valeurs.
- ▶ Implémentation d'un problème d'allocation séquentielle décentralisée.
- ▶ Résultats qui montrent l'impact du coefficient d'empathie.

Perspectives

- ▶ Représentation d'autres type de compromis.
- ▶ Représentation des intentions.
- ▶ Approximation des politiques optimales.
- ▶ Observabilité partielle.

-  Grandi, U., Lorini, E., Parker, T., & Alami, R. (2022). Logic-Based Ethical Planning (A. Dovier, A. Montanari, & A. Orlandini, Eds.). In A. Dovier, A. Montanari, & A. Orlandini (Eds.), *21st International Conference of the Italian Association for Artificial Intelligence: Advances in Artificial Intelligence (AIXIA 2022)*, Springer International Publishing.
-  Svegliato, J., Nashed, S., & Zilberstein, S. (2021). Ethically compliant sequential decision making. In *Aaai conference on artificial intelligence*, Association for Computing Machinery.
-  De Moura, N., Chatila, R., Evans, K., Chauvier, S., & Dogan, E. (2020). Ethical decision making for autonomous vehicles. In *IEEE intelligent vehicles symposium*.
-  Rodriguez-Soto, M., Lopez-Sanchez, M., & Rodriguez-Aguilar, J. A. (2020). A structural solution to sequential moral dilemmas. In *19th international conference on autonomous agents and multiagent systems*.
-  Berreby, F., Bourgne, G., & Ganascia, J.-G. (2017). A declarative modular framework for representing and applying ethical principles. In *16th international conference on autonomous agents and multiagent systems*.
-  Abel, D., MacGlashan, J., & Littman, M. L. (2016). Reinforcement learning as a framework for ethical decision making. In *Acm conference on ai, ethics, and society (workshop)*.
-  Cointe, N., Bonnet, G., & Boissier, O. (2016). Ethical judgment of agents' behaviors in multi-agent systems. In *15th international conference on autonomous agents and multiagent systems*.
-  Timmons, M. (2012). *Moral theory: An introduction*. Rowman & Littlefield Publishers.
-  Bench-Capon, T. (2003). Persuasion in practical argumentation using value-based argumentation frameworks. *Journal of Logic and Computation*, 13(3), 429–448.
-  Nair, R., Tambe, M., Yokoo, M., Pynadath, D., & Marsella, S. (2003). Taming decentralized POMDPs: Towards efficient policy computation for multiagent settings. In *18th international joint conference on artificial intelligence*.
-  Shapley, L. S. (1953). Stochastic games. In *Proceedings of the national academy of sciences of the united states of america (pnas)*.

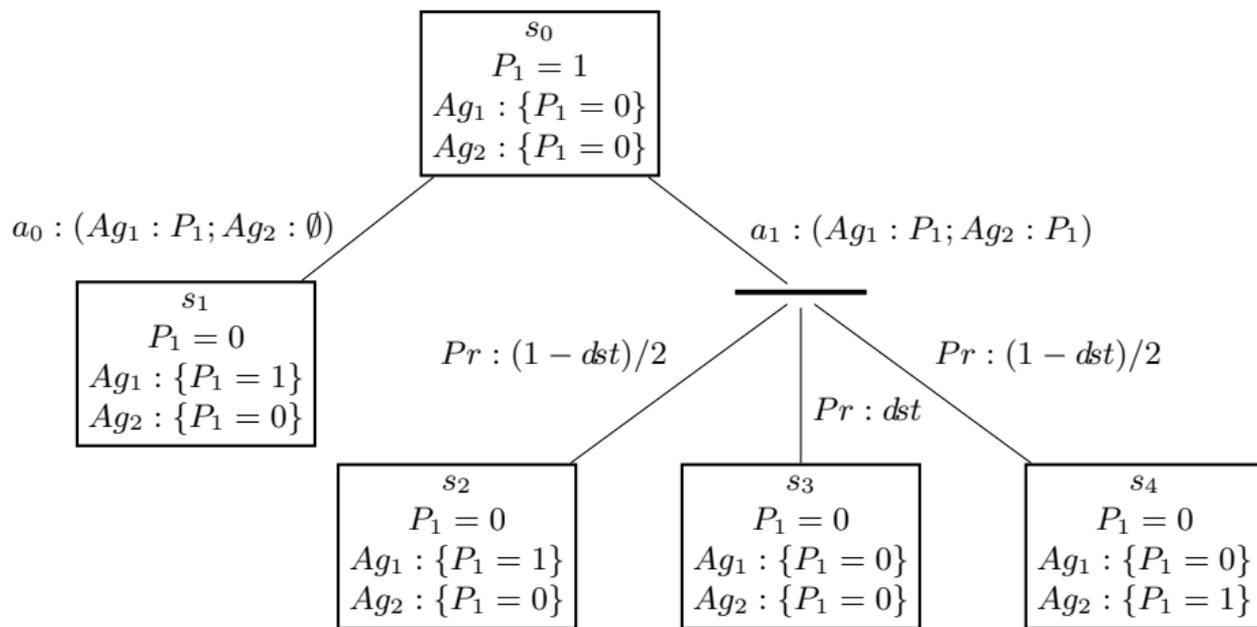


Figure: Exemple de transition lorsque les deux agents sont du même type.