

# APPRENTISSAGE DE DÉCISIONS ALIGNÉES SUR DES VALEURS HUMAINES, ET HUMAINS DANS LA BOUCLE

---

Rémy Chaput

2025/03/28

<https://rchaput.github.io/talk/gt-ace-2025/>



LIVE AND  
DISCOVER



# Contexte

# Contexte

- Domaine de l'éthique computationnelle
- On cherche à apprendre des comportements **alignés sur des valeurs humaines**
  - Valeurs = écologie, inclusivité, ...
  - **Différents humains** ont des **préférences différentes** sur **différentes valeurs** dans **différents contextes**
- L'apprentissage par renforcement (*RL*) est une méthode intéressante
  - Et plus particulièrement l'apprentissage multi-objectif (*MORL*) ([Deschamps, Chaput, and Matignon 2024](#))
- Éthique => vient des humains
  - Nécessité d'intégrer l'humain dans la boucle

Cas d'usage

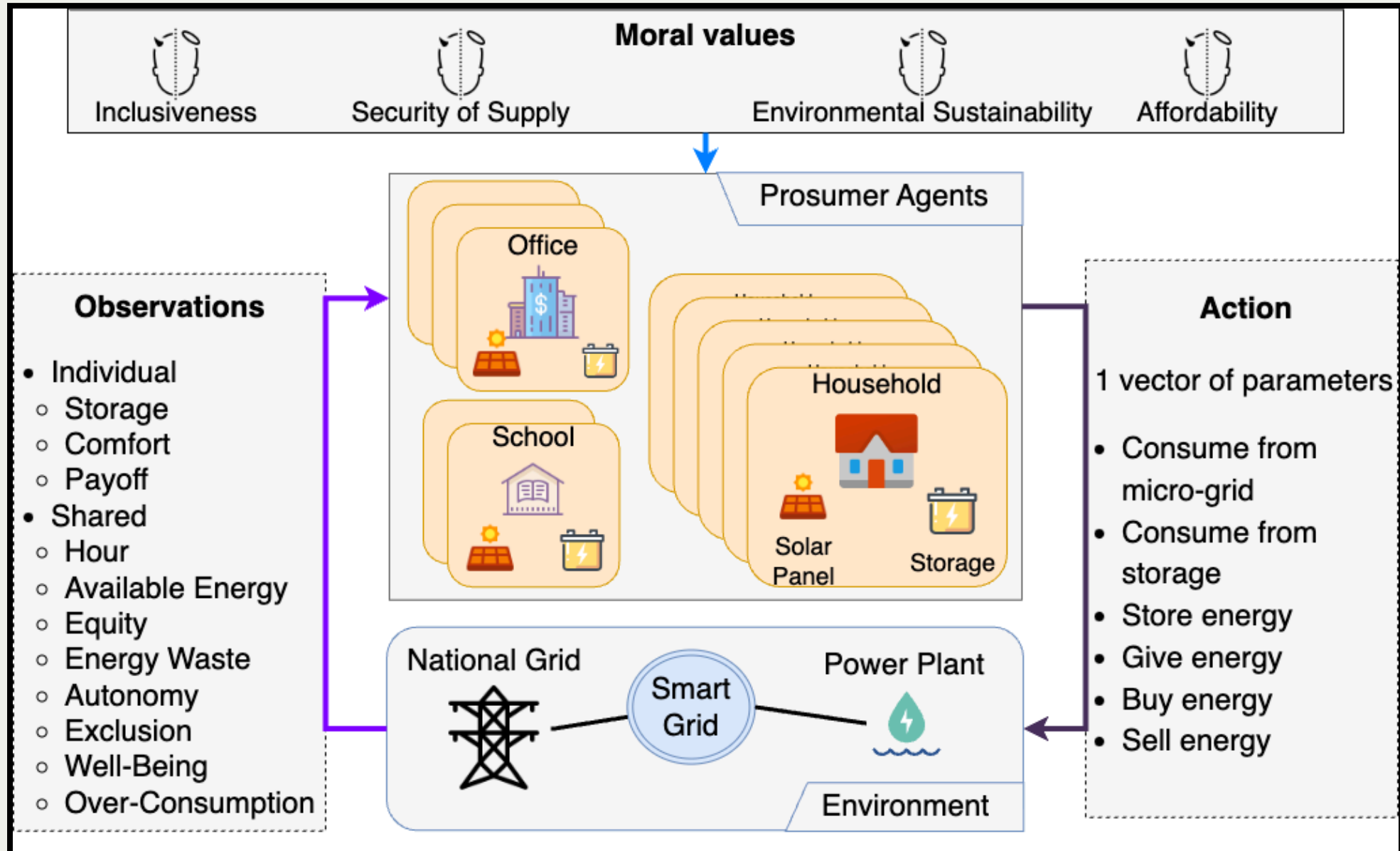
# Les cas d'usage en éthique computationnelle

# Smart Grid

- Proposition d'un cas d'usage **complexe**, **réutilisable** et **extensible**
- Répartition de l'énergie au sein d'une *smart grid*
  - Agents artificiels qui décident de la gestion de l'énergie (stock, consommation, achat/vente, ...) pour représenter au mieux les utilisateurs
  - Actions et observations continues et multi-dimensionnelles
- Projet OpenSource : <https://github.com/ethicsai/ethical-smart-grid/>
  - Documentation : <https://ethicsai.github.io/ethical-smart-grid/>

Scheirlinck, C., Chaput, R., & Hassas, S. (2023). Ethical Smart Grid: a Gym environment for learning ethical behaviours. Journal of Open Source Software, 8(88), 5410. <https://doi.org/10.21105/joss.05410>

# Smart Grid



# À vous de jouer !

Nous espérons que ce cas d'usage vous intéressera 😊

- Quelques exemples de code (à améliorer...)
- Conçu pour être paramétrable
  - Nombre d'agents
  - Types d'agents
  - Conditions de l'environnement : énergie disponible, durée, ...
- Possibilité d'extensions :
  - Autres types d'agents
  - Observations différentes
  - Autres éléments de l'environnement ...



# L'approche QSOM-MORL

# Apprentissage de politiques

- On considère plusieurs valeurs morales (bien-être, équité, écologie, affordability)
- On utilise du MORL pour apprendre les **intérêts** des actions possibles dans chaque situation par rapport à chaque valeur morale
- On obtient des vecteurs d'intérêts pour chaque paire situation/action
  - $Q(s1, a1) = [3, 4, 3.5, 3]$
  - $Q(s1, a2) = [1, 2, 3.5, 3]$
  - $Q(s1, a3) = [5, 3, 2.5, 3]$
  - ici, on remarque que  $Q(s1, a1)$  est meilleure que  $Q(s1, a2)$  sur les deux premières valeurs morales, et aussi bonne sur les deux dernières




# Apprentissage de politiques

- Idéalement, la politique optimale est celle qui choisit la meilleure action à chaque situation ...
  - ? Que veut dire “meilleure” quand on compare des vecteurs ?
- Façon “simple” : on considère des poids sur chaque composante et on pondère !
  - => ⚠ C’est difficile à appliquer pour l’éthique :
    - difficile pour un utilisateur de donner des poids à priori
    - différentes préférences selon le contexte (ex : été vs hiver)

# Détection de dilemmes

- Notre façon : on définit des “**niveaux de préférences éthiques**”
  - Représentent le degré de satisfaction attendu pour chaque valeur morale
  - Ex : [50%, 75%, 50%, 60%]
  - Différent pour chaque utilisateur
- On compare les intérêts de chaque action à leurs maximums théoriques
- Si le **ratio** intérêt appris / intérêt maximum d'une action **dépassent les niveaux**, pour chacune des valeurs morales, on effectue cette action
- Sinon ... => c'est un **dilemme** !

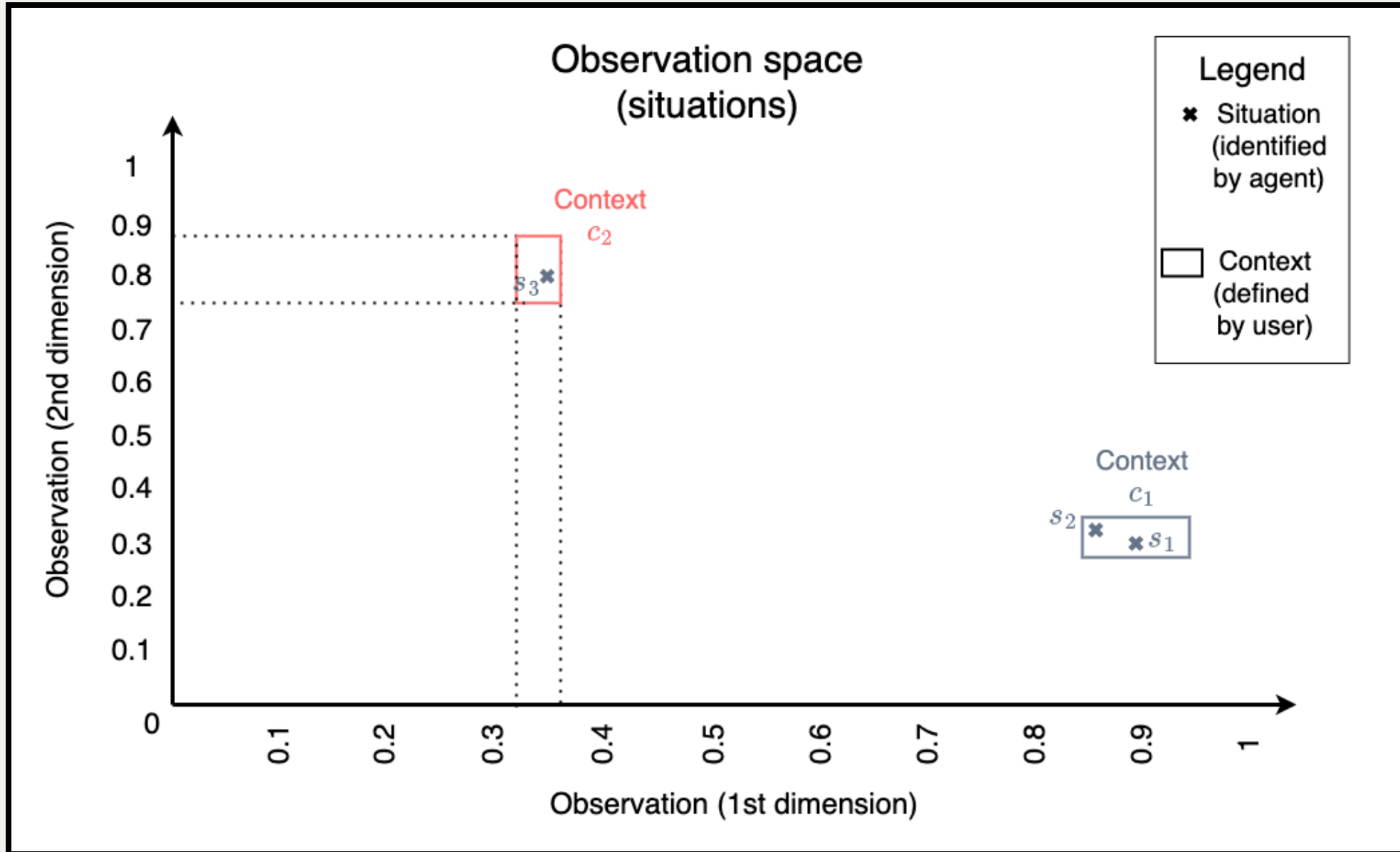
# Différents utilisateurs reconnaissent les dilemmes différemment

| Action | Interests<br>$Q(a_i)$ | Theoreticals<br>$Q^{th}(a_i)$ | Ratio<br>$\frac{Q(a_i)}{Q^{th}(a_i)}$ |   |
|--------|-----------------------|-------------------------------|---------------------------------------|---|
| $a_1$  | [3, 4, 3.5, 3]        | [5, 5, 5, 5]                  | [60%, 80%, 70%, 60%]                  | <br>Human thresholds $\zeta_1$<br>[50%, 75%, 50%, 60%]<br>Acceptable |
| $a_3$  | [5, 3, 2.5, 3]        | [6, 6, 6, 6]                  | [83%, 50%, 42%, 50%]                  | <br>Human thresholds $\zeta_2$<br>[80%, 45%, 20%, 50%]<br>Acceptable |
|        |                       |                               |                                       | <br>Human thresholds $\zeta_3$<br>[75%, 70%, 0%, 60%]<br>Dilemma   |

# Résolution de dilemmes par préférences utilisateurs

- Humain dans la boucle : quand un dilemme est identifié, on **demande à l'utilisateur** quelle action effectuer
- Problème : si trop de dilemmes, on risque de demander trop souvent !
- => On regroupe les dilemmes similaires entre eux
  - Notion de **contexte** : bornes sur les observations d'une situation
  - Défini par l'utilisateur lorsqu'un (nouveau) dilemme est identifié

# Dilemmas et contextes



# Retour d'expérience



# Questionnaire d'utilisation

- Nous avons effectué l'an dernier un questionnaire sur l'utilisabilité de notre approche
- Plusieurs parties :
  - Description du cas d'usage *Smart Grid*
  - Configuration de base : niveaux de préférence éthique sur les valeurs morales, seuil de similarité sur les actions
  - Utilisation factice du système sur 4 situations
    - Présentation de la situation (observations)
    - Est-ce que l'utilisateur est d'accord avec le fait que la situation a été classée comme dilemme / non-dilemme
    - Présentation des actions possibles
    - Choix de l'action préférée
    - Définition du contexte

# Difficulté de configurer

- Les répondants ont eu du mal à “configurer” le système
  - Beaucoup d’explications / texte à lire ...
- Compliqué de définir des niveaux de préférence éthique
  - Hors contexte, comment donner des nombres ?
  - Un répondant propose d’extraire les préférences via des choix
  - => Toute la littérature sur l’élicitation de préférences !
  - Attention : nous nous mentons souvent à nous-même concernant l’éthique ...

# Dilemmes : 2 avis opposés

- Une majorité des utilisateurs préfère que les situations soient classés comme des dilemmes
  - *“Je préfère résoudre le dilemme moi-même, pas confiance dans une évaluation éthique dans ce cas”*
  - *“C’est pas une dictature”*
  - 75 réponses “Convient plutôt / Convient tout à fait”
- Mais une portion non-négligeable aurait préféré que ce soit automatisé !
  - *“Je préfère que ce soit automatisé. J’ai autre chose à faire!”*
  - 17 réponses “Ne convient plutôt pas / Ne convient pas du tout”
- => Importance de l’humain dans la boucle ... mais penser aussi à ceux qui ne veulent pas y être !

# Adaptation à l'humain

- Les répondants auraient voulu pouvoir modifier les actions directement
  - Quel impact sur le reste de la politique ??
  - En particulier quand la politique est prévue comme une séquence d'actions optimale sous réserve qu'on suive le reste de la politique ...
  - Cf. fonction de valeur :

$$V_{\pi}(s) = \sum_a \pi(a|s) \sum_{s'} Pr(s'|s, a)(R(s, a, s') + \gamma V_{\pi}(s'))$$

# Conclusion

# Pistes

- MORL avec préférences non-linéaires + Continuous RL pour gérer l'évolution potentielle des préférences
- Prise en compte de l'incertitude
  - Sur les données en entrée : vie privée donc observations partielles
  - Sur les intérêts appris (multi-objectif donc difficile de mettre à jour)
  - Mais aussi sur les actions effectuées : politiques capables de s'ajuster
- Explicabilité des systèmes d'IA
  - Nécessaire pour garder le contrôle dessus ...
  - Comment choisir une action si on ne comprend pas ce qu'elle va faire / impliquer ?

## *Key takeaway*

- Éthique => importance de l'humain dans la boucle
  - La majorité des utilisateurs (dans notre questionnaire) voulaient pouvoir choisir
- S'adapter à l'utilisateur
  - Système compréhensible pour être utilisable (vocabulaire, concepts, etc.)
  - Attention à la surcharge cognitive et/ou sur-sollicitation
  - Permettre également de “laisser faire”
- Comment planifier une politique optimale si l'utilisateur peut demander à changer une action ?

# Merci de votre attention !

Des questions ?

Contacts :

- Rémy Chaput [remy.chaput@cpe.fr](mailto:remy.chaput@cpe.fr)
- Projet ANR AccelerAI [accellerai@liris.cnrs.fr](mailto:accellerai@liris.cnrs.fr)
  - Porteuse du projet : Laetitia Matignon  
[laetitia.matignon@liris.cnrs.fr](mailto:laetitia.matignon@liris.cnrs.fr)

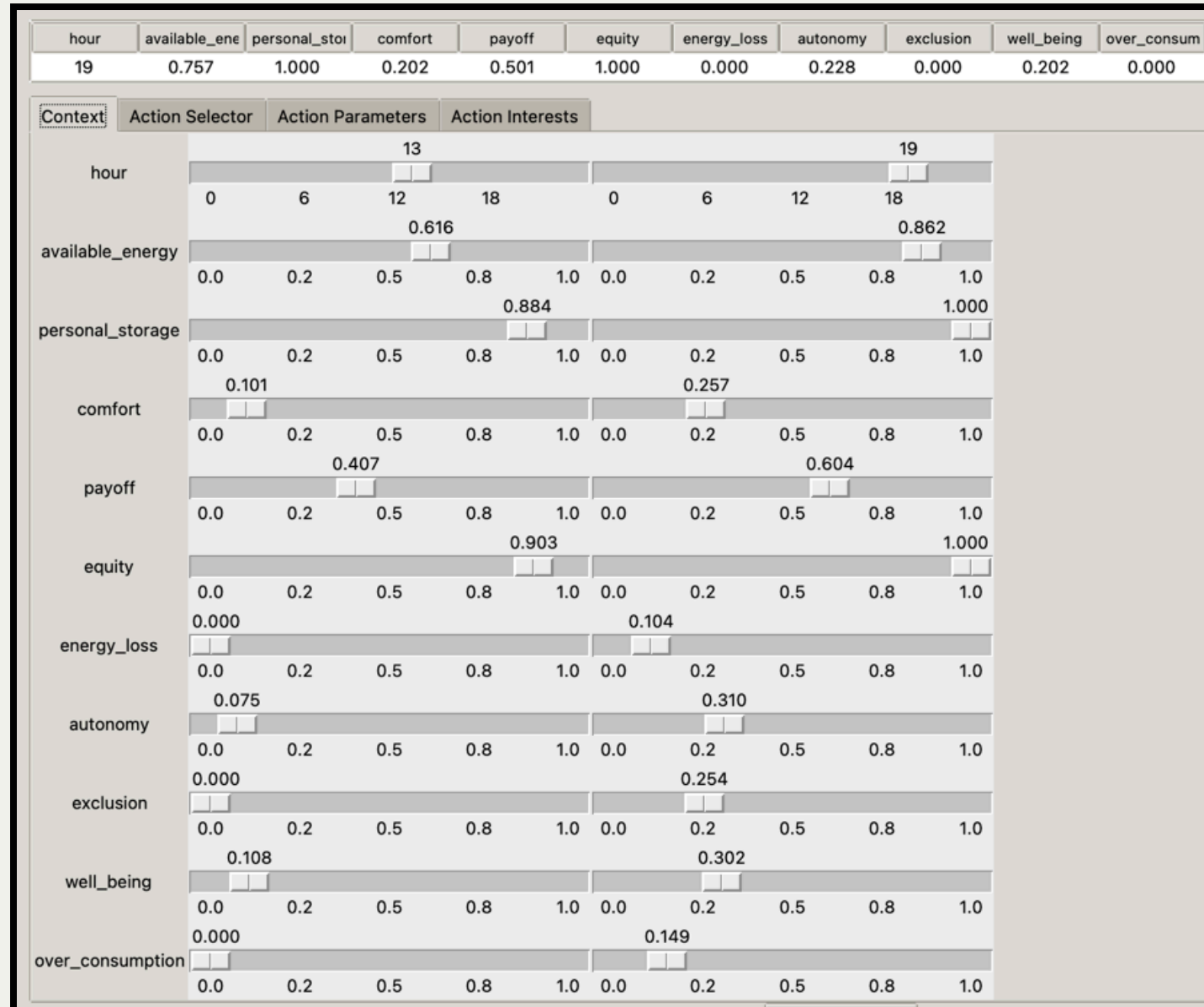


# References

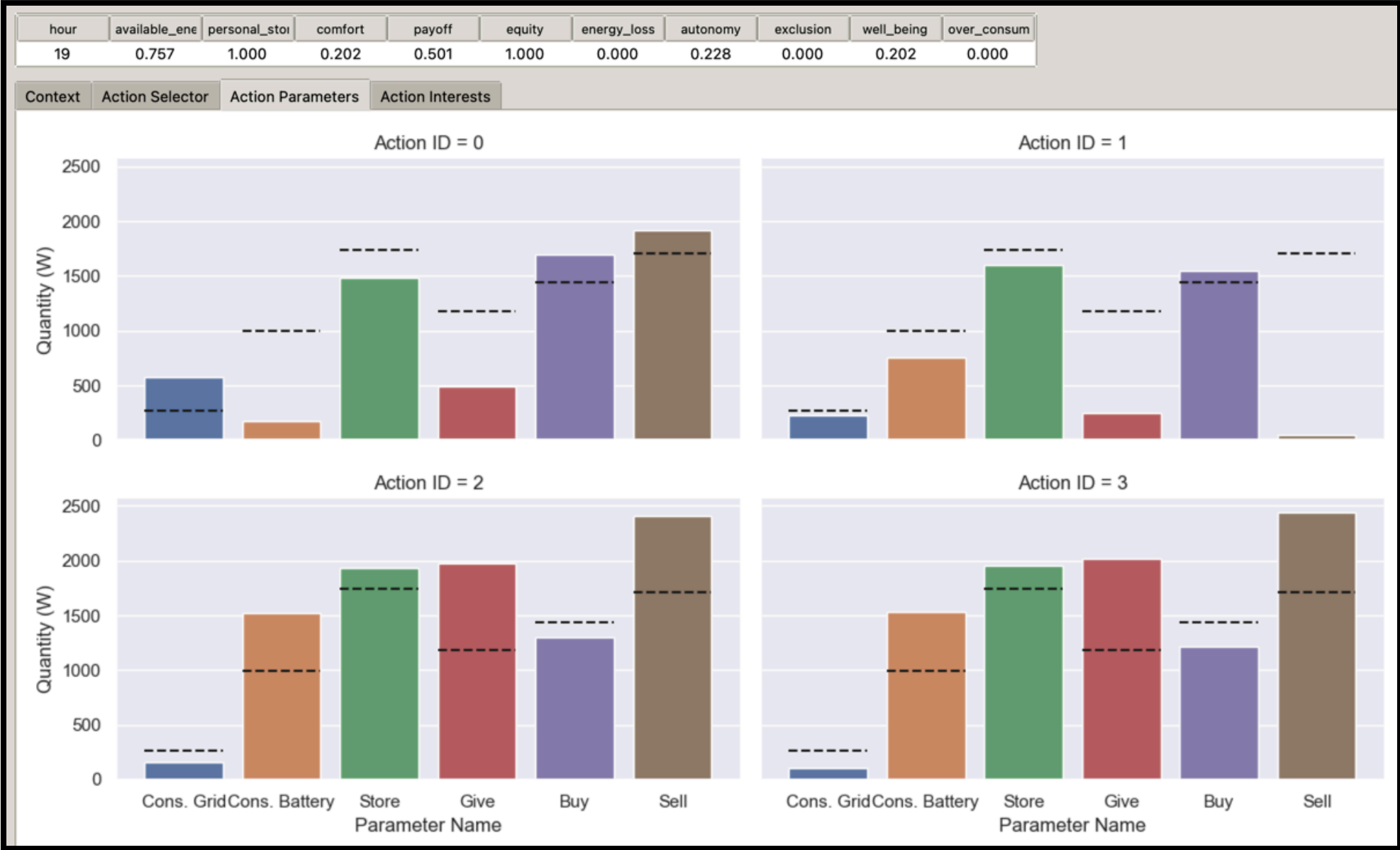
- Anderson, Michael, Susan Leigh Anderson, and Vincent Berenz. 2019. "A Value-Driven Eldercare Robot: Virtual and Physical Instantiations of a Case-Supported Principle-Based Behavior Paradigm." *Proceedings of the IEEE* 107 (3): 526–40. <https://doi.org/10.1109/JPROC.2018.2840045>.
- Chaput, Rémy, Laetitia Matignon, and Mathieu Guillermin. 2023. "Learning to Identify and Settle Dilemmas Through Contextual User Preferences." In *2023 IEEE 35th International Conference on Tools with Artificial Intelligence (ICTAI)*, 474–79. <https://doi.org/10.1109/ICTAI59109.2023.00075>.
- Deschamps, Timon, Rémy Chaput, and Laëtitia Matignon. 2024. "Multi-objective reinforcement learning: an ethical perspective." In *Multi-Objective Decision Making Workshop*. Santiago de Compostela, Spain: ECAI. <https://hal.science/hal-04711682>.
- Haas, Julia. 2020. "Moral Gridworlds: A Theoretical Proposal for Modeling Artificial Moral Cognition." *Minds and Machines*, April. <https://doi.org/10.1007/s11023-020-09524-9>.
- LaCroix, Travis, and Alexandra Sasha Luccioni. 2022. "Metaethical Perspectives on 'Benchmarking' AI Ethics," no. arXiv:2204.05151 (April). <https://doi.org/10.48550/arXiv.2204.05151>.

# Annexes

# Interface de résolution de dilemmes



# Interface de résolution de dilemmes



# Interface de résolution de dilemmes



# Interface de résolution de dilemmes

| hour | available_ene | personal_stoi | comfort | payoff | equity | energy_loss | aut |
|------|---------------|---------------|---------|--------|--------|-------------|-----|
| 19   | 0.757         | 1.000         | 0.202   | 0.501  | 1.000  | 0.000       | 0.  |

| Context   | Action Selector | Action Parameters | Action Interests |
|---|-----------------|-------------------|------------------|
| <div><div>Action ID = 0</div><div><input checked="" type="radio"/> Parameters = [0.23111555 0.06819946 0.59250098 0.19501867 0.67720321 0.76896747]</div><div>Interests = [5.70397815 6.67034231 6.67074222 0.65284908]</div></div> |                 |                   |                  |
| <div><div>Action ID = 1</div><div><input type="radio"/> Parameters = [0.0886732 0.30100162 0.64076246 0.09730741 0.62050321 0.01911589]</div><div>Interests = [2.31330539 2.09347349 7.0866135 0.24543208]</div></div>              |                 |                   |                  |
| <div><div>Action ID = 2</div><div><input type="radio"/> Parameters = [0.06320528 0.60990433 0.77258426 0.79014815 0.51986592 0.96462507]</div><div>Interests = [2.45198313 2.9457167 3.97402727 1.61318562]</div></div>             |                 |                   |                  |
| <div><div>Action ID = 3</div><div><input type="radio"/> Parameters = [0.041645 0.61255743 0.78164123 0.80839148 0.48636543 0.97873474]</div><div>Interests = [2.76133183 2.84486227 4.37412502 1.77183498]</div></div>              |                 |                   |                  |