







On Admissible Behaviours for Goal-Oriented Decision-Making of Value-Aware Agents

Andrés Holgado-Sánchez^(✉) , Joaquín Arias , Mar Moreno-Rebato ,
and Sascha Ossowski 

CETINIA, Universidad Rey Juan Carlos de Madrid, 28933 Móstoles, Spain
{andres.holgado,joaquin.arias,mar.rebato,sascha.ossowski}@urjc.es

Abstract. The emerging field of *value awareness engineering* claims that software agents and systems should be value-aware, i.e. they should be able to explicitly reason about the value-alignment of their actions. Values are often modelled as preferences over states or actions which are then extended to plans. In this paper, we examine the effect of different groundings of values depending on context and claim that they can be used to prune the space of courses of actions that are aligned with them. We put forward several notions of such value-admissible behaviours and illustrate them in the domain of water distribution.

Keywords: Value alignment · Value-admissible behaviours · Value awareness engineering · Water distribution

1 Introduction

A key requirement for trustworthy AI is to consider ethical aspects in the design and implementation of AI systems. In particular, it is considered of utmost importance that autonomous AI agents and systems include a systematic way of aligning their decisions with human values. While value-based decision-making is a widely discussed problem in sociology, only recently it has found its way into computer science [17]. The emerging field of *value awareness engineering* [8] claims that software agents and systems should be value-aware, i.e. they should be able to explicitly reason about the value-alignment of their actions.

Proposals for modelling value-based decision processes of autonomous agents are often based on preferences over states or actions [7,9], which are then extended to sequential decisions. Other approaches [2,3] set out from observed sequences of actions (plans) and then learn preferences over states or actions through (inverse) reinforcement learning [10].

In this paper, we are concerned with the role of values in plan selection of autonomous value-aware agents. In particular, we argue that values not only induce preferences over plans, but may also be used to discard certain courses of actions right away depending on a particular value grounding. For this purpose,

we put forward several notions of value-admissible behaviours, and illustrate them with regard to different groundings of the value of equity in water distribution, taking into account real-world (legal) restrictions.

This paper is structured as follows. Section 2 presents a discussion of related work. Section 3 introduces the value-related world model for this paper, and puts forwards our notion of value-admissible plans. In Sect. 4 we present a use case regarding equitable domestic water distribution in a drought scenario, providing legal considerations around the value of equity. We also describe and analyze the results of applying the proposed value-alignment framework to the example. Finally, Sect. 5 presents our conclusions and points to future lines of work.

2 Related Work

The practical reasoning community was among the first to formally represent values for computation. Weide et al. [17] introduced *value preferences* represented as *agent perspectives* that consist of preorder relationships between states to represent the agent’s ideas on how states promote or demote certain values. However, they use that preference to perform actions in reasoning schemes and do not analyze sequences of decisions.

An approach more concerned about abstraction and generality of value representation is introduced by Montes and Sierra [9]. It conceives states as representative of values through a function evaluation, and relies on a taxing example in order to illustrate a more general framework for optimizing value-alignment of normative systems. Still, their analysis does not consider the effects of choosing different value semantics functions or other criteria that would characterize value-admissible plans. Similarly, Lera-Leri et al. [7] proposed an extended formalization of a value system where the focus is put on numerically assessing the value of both *taking* or *not taking* actions (instead of states). This framework, though indeed useful for the value system aggregation problem is, again, not focusing on analyzing sequences of decisions/actions.

Techniques on reinforcement learning (RL) [16] are considered state of the art in most decision-making scenarios, though human values have been introduced scarcely into those systems. There are examples of policies learning values such as fairness jointly with efficiency in multi-agent systems [6]. The approach defines a suitable special reward function based on the Coefficient of Variation (CV) that, for each agent, intends to maximize its default bounded reward subject to that reward being similar to the other’s, in a resource allocation problem. A similar approach was developed in [5] to consider equality in social dilemmas. Finally, [14] brings forward a powerful model for both multi-value-aware and multi-norm-compliant MDPs, but it relies heavily on the algorithmic value concept in RL to define the criteria of best value-promoting plans.

As specifying rewards manually requires domain expertise and is a process prone to optimization, IRL (Inverse Reinforcement Learning) [10] has been used which learns the reward from value-aligned trajectories. However,

Arnold et al. [1] show that IRL by itself may not be adequate for agents to learn values, suggesting the use of an external process to actually infer the norms or guidelines shaping the value-aligned decisions.

3 Value Aligning Sequences of Decisions

We define goal-oriented decision-making as a model based on a Multi-Agent System (MAS) [9], where the world is modelled as a labelled transition system, called **decision world** $(\mathcal{S}, \mathcal{A}, \mathcal{T})$ with the following elements.

- **States** \mathcal{S} , representing the MAS completely in each situation.
- **Actions** \mathcal{A} , representing the MAS joint actions or decisions.
- **Transitions** $\mathcal{T} \subset \mathcal{S} \times \mathcal{A} \times \mathcal{S}$, representing available actions connecting each pair of states. We will denote them with $s \xrightarrow{a} t$, where $s, t \in \mathcal{S}$ and $a \in \mathcal{A}$.
- **Paths** \mathcal{P} , representing joint transitions (sequences of decisions), e.g. a path of length n from s_0 to s_n would be represented as: $P = s_0 \xrightarrow{a_1} s_1 \xrightarrow{a_2} \dots \xrightarrow{a_n} s_n$.
- **Goal States** $\mathcal{G} \subset \mathcal{S}$, representing states where agents satisfy their needs or aims in the problem.
- **Plans**, representing paths that we consider solutions to our problems, i.e. those going from a given initial state s_0 , to a goal state $s_g \in \mathcal{G}$.

We are interested in identifying which plans adhere the better with a value v under consideration. We assume v is firstly grounded in states for then, constructing path-level criteria.

3.1 State-Level Alignment: Value Preferences

Following Weide et al. [17] or Sierra et al. [15], we assume a value preference among states based on a preorder relation \sqsubseteq_v , which we call **perspective** or **value preorder**, i.e. given s and s' , two states, $s \sqsubseteq_v s'$ means that s' is at least as preferred as s w.r.t. the value v .

Another approach is using a numeric value to quantify the above relation. Citing [9], the semantics of a value v in state s is an *unbounded semantics function* $f_v : \mathcal{S} \rightarrow \mathbb{R}$, where f_v is directly proportional to the promotion of v .¹ The relationship between those approaches is fairly straight-forward: $s' \sqsubseteq_v s'' \iff f_v(s') \leq f_v(s'')$. Examples of statistical functions that can be used to define semantics functions for the value of equity are the following:

1. **Maximum-Minimum difference (Mn)**: Difference between maximum and minimum values of the state. Inversely proportional to equality.
2. **Sample Standard Deviation (SSD)**: Standard deviation as dispersion metric is inversely proportional to equality.

¹ Original definition from Montes and Sierra [9] assumes that the range of all value semantics functions is bounded in $[-1, 1]$, so $f_v(s) \approx -1, 0, +1$ indicates that state s strongly opposes, is neutral or strongly promotes the value v , respectively. This would represent an (unnecessary strict) *absolute* value promotion metric.

3. Median Absolute Deviation (**MAD**). It is a robust version of the SSD, unaffected by outsiders.
4. Coefficient of Variation (**CV**) [6]. Defined as the sample standard deviation over the mean (in absolute value). Values closer to 0 mean greater equality.
5. Gini Index (**GI**) [5, 9]². Inequality in an economic system is usually represented with this function as it has unique important properties [12].

3.2 Plan-Level Alignment and Admissibility

In literature, the value-alignment of a path (and a of plan, by extension) is given by a human [2] or calculated by aggregating values of states [9] or actions [7]. However, it is important to notice that, from the point of view of the decision-making of a value-aware agent, not all courses of action need to be considered. For instance, in a water distribution scenario, all assignments that, at some point in time, leave stakeholders without a minimum amount of water necessary for basic needs, should not be considered even if they lead to a final state in which water distribution is equitable. These “lower bounds” on the value alignment determine the paths that are *admissible* under a certain value. They can either be determined in absolute or in relative terms, and based on preference preorders or semantics functions, as we will argue in the sequel.

Given an aggregation function *agg*, and a semantics function f_v , we define the semantics of a value for a path $P = s_0 \xrightarrow{a_1} \dots \xrightarrow{a_n} s_n$ as: $agg_v(P) = agg(\{f_v(s_0), \dots, f_v(s_n)\})$. This is called its **aggregated alignment**. Examples of aggregation functions (*agg*) are the mean, the (discounted) sum, the maximum, etc. This aggregation concept was already mentioned as a modelling aspect in [9].

Value-admissible behaviours for a value v are given by a constraint criterion on the set of all plans \mathcal{P} . It characterizes the subset of plans $B(\mathcal{P}, \sqsubseteq_v)$ that are admissibly aligned with the value, based on state/action-level alignment \sqsubseteq_v . In this paper we are concerned with three very general classes of such behaviours:

- a) **Local behaviour**. Admits plans which are constructed by only visiting the next states that are the most preferable:

$$B_{local}(\mathcal{P}, \sqsubseteq_v) = \{P \in \mathcal{P} \mid \forall s \xrightarrow{a} t \in P, \exists s \xrightarrow{a'} t' \in Q \in \mathcal{P} \cdot t \neq t' \wedge t \sqsubseteq_v t'\}$$

- b) **Goal behaviour**. Admits plans leading to the goal states that are the most preferable. Here, $out(P)$ denotes the final and goal state of P .

$$B_{goal}(\mathcal{P}, \sqsubseteq_v) = \{P \in \mathcal{P} \mid \exists Q \in \mathcal{P} \cdot out(Q) \neq out(P) \wedge out(P) \sqsubseteq_v out(Q)\}$$

² Note that the $[-1, 1]$ -bounded semantics function used in [9] is defined in terms of the Gini index, i.e., $f_{eq} = 1 - 2 \cdot GI(s)$. Similarly, the rest of the semantics functions we have enumerated can be bounded to that interval if needed. For this theory, we just consider these functions as *quantifiers* of value preorders.

- c) **Aggregated behaviour.** This strategy admits plans with the highest overall alignment according to an *agg* aggregation function.

$$B_{agg_v}(\mathcal{P}, \sqsubseteq_v) = \{P \in \mathcal{P} \mid \nexists Q \in \mathcal{P} \setminus \{P\} \cdot agg_v(P) \leq agg_v(Q)\}$$

Requiring value-admissibility of such behaviours obviously reduces the space of plans that a value-aware agent can choose from. In some situations (e.g. in Sect. 4.2 while using certain semantics functions) this may even lead to a unique admissible plan. Therefore, we can introduce some relaxation over the above criteria, by admitting some more states of plans that are admissibly close to abiding to them. This relaxation can be more easily stated by quantifying the preorder, i.e. using (not necessarily bounded) semantics functions. As an example, we detail the *epsilon-local* behaviour:

ϵ -Local Behaviour. Given a set of plans \mathcal{P} , $\epsilon \in \mathbb{N}$, and the semantics function for a value v , f_v , the ϵ -local behaviour, B_ϵ is defined as:

$$B_\epsilon(\mathcal{P}, f_v) = \{P \in \mathcal{P} \mid \forall s \xrightarrow{a} t \in P, f_v(t) \geq \max\{f_v(t') \mid \exists t \xrightarrow{a'} t' \in Q \in \mathcal{P}\} - \epsilon\}$$

This behaviour extends the local one by admitting not only the next most preferable state(s) but the ϵ -most preferred at each step; i.e., among the next possible states, we would admit traversing those with up to an ϵ decrement in semantics value w.r.t the most valued one(s).

4 Example: Equity in Water Distribution

To illustrate our approach to value alignment, we draw upon a use case in the domain of water distribution. This domain has been explored deeply, i.e. with socio-cognitive agents [11], though with no value-awareness in mind yet. In the following we first summarise legal aspects and values related to water use, and then present a simple example considering a situation of water distribution in a drought scenario, where the value of equity is to be maintained.

4.1 Legal and Values Considerations for Water Distribution

Preserving values in the context of water distribution is indeed of the maximum importance and representative of general situations. At the European level, the Parliamentary Assembly of the Council of Europe declared that access to water must be recognized as a fundamental human right because it is essential for life on the planet and it is a resource that must be shared by humanity³. Providing such access is, in turn, a commitment under the UN Sustainable Development Goal No. 6 of the 2030 Agenda “Ensure availability and sustainable management of water and sanitation for all”.⁴

³ Council of Europe Parliamentary Assembly Resolution No. 1693 (2009).

⁴ SDG 6 of the United Nations 2030 Agenda for Sustainable Development <https://www.un.org/sustainabledevelopment>.

As we have seen above, water is an essential good for human life, so universal access to it must be guaranteed; but water is also a scarce resource with economic value, which contributes simultaneously to social, environmental, and economic objectives.⁵ Currently, the water volume allocation for agriculture is 70%. In water stress scenarios, it will undoubtedly be necessary to reallocate this percentage to other uses⁶ and, consequently, to improve water management, including digitization in this sector. This will require a better allocation of water in situations of scarcity and theorizing about different models.

In Spain, the average household water consumption was 133 litres per inhabitant per day.⁷ The main use of water is irrigation and agricultural use, which accounts for approximately 80.5% of this demand, followed by urban supply, which represents 15.5%. The remainder is for industrial use [4]. Of all the water uses, the priority is urban water supply.⁸ The regulations have established that the net or average consumption endowment, as a minimum objective, must be at least 100 litres per inhabitant per day.⁹

From the legal point of view, water (surface and groundwater) is a public good (i.e., it is not subject to private ownership). Urban water supply is configured as a public service, extensively regulated (including its price through the corresponding tariff) and, as such, it has the characteristics inherent to such services: equal access, provision, and quality, the existence of basic common conditions, universality and continuity, solidarity, transparency and control with user participation [13]. In turn, the legislation establishes general principles applicable to water management, from which stand out management unit, integral treatment, deconcentration, decentralization, coordination, efficiency and user participation.¹⁰

4.2 Use Case

In our use case, in a situation of drought, water needs to be distributed from a reservoir to 4 equally populated and distant villages using a tanker vehicle with 11kl of capacity. We consider a goal of distributing a total of 44kl from the reservoir to the villages. For each trip, we must decide which village is visited and supplied with water. For simplicity, the vehicle always discharges the entire capacity of its tank when arriving to a village.

The problem can be modelled with the basic elements of the decision world introduced in Sect. 3 as follows:

⁵ <https://www.oecd.org/water/Recomendacion-del-Consejo-sobre-el-agua.pdf>.

⁶ In agriculture (<https://www.bancomundial.org/es/topic/water-in-agriculture>).

⁷ Statistics on Water Supply and Sanitation Year 2020, see https://www.ine.es/prensa/essa_2020.pdf.

⁸ Royal Decree 1/2001, of July 20, approving the Revised Text of the Water Law, Article 60.

⁹ Royal Decree 3/2023, of January 10, establishing the technical-sanitary criteria for the quality of drinking water, its control, and supply, Article 9.

¹⁰ Royal Decree 1/2001, of July 20, approving the Revised Text of the Water Law, Article 14.

- **States:** a state is a list of four values where each value represents the amount of water delivered to each village, i.e. $[11, 11, 0, 0]$.
- **Actions:** an action indicates the village visited by the vehicle, identified by a number from 1 to 4 (one for each village).
- **Transitions:** Depending on goals, we will have different transitions, though they all model that the truck delivers its $11kl$ to the village indicated by the action. An example of a transition would be $[0, 0, 11, 11] \xrightarrow{4} [0, 0, 11, 22]$.

Depending on the particular context, the value of equity in this scenario can be grounded in different semantics functions. We intend to examine the impact of choosing a specific semantics function in relation to different notions of value-admissible behaviour. For this purpose, we consider three different semantics functions inspired by statistical ones from Sect. 3: $f_1 = -Mm$, $f_2 = -SSD$ and $f_3 = -MAD$, and the three main behaviours proposed in Sect. 3.2, i.e., Local, Goal, and Aggregated (considering the **sum** as the aggregation function, no “*epsilon*” versions considered yet).

Figure 1 shows the state-transition diagram for our use case. For simplicity, we collapsed states ordering the variables from highest to lowest, so each path represents much more distributions, but all equivalent in the end. In the figure, plans pertaining to local behaviours are represented by red edges, the ones from the aggregated behaviours by blue edges, and goal states are indicated by green nodes.

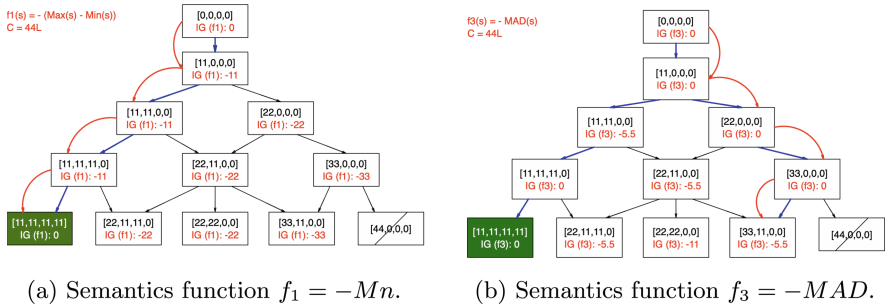


Fig. 1. Plans admissible to deliver 44kl under local, goal, and aggregated behaviours. The *local plans* are represented by red edges; the aggregated plans, by blue edges; and the goal plans are those going from the initial state to the green nodes, which mark the most value-aligned goal states. (Color figure online)

In line with the discussion put forward in Sect. 4.1, we can assume that the *local behaviour* is strongly aligned with the value of equity, as an agent adhering to that type of behaviour can justify its actions by claiming that it is always promoting equity to the best it can at each moment. Figure 1a shows that, for f_1 , there is one single plan admissible under all the behaviours. Indeed, the goal-admitted plan (reaching the green state) coincides with the plans admitted by the local and aggregated behaviours.

By contrast, Fig. 1b indicates that considering the semantics function f_3 there is only one goal-admissible plan, which does not coincide with the local-admissible plan. Both plans are aggregate-admissible.

It is worth noting that with Schur-Concave semantics functions, such as $f_2 = -\text{SSD}$ or the modified Gini Index from [9], the three behaviours admit the same single plan (Fig. 1a). This plan certifies a local behaviour which is aligned with the value while keeping the highest overall final alignment score according to the semantics of the value. This situation, however, is not the norm, as under f_3 , the local plan reaches the state [33, 11, 0, 0] instead of [11, 11, 11, 11]. This means that we cannot generally assume that just adhering to equitable principles will lead us into the most equitable goal. Still, the local plan is also admissible as aggregated behaviour, so it does preserve equity in that sense.

In general, the left blue plan can probably be conceived as most aligned with the value of equity (it is a goal plan, therefore reaches an equitable goal, and it is also part of the aggregated behaviours, so achieves overall equity). But notice that, to follow that plan, the second decision would need to be less equity-aligned than the other option, implying that it might not fully comply with legal requirements.

These problems may be addressed using the ϵ -local behaviour introduced in Sect. 3.2. With this behaviour, under certain circumstances, a small enough $\epsilon > 0$ may be tolerated (e.g. when all the villages get the Spanish minimum legal amount of water per inhabitant: 133l/inhab.) in the hope for finding better future alignment.

5 Conclusions and Future Work

In this paper, we analyze water allocation and human rights legislation to analyze value-aligned decision-making in a water scarcity scenario where preserving the value of equity is a legal requirement. Based on recent work, we formalize the value alignment problem with state-level preferences and semantics functions, characterizing not only the aggregated alignment of general paths but also plan value-admissibility criteria with the concept of behaviours. With a small water distribution in a drought situation example, we observe that a behaviour that conforms to the legislation (trying to preserve equity in each action) may lead to less equal states in the long term. As such, we ended up proposing a relaxed behaviour that could contemplate better future equity-aligned decisions without losing the law’s intentions regarding the value.

In future work, we propose using reinforcement learning considering value-admissibility behaviours. Different tasks can be investigated, such as learning an approximately optimal policy adhering to different behaviours simultaneously or one that adheres to the ϵ -local behaviour while maximizing others (e.g. aggregated/goal behaviour). Lastly, we highlight the problem of defining suitable value-aligned aggregation functions for generic (goal-oriented) decision-making problems.

Acknowledgements. This work has been supported by grant VAE: TED2021-131295B-C33 funded by MCIN/AEI/ 10.13039/501100011033 and by the “European Union NextGenerationEU/PRTR”, by grant COSASS: PID2021-123673OB-C32 funded by MCIN/AEI/ 10.13039/501100011033 and by “ERDF A way of making Europe”, and by the AGROBOTS Project of Universidad Rey Juan Carlos funded by the Community of Madrid, Spain.

References

1. Arnold, T., Kasenberg, D., Scheutz, M.: Value alignment or misalignment - what will keep systems accountable? In: AAAI Workshop on AI, Ethics, and Society (2017)
2. Christiano, P., Leike, J., Brown, T.B., Martic, M., Legg, S., Amodei, D.: Deep reinforcement learning from human preferences (2023)
3. Fürnkranz, J., Hüllermeier, E., Cheng, W., Park, S.H.: Preference-based reinforcement learning: a formal framework and a policy iteration algorithm. *Mach. Learn.* **89**, 123–156 (2012). <https://doi.org/10.1007/s10994-012-5313-8>
4. Government, S.: Strategic project for economic recovery and transformation of digitalization of the water cycle. Report 2022. Technical report, Ministry for the Ecological Transition and Demographic Challenge (2022)
5. Guo, T., Yuan, Y., Zhao, P.: Admission-based reinforcement-learning algorithm in sequential social dilemmas. *Appl. Sci.* **13**(3) (2023). <https://doi.org/10.3390/app13031807>. www.mdpi.com/2076-3417/13/3/1807
6. Jiang, J., Lu, Z.: Learning fairness in multi-agent systems. In: *Advances in Neural Information Processing Systems*, vol. 32 (2019)
7. Lera-Leri, R., Bistaffa, F., Serramia, M., Lopez-Sanchez, M., Rodriguez-Aguilar, J.: Towards pluralistic value alignment: aggregating value systems through l_p -regression. In: *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2022, Richland, SC*, pp. 780–788. International Foundation for Autonomous Agents and Multiagent Systems (2022)
8. Montes, N., Osman, N., Sierra, C., Slavkovik, M.: Value engineering for autonomous agents. *CoRR* abs/2302.08759 (2023). <https://doi.org/10.48550/arXiv.2302.08759>
9. Montes, N., Sierra, C.: Synthesis and properties of optimally value-aligned normative systems. *J. Artif. Intell. Res.* **74**, 1739–1774 (2022). <https://doi.org/10.1613/jair.1.13487>
10. Ng, A.Y., Russell, S.J.: Algorithms for inverse reinforcement learning. In: *Proceedings of the Seventeenth International Conference on Machine Learning*, pp. 663–670 (2000)
11. Perello-Moragues, A., Poch, M., Sauri, D., Popartan, L.A., Noriega, P.: Modelling domestic water use in metropolitan areas using socio-cognitive agents. *Water* **13**(8) (2021). <https://doi.org/10.3390/w13081024>. www.mdpi.com/2073-4441/13/8/1024
12. Plata-Pérez, L., Sánchez-Pérez, J., Sánchez-Sánchez, F.: An elementary characterization of the Gini index. *Math. Soc. Sci.* **74**, 79–83 (2015)
13. PricewaterhouseCoopers: La gestión del agua en España. análisis y retos del ciclo urbano del agua (2018). www.pwc.es/es/publicaciones/energia/assets/gestion-agua-2018-espana.pdf

14. Rodriguez-Soto, M., Serramia, M., Lopez-Sanchez, M., Rodriguez-Aguilar, J.A.: Instilling moral value alignment by means of multi-objective reinforcement learning. *Ethics Inf. Technol.* **24**, 9 (2022). <https://doi.org/10.1007/s10676-022-09635-0>
15. Sierra, C., Osman, N., Noriega, P., Sabater-Mir, J., Perelló, A.: Value alignment: a formal approach. CoRR abs/2110.09240 (2021). arxiv.org/abs/2110.09240
16. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. MIT Press, Cambridge (2018)
17. van der Weide, T.L., Dignum, F., Meyer, J.-J.C., Prakken, H., Vreeswijk, G.A.W.: Practical reasoning using values. In: McBurney, P., Rahwan, I., Parsons, S., Maudet, N. (eds.) ArgMAS 2009. LNCS (LNAI), vol. 6057, pp. 79–93. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-12805-9_5