ON REWARD DESIGN FOR ALIGNED DECISION-MAKING UNDER UNCERTAINTY

Anonymous authors
Paper under double-blind review

ABSTRACT

In human–AI cooperation, a utility function elicited from a human decision maker is routinely repurposed as a reward for the artificial agent. When the agent communicates set-valued predictions, extending the original utility to a reward over imprecise information is nontrivial, as the expanded utility has many more degrees of freedom than the original. Existing methods assume that the value of an imprecise piece of information reduces to an aggregate over its constituent precise states, an assumption that is frequently violated. We propose a general framework that makes no such reduction. Decider's behavior under uncertainty is explicitly modeled via *selection strategies* induced by decision criteria. We introduce the notion of *validity*, which ensures that a criterion respects the rationality encoded in the utility function, and the stronger notion of *alignment*, which guarantees that the resulting reward faithfully preserves the original preferences. We prove constructive compatibility results between classical criteria (Laplace, maximin, maximax, minimax regret) and standard aggregators (min, max, avg), and provide an algorithm to build aligned rewards. Our framework accommodates arbitrary sets of terminal acts, bridges the gap between utility elicitation and reward design for set-valued information, and generalizes naturally to richer uncertainty representations.

1 INTRODUCTION

Human-machine cooperation typically relies on defining **reward** schemes that encapsulate the human's goals Sutton & Barto (1998). A typical scenario is one where the human (which we call the Decision Maker, or DM) cannot access all relevant information to make a decision and may rely on information provided by an artificial agent (which we call the Information Assistant, or IA) Bansal et al. (2019). The standard procedure in this case is eliciting/learning a **utility** function from the DM and using it as a reward function for optimizing the IA's policy Sutton (2022); Christiano et al. (2023); Marklund et al. (2026).

When the IA is allowed to provide a richer representation of uncertainty, this conflation between the original utility and the reward can become non-trivial. For example, when the IA provides a disjunctive **set-valued prediction** Couso & Dubois (2014), extending the original utility function to an extended utility that can serve as reward has been extensively studied, and multiple approaches have been proposed Del Coz et al. (2009); Zaffalon et al. (2012); Yang et al. (2017); Ma & Denœux (2021); Imoussaten & Jacquin (2022); de Souza et al. (2025). However, these methods generally assume that the value of an imprecise piece of information can be evaluated by aggregating the values of the corresponding precise pieces. This assumption is not always valid: the expanded utility function has many more degrees of freedom than the original one Kunitomo-Jacquin et al. (2025).

We propose a **general framework** for extending a utility function to a reward function without assuming that the value of imprecise information reduces to an aggregate of precise values. Our framework allows the DM to express arbitrary terminal acts and models their behavior under uncertainty through selection strategies induced by decision criteria. We introduce the stronger notion of **alignment**, which guarantees that the resulting reward faithfully preserves the rationality encoded in the original utility, and show that it is frequently violated by existing approaches. We prove constructive compatibility results between classical criteria and standard aggregators, and provide an

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

algorithm to build aligned rewards. Our extension thus bridges the gap between utility elicitation and reward design for set-valued information.

The paper is organized as follows. Section 2 formalizes the problem. Section 3 reviews existing approaches and their limitations. Section 4 introduces selection strategies. Section 5 defines criteria and validity. Section 6 examines classical criteria under our framework. Section 7 defines alignment. Section 8 presents our constructive results. Section 9 provides some empirical validation. Section 10 concludes. The Appendix contains all proofs and additional results.

2 PROBLEM FORMULATION

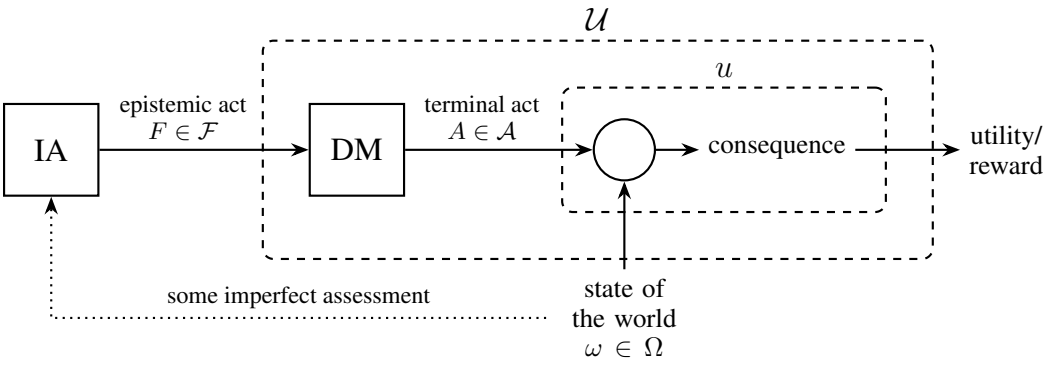


Figure 1: Illustration of the problem formulation. The IA performs an epistemic act $F \in \mathcal{F}$ (in our examples a set-valued prediction) that imperfectly assesses the state of the world $\omega \in \Omega$. Based on F, the DM selects a terminal act $A \in \mathcal{A}$, which together with ω determines a consequence whose utility is $u(A, \omega)$. The reward function $\mathcal{U}$ extends u to non-terminal acts. The reward $\mathcal{U}(F, \omega)$ should capture the utility to the DM of receiving information F when the true state is ω.

Let Ω denote the finite set of possible **states of the world** and consider two entities, an **Information Assistant (IA)** and a **Decision Maker (DM)**. The IA can (imperfectly) assess the actual state of the world and communicate this information to the DM. We denote by $\mathcal{F} \subseteq 2^{\Omega} \setminus \{\emptyset\}$ the set of possible pieces of information that the IA can provide and call the action of providing such information an **epistemic act**.

We consider a disjunctive interpretation of the epistemic acts: each $F \in \mathcal{F}$ represents the set of states that the IA considers possible given its assessment. When the IA provides the information F, it communicates to the DM that "the actual state of the world is in F, but no more specific information is available".

In response to this information, the DM selects a single act to perform. We denote by $\mathcal{A} \subseteq \mathcal{F}$ the set of available actions for the DM, which we call **terminal acts**[1]. Performing a terminal act $A \in \mathcal{A}$ when the world is in state $\omega \in \Omega$ yields a unique, known consequence, which we represent by convention as the pair in $\mathcal{A} \times \Omega$. Consequences provide different degrees of satisfaction to the DM, and this satisfaction is encoded by a **utility function**

$$u : \mathcal{A} \times \Omega \to [0, 1],$$

which we assume is known in advance.

Our goal is to evaluate the reward that should be assigned to any possible epistemic act of the IA so that its preferences align with the DM. That is, we want to define a function

$$\mathcal{U} : \mathcal{F} \times \Omega \to [0, 1]$$

[1]In decision theory, acts are mappings from states of the world to a set of consequences Savage (1954). In this work, by denoting $\mathcal{A}$ the set of terminal acts, we mean that the set of acts available to the DM can be written as $\tilde{\mathcal{A}} = \{\delta_B\}_{B \in \mathcal{A}}$, where the DM always performs the act δ_A when the information provided by the IA corresponds to $A \in \mathcal{A}$. We abuse notation and simply denote the act δ_A by A.

that coincides with u on $\mathcal{A} \times \Omega$, such that $\mathcal{U}(F, \omega)$ equals the utility to the DM of having the information represented by F when the actual state of the world is ω. Throughout the paper, we call $\mathcal{U}$ the **reward function** and $\mathcal{A}^c = \mathcal{F} \setminus \mathcal{A}$ the set of non-terminal acts. Figure 1 illustrates the problem formulation.

To make the framework concrete, we introduce a medical diagnosis example.

Example 2.1. Consider a patient presenting with symptoms consistent with four admissible diagnoses (possible states of the world), two cardiac diseases and two respiratory diseases:

$$\Omega = \{C_1, C_2, R_1, R_2\}.$$

The medical team acts as the DM. It can list the available treatments (the terminal acts) and evaluate the corresponding consequences. The available terminal acts are

$$\mathcal{A} = \{\Omega, \{C_1, C_2\}, \{R_1, R_2\}, \{C_1\}, \{C_2\}, \{R_1\}, \{R_2\}\},$$

corresponding respectively to a general protocol valid for all diseases (Ω), a general cardiac treatment ($\{C_1, C_2\}$), a general respiratory treatment ($\{R_1, R_2\}$), and the immediate treatment of a specific disease (each of $\{C_1\}, \{C_2\}, \{R_1\}, \{R_2\}$).

As in the general formulation, $\mathcal{A} \subseteq \mathcal{F}$, so every treatment is also a piece of information that the IA may communicate. For instance, the act $\{C_1, C_2\}$ denotes both the prediction "the disease is either C_1 or C_2" (when read as an epistemic act) and the decision to apply a general cardiac treatment (when read as a terminal act). Table 1 makes this correspondence explicit by reading each set $A \in \mathcal{A}$ in both roles, while Table 2 encodes the DM's satisfaction for each consequence.

$A \in \mathcal{A}$	epistemic act	terminal act
$\{C_1\}$	pred. "C_1"	treat C_1
$\{C_2\}$	pred. "C_2"	treat C_2
$\{R_1\}$	pred. "R_1"	treat R_1
$\{R_2\}$	pred. "R_2"	treat R_2
$\{C_1, C_2\}$	pred. "C_1 or C_2"	cardiac treat.
$\{R_1, R_2\}$	pred. "R_1 or R_2"	respiratory treat.
Ω	ignorance	general protocol

Table 1: Terminal acts of the medical diagnosis example in their two roles. Each act is both a prediction the IA may communicate (epistemic act) and a treatment the DM applies in response (terminal act). The first four acts target a single disease, the next two a disease category, and the last abstains from predicting while applying a general protocol.

$u(A, \omega)$	C_1	C_2	R_1	R_2
$\{C_1\}$	1	0	0	0
$\{C_2\}$	0	1	0	0
$\{R_1\}$	0	0	1	0
$\{R_2\}$	0	0	0	1
$\{C_1, C_2\}$	0.9	0.9	0	0
$\{R_1, R_2\}$	0	0	0.9	0.9
Ω	0.8	0.8	0.8	0.8

Table 2: Utility function for the medical diagnosis example. Treating a specific disease yields the best outcome in that state and the worst in all others; broader treatments trade off accuracy for robustness across disease categories.

The medical team is assisted by an IA machine-learning system that can provide the DM, as information, any disjunctive set in $\mathcal{F} = 2^\Omega \setminus \{\emptyset\}$. Our goal is to evaluate the reward that should correspond to each possible epistemic act. For instance, the initial set of terminal acts does not allow us to straightforwardly assign a reward when the available information corresponds to the disjunctive set $F = \{C_1, R_1\}$ (i.e., when the IA is hesitating between a specific cardiac and a specific respiratory disease, while ruling out the remaining two). We aim to define $\mathcal{U}(F, \omega)$ for all non-empty $F \subseteq \Omega$ and $\omega \in \Omega$.

3 Related Works: Existing Approaches and Their Limitations

A common approach in the literature is to define the reward function $\mathcal{U}$ by aggregating the utilities of the underlying precise pieces. To this end, one fixes a symmetric (invariant to permutation) function $\mathrm{agg} : \bigcup_{k \geq 1} [0,1]^k \to [0,1]$ such that $\mathrm{agg}(x, x, ..., x) = x$, and sets

$$\mathcal{U}(F, \omega) = \mathrm{agg}(u(\omega_1, \omega), \ldots, u(\omega_n, \omega)), \tag{1}$$

for any $F = \{\omega_1, \ldots, \omega_n\} \in \mathcal{F}$ and $\omega \in \Omega$.

The simplest aggregators are the average, the minimum, and the maximum, but many others exist. Yang et al. (2017) proposes p-discounted costs, generalizing the average, with a parameter r calibrating the reward for cautiousness/mistake-aversion. Ma & Denœux (2021) defines the aggregator using Ordered Weighted Average (OWA) operators, where a single parameter γ models the decision maker's tolerance for imprecision, ranging from the average to the maximum at the extremes.

We identify two main issues in the existing literature.

1. These methods assume $\mathcal{A} = \{\{\omega\}\}_{\omega\in\Omega}$, which need not hold: applying any of them to Example 2.1 would exclude the terminal acts $\{C_1, C_2\}$, $\{R_1, R_2\}$, and Ω from the evaluation of $\mathcal{U}$. To the best of our knowledge, no prior work addresses this issue while allowing a general $\mathcal{A} \subseteq \mathcal{F}$.
2. Even when $\mathcal{A} = \{\{\omega\}\}_{\omega\in\Omega}$, this aggregation approach is not always appropriate. To model the DM's behavior, it presupposes a specific (and frequently wrong) assumption: that, *given an imprecise piece of information F, the DM will select some $\omega \in F$ and act as if that state were the truth*. In effect, the DM is assumed to decide solely from the states in F, disregarding their own utility function.

The following example illustrates the second point.

Example 3.1. Consider $\mathcal{A} = \{\{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_4\}\}$, $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, and the utility function given in Table 3 with some $p \in (\frac{1}{2}, 1)$.

$u(A,\omega)$	ω_1	ω_2	ω_3	ω_4
$\{\omega_1\}$	1	0	0	0
$\{\omega_2\}$	0	1	0	0
$\{\omega_3\}$	p	p	1	0
$\{\omega_4\}$	0	0	0	1

Table 3: Utility function for Example 3.1.

Consider the two epistemic acts $F_1 = \{\omega_1, \omega_2\}$ and $F_2 = \{\omega_3, \omega_4\}$. By Equation (1), we have $\mathcal{U}(F_1, \omega_1) = \mathcal{U}(F_2, \omega_3) = \text{agg}(1, 0)$, yet they correspond to very different situations.

With F_2, the DM rules out ω_1 and ω_2 and is left with $\{\omega_3\}$ or $\{\omega_4\}$ as the only reasonable options, any other act yielding a sure loss. Both amount to a gamble between the best and worst outcomes.

With F_1, the DM is likewise uncertain between two states and can gamble on $\{\omega_1\}$ or $\{\omega_2\}$. But here choosing $\{\omega_3\}$ yields a sure utility of p for every $\omega \in F_1$. For sufficiently high p, a risk-averse DM will prefer $\{\omega_3\}$ over a gamble. Thus F_1 is worth p to the DM, strictly more than F_2. Using the same aggregation function for both acts is therefore inadequate: it disregards the structure of u and leads to misalignment.

Other works Zaffalon et al. (2012); Del Coz et al. (2009); Imoussaten & Jacquin (2022); de Souza et al. (2025) define $\mathcal{U}$ directly, without an underlying utility function u, typically from the cardinality of the set and a risk-seeking or risk-averse parameter. While these approaches yield well-defined rewards, their practical usefulness is limited because they constrain the expressiveness of the DM's preferences.

In Yang et al. (2017); Kunitomo-Jacquin et al. (2025), the authors propose desirable properties for any reward function $\mathcal{U}$, which we argue to be too restrictive in the general case. For example, both stress that error-averse behavior is desirable (that is, $\omega \in F, \omega \notin F' \implies \mathcal{U}(F, \omega) > \mathcal{U}(F', \omega)$). This property can be useful in some cases but is certainly not general. For instance, in Example 3.1, it is certainly rational for some DM to have a preference such that $\mathcal{U}(\{\omega_3\}, \omega_1) > \mathcal{U}(\{\omega_1, \omega_4\}, \omega_1)$.

4 Introducing Strategies

To extend u to a reward function $\mathcal{U}$ defined on all epistemic acts, we must model how the DM behaves when given imprecise information. The DM can evaluate utilities only for terminal acts, but when the IA provides some $F \in \mathcal{A}^c$, the DM must still ultimately perform a terminal act. We

represent via a **selection strategy** the terminal acts considered admissible by the DM under each piece of information.

Definition 4.1. A **selection strategy** is a map $\tau : \mathcal{F} \to 2^{\mathcal{A}} \setminus \emptyset$ that assigns to each epistemic act $F \in \mathcal{F}$ a set $\tau(F) \subseteq \mathcal{A}$ of terminal acts.

The set $\tau(F)$ collects the terminal acts that the DM considers acceptable, or potentially optimal, when the information provided by the IA corresponds to F.

Definition 4.2. A selection strategy τ is **valid** (under u) if it is the identity on terminal acts, i.e., if $\tau(A) = \{A\}$ for all $A \in \mathcal{A}$.

Validity reflects the modeling assumption introduced in Section 2: when the IA provides information corresponding to a terminal act $A \in \mathcal{A}$, the DM performs A. For non-terminal acts $F \in \mathcal{A}^c$, the strategy may return several terminal acts, capturing the idea that multiple choices may be reasonable. If a strategy τ satisfies only the weaker condition $A \in \tau(A)$ for all $A \in \mathcal{A}$, we call it **validable**. A validable strategy can be made valid by removing from each $\tau(A)$ any act other than A.

Our approach generalizes the implicit assumption of prior work discussed in Section 3. There, the DM, when given an imprecise piece of information F, is assumed to pick some $\omega \in F$. In our framework, the DM finally selects a terminal act $A \in \tau(F)$, which may depend on the structure of u and is not necessarily singleton from F. Analogously to Equation (1), we aim to define $\mathcal{U}$ by aggregating the utilities of the selected terminal acts. We are thus left with two problems:

1. How to define valid(able) strategies that faithfully capture the DM's behavior.
2. How to define proper aggregation schemes that produce a coherent reward.

We address these in the following.

5 Introducing Criteria

A natural way to define selection strategies is through **criteria**, scoring rules that rank terminal acts given the information F.

Definition 5.1. We call **criterion** a family $\mathbf{g} = \{\mathbf{g}_F^u\}_{F \in \mathcal{F}}$ of maps $\mathbf{g}_F^u : \mathcal{A} \to \mathbb{R}$.

A criterion $\mathbf{g}$ induces a family of weak orders ${}^{\mathbf{g}}\succeq_F^u$ over $\mathcal{A}$ via

$$\mathbf{g}_F^u(A) \geq \mathbf{g}_F^u(B) \iff A \, {}^{\mathbf{g}}\succeq_F^u B. \tag{2}$$

The associated strict order is denoted by ${}^{\mathbf{g}}\succ_F^u$. For any two sets of acts $S_1, S_2 \subseteq \mathcal{A}$, we also write $S_1 \, {}^{\mathbf{g}}\succeq_F^u S_2$ if the relation holds for all $A \in S_1$ and $B \in S_2$.

Definition 5.2. Let $\mathbf{g}$ be a criterion. An act $A \in \mathcal{A}$ is **$\mathbf{g}$-disposable** (with respect to u) if there exists $B \in \mathcal{A}$, $B \neq A$, such that $B \, {}^{\mathbf{g}}\succ_A^u A$.

Thus, a disposable act $A \in \mathcal{A}$ is one that the criterion itself discards when the information provided is precisely A.

A strategy can be derived directly from a criterion: $\tau_{\mathbf{g}} : \mathcal{F} \to 2^{\mathcal{A}}$ selects the maximal elements under this order,

$$\tau_{\mathbf{g}}(F) = \arg\max_{A \in \mathcal{A}} \mathbf{g}_F^u(A), \qquad F \in \mathcal{F}. \tag{3}$$

The link between disposability and strategy validity is immediate:

Lemma 5.3. *A criterion $\mathbf{g}$ has no $\mathbf{g}$-disposable acts if and only if the induced strategy $\tau_{\mathbf{g}}$ is validable.*

See Appendix A for the proof. Indeed, if A is $\mathbf{g}$-disposable, then $\tau_{\mathbf{g}}$ discards A under information A.

We accordingly call a criterion $\mathbf{g}$ **valid** (with respect to u) if it has no disposable acts. When choosing a criterion to define a strategy, a rational DM will restrict attention to valid ones. Given a valid criterion $\mathbf{g}$, we denote by

$$\tau_{\mathbf{g}}^*(F) = \begin{cases} \{A\}, & \text{if } F = A \in \mathcal{A}, \\ \tau_{\mathbf{g}}(F), & \text{if } F \in \mathcal{A}^c, \end{cases} \tag{4}$$

the (always valid) strategy that can be obtained from $\mathbf{g}$.

6 SOME EXAMPLES

We present some criteria that are classical for decision under uncertainty Savage (1954). For all $F \in \mathcal{F} \subseteq 2^\Omega$ and $A \in \mathcal{A}$, the **average** (Laplace), **maximin**, **maximax**, and **minimax-regret** criteria are respectively defined by

$$\begin{aligned}
\mathbf{avg}^u_F(A) &= \frac{1}{|F|}\sum_{\omega\in F} u(A,\omega),\\
\mathbf{min}^u_F(A) &= \min_{\omega\in F} u(A,\omega),\\
\mathbf{max}^u_F(A) &= \max_{\omega\in F} u(A,\omega),\\
\mathbf{reg}^u_F(A) &= \min_{\substack{B\in\mathcal{A}\\ \omega\in F}} (u(A,\omega)-u(B,\omega)).
\end{aligned}$$

Another particularly natural criterion, which we call the **non-dominated**, selects all terminal acts that are not Pareto-dominated, within the information set F, by any other terminal act for the preferences in u:

$$\mathbf{nd}^u_F(A) = \begin{cases} 1, & \text{if } \nexists B \in \mathcal{A} : u(B,\omega) \geq u(A,\omega) \text{ for all } \omega \in F,\\ & \text{and } u(B,\omega') > u(A,\omega') \text{ for some } \omega' \in F,\\ 0, & \text{otherwise.}\end{cases} \tag{5}$$

The structure of u already encodes some information about the DM's behavior that not all criteria preserve. We illustrate this with the medical diagnosis example, where some classical criteria turn out not to be valid with respect to the given u.

Example 6.1. Consider the medical diagnosis utility function u from Example 2.1. The criteria **nd**, **avg**, **min**, and **reg** are valid with u, whereas **max** is *not*. Indeed, taking $F = \{C_1, C_2\}$ we have

$$\mathbf{max}^u_F(\{C_1,C_2\}) = 0.9 \quad \text{but} \quad \mathbf{max}^u_F(\{C_1\}) = 1.$$

Hence $\{C_1\} \overset{\mathbf{max}}{\succ}{}^u_F \{C_1,C_2\}$, so $\{C_1,C_2\}$ is **max**-disposable and **max** is not valid with u.

Selecting a valid criterion is not yet sufficient to obtain a satisfactory reward function $\mathcal{U}$. Once a strategy τ is chosen, one must still aggregate the utilities of the selected terminal acts to define $\mathcal{U}$. The following example shows that, even when τ is valid, a naive aggregation can induce preferences that contradict the rationality encoded in u.

Example 6.2. Consider $\mathcal{A} = \{\{\omega_1\},\{\omega_2\},\{\omega_3\},\{\omega_1,\omega_2\}\}$, $\Omega = \{\omega_1,\omega_2,\omega_3\}$, and the utility function given in Table 4 with $p \in (0, \frac{1}{3})$.

$u(A,\omega)$	ω_1	ω_2	ω_3
$\{\omega_1\}$	1	0	0
$\{\omega_2\}$	0	1	0
$\{\omega_3\}$	0	0	1
$\{\omega_1,\omega_2\}$	p	p	0

Table 4: Utility function for Example 6.2.

The non-dominated strategy is validable with respect to u, and we have $\tau_{\mathbf{nd}}(\Omega) = \{\{\omega_1\},\{\omega_2\},\{\omega_3\},\{\omega_1,\omega_2\}\}$. Constructing $\mathcal{U}$ by averaging over $\tau^*_{\mathbf{nd}}(F)$ yields

$$\mathcal{U}(\Omega,\omega_1) = \mathcal{U}(\Omega,\omega_2) = \frac{1+p}{4}, \qquad \mathcal{U}(\Omega,\omega_3) = \frac{1}{4}.$$

Since $\mathcal{U}(\Omega,\omega) > u(\{\omega_1,\omega_2\},\omega)$ for every $\omega \in \Omega$ when $p < \frac{1}{3}$,

$$\Omega \overset{\mathbf{nd}}{\succ}{}^{\mathcal{U}}_{\{\omega_1,\omega_2\}} \{\omega_1,\omega_2\},$$

and $\{\omega_1,\omega_2\}$ becomes **nd**-disposable with respect to $\mathcal{U}$.

In Example 6.2, the obtained reward ranks the ignorance epistemic act Ω strictly above the terminal act $\{\omega_1, \omega_2\}$ for every possible state of the world. If the DM accepts this reward, they must conclude that $\{\omega_1, \omega_2\}$ should never be chosen, and then u would not faithfully represent their preferences. Conversely, if the DM insists that u is rational and no terminal act is disposable, then $\mathcal{U}$ cannot serve as a valid extension of u. More generally, whenever a reward makes a strategy that was valid under u no longer validable under $\mathcal{U}$, the DM's preferences are not encoded in $\mathcal{U}$.

Validity of the strategy alone thus does not guarantee that the aggregated reward preserves the rationality encoded in u: a naive aggregation, even over a well-chosen strategy, can reintroduce exactly the inconsistency that validity was meant to avoid. In the next section we discuss the stronger notion of **alignment**, which ensures that the reward faithfully reflects the preferences encoded in both u and the chosen valid criterion.

7 Constructing Aligned Rewards

A valid criterion $\mathbf{g}$ still leaves open how to aggregate the utilities of the selected terminal acts into a reward function $\mathcal{U}$. As the previous examples show, doing so naively can reintroduce the very inconsistencies that validity was meant to prevent. We therefore require a stronger condition.

Definition 7.1. Let $\mathbf{g}$ be a criterion valid under u and τ be a strategy. We say a reward $\mathcal{U} : \mathcal{F} \times \Omega \to [0,1]$ is **g-aligned** with u under τ if for all $A \in \mathcal{A}^c$, $F \in \mathcal{F}$ and $B, C \in \mathcal{A}$,

$$B \,{}^{\mathbf{g}}\!\succeq^{u}_{F} \tau(A) \,{}^{\mathbf{g}}\!\succeq^{u}_{F} C \;\Rightarrow\; B \,{}^{\mathbf{g}}\!\succeq^{\mathcal{U}}_{F} A \,{}^{\mathbf{g}}\!\succeq^{\mathcal{U}}_{F} C.$$

Thus, alignment requires that the preferences encoded in u via $\mathbf{g}$ are preserved when $\mathcal{U}$ replaces u. The following result shows that it implies the non-disposability requirement discussed in Section 6.

Theorem 7.2. *Let $\mathbf{g}$ valid under u. If $\mathcal{U}$ is $\mathbf{g}$-aligned with u under $\tau^*_{\mathbf{g}}$, then there are no $\mathbf{g}$-disposable acts under $\mathcal{U}$.*

Theorem 7.2 (proved in Appendix A) guarantees that an $\mathbf{g}$-aligned reward never creates $\mathbf{g}$-disposability. The pathologies of Examples 3.1 and 6.2 therefore cannot arise.

8 Aggregation-Based Rewards

We now turn to the constructive question: given a valid criterion $\mathbf{g}$, how to build an aligned reward. Following the aggregation paradigm of Equation (1), we define rewards by aggregating the utilities of the terminal acts selected by $\tau^*_{\mathbf{g}}$.

Let $\mathrm{agg} : \bigcup_{k\geq 1}[0,1]^k \to [0,1]$ be a symmetric aggregation function such that $\mathrm{agg}(x, x, ..., x) = x$. For a strategy τ, the corresponding aggregated reward $\mathcal{U}^{\tau}_{\mathrm{agg}}$ is defined, for all $F \in \mathcal{F}$ and $\omega \in \Omega$, by

$$\mathcal{U}^{\tau}_{\mathrm{agg}}(F, \omega) = \mathrm{agg}\big(u(A_1, \omega), \ldots, u(A_n, \omega)\big), \qquad \tau(F) = \{A_1, \ldots, A_n\}. \tag{6}$$

For a terminal act $A \in \mathcal{A}$, we have $\tau^*_{\mathbf{g}}(A) = \{A\}$, so $\mathcal{U}^{\tau^*_{\mathbf{g}}}_{\mathrm{agg}}(A, \omega) = u(A, \omega)$, as required.

The following theorems establish this compatibility for some of the classical criteria introduced in Section 6.

Theorem 8.1. *If **avg** is valid under u, then $\mathcal{U}^{\tau^*_{avg}}_{avg}$ is **avg**-aligned.*

Theorem 8.2. *If **min** is valid under u, then $\mathcal{U}^{\tau^*_{min}}_{min}$ is **min**-aligned.*

Theorem 8.3. *If **max** is valid under u, then $\mathcal{U}^{\tau^*_{max}}_{max}$ is **max**-aligned.*

Theorem 8.4. *If **reg** is valid under u, then $\mathcal{U}^{\tau^*_{reg}}_{min}$ is **reg**-aligned.*

In particular, Theorems 8.1–8.4 (proofs in Appendix A) show that the average criterion is compatible with the average aggregator, the maximin and minimax-regret criteria are both compatible with the minimum aggregator, and the maximax criterion is compatible with the maximum aggregator. Whenever validity can be verified under one of these criteria, an aligned reward is guaranteed to exist.

Algorithm 1 summarizes the overall procedure. The DM first identifies which criteria are valid under u, then selects one of them together with a compatible aggregator. The reward is built by assembling the strategy induced by the criterion and aggregating the resulting utilities.

Algorithm 1: Construct an aligned reward.

Require: Utility $u : \mathcal{A} \times \Omega \to [0, 1]$, a set of criteria $\mathcal{G}$, a set of aggregators Λ, and a compatibility map $\mathcal{C} : \mathcal{G} \to 2^{\Lambda}$

Ensure: A reward $\mathcal{U} : \mathcal{F} \times \Omega \to [0, 1]$

$\mathcal{G}_{\text{val}} \leftarrow \{\mathbf{g} \in \mathcal{G} : \mathbf{g} \text{ is valid under } u\}$
ask the DM to choose $\mathbf{g} \in \mathcal{G}_{\text{val}}$ and $\text{agg} \in \mathcal{C}(\mathbf{g})$
for all $F \in \mathcal{A}$ **do**
  $\mathcal{U}(F, \omega) \leftarrow u(F, \omega)$
end for
for all $F \in \mathcal{A}^c$ **do**
  $\tau^*_{\mathbf{g}}(F) \leftarrow \arg\max_{A \in \mathcal{A}} \mathbf{g}^u_F(A)$
  $\{A_1, \ldots, A_n\} \leftarrow \tau^*_{\mathbf{g}}(F)$
  $\mathcal{U}(F, \omega) \leftarrow \text{agg}\big(u(A_1, \omega), \ldots, u(A_n, \omega)\big)$
end for
return $\mathcal{U}$

The compatibility map $\mathcal{C}$ can be populated with the pairings established in Theorems 8.1–8.4.

9 Illustrative Experiment

A concern that may arise about the practical relevance of our present approach is the question of how widespread the problem is. The pathologies of Examples 3.1 and 6.2 could in principle be dismissed as ad hoc constructions, exceptions to a general rule of "well-behaved" utilities. To test this hypothesis, we conducted an experiment to measure the frequency of disposable acts over a broad random sample of utility functions. We sampled 100 utility functions u on $\mathcal{A} = \{\{\omega\}\}_{\omega \in \Omega}$, for $|\Omega| \in \{2, 3, 4, 5\}$ and satisfying the non-disposability of terminal acts, i.e., $u(\{\omega\}, \omega) > u(\{\omega\}, \omega')$ for $\omega \neq \omega'$. For these utility functions, the criteria **avg**, **min**, **max**, and **reg** are always valid, and it is always possible to extend each of them to $\mathcal{F} = 2^{\Omega} \setminus \{\emptyset\}$ via Algorithm 1, ensuring that the resulting reward $\mathcal{U}$ is aligned and has no disposable acts.

For comparison, we applied the naive approach described in Section 3 to these utility functions and computed the ratio of epistemic acts that become disposable under the induced reward for different aggregators. Figure 2 reports these ratios for different decider criteria discussed. The naive approach produces a substantial proportion of disposable epistemic acts under every aggregator and criterion, and the ratio increases with $|\Omega|$, exceeding 90% for the $\max$ and $\min$ aggregators already at $|\Omega| = 5$. The problem we address is then not a marginal or adversarial phenomenon. In higher dimensions, most epistemic acts under a naive reward are disposable, and these rewards fail consistently to capture the DM's preferences.

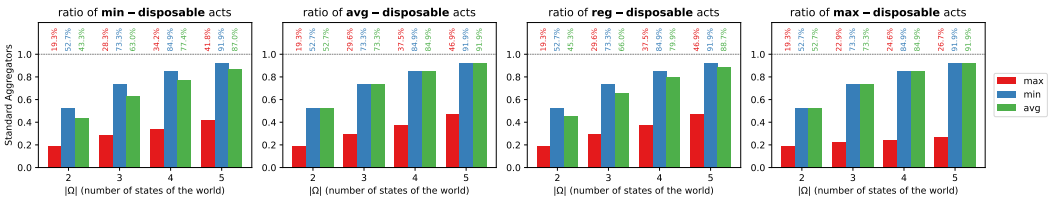


Figure 2: Ratio of disposable epistemic acts when naively applying the three standard aggregators ($\max$, $\min$, and $\max$) as a function of the number of states of the world. Each bar reports the average fraction of epistemic acts that are **g**-disposable under the corresponding decider criterion, over 100 randomly generated utility functions. Columns correspond to the maximin, Laplace, minimax-regret, and maximax decider criteria, discussed in Section 6. In higher dimensions, the naive approach leads to most epistemic acts being disposable.

Appendix C presents additional results. We first show that the more complex aggregators introduced by Yang et al. (2017); Ma & Denœux (2021) do not fundamentally change the picture. We then depart from the ideal setting of an aligned reward by measuring the ratio of disposable acts under a misuse of Algorithm 1, in which the DM does not identify their actual criterion. Even in this case, the ratio of disposable acts remains significantly lower than under the naive approach.

10 Conclusion and Perspectives

We introduced a general framework for extending a utility function to a reward over set-valued information. The core idea was introducing selection strategies, which incorporate into the reward the DM's behavior when faced with imprecise information. This avoided the implicit assumptions that constrain prior work. Building on criteria, we defined valid strategies that respect the rationality encoded in the utility function. Via compatible aggregators, we then constructed aligned rewards that preserve the DM's preferences.

A practical concern is the case where none of the criteria paired with aggregators in Theorems 8.1–8.4 is valid for a given u. We emphasize that this never occurs under the assumptions standard in the reviewed literature: $\mathcal{A} = \{\{\omega\}\}_{\omega\in\Omega}$ and $u(\{\omega\}, \omega) > u(\{\omega\}, \omega')$ for $\omega \neq \omega'$. Moreover, our method applies well beyond these conditions, as the medical diagnosis example demonstrates. The search for new valid criteria and compatible aggregators is, however, an interesting direction for future work. A natural extension would be to introduce parameters that interpolate between different valid strategies and criteria while preserving alignment, yielding a single parametric family that captures a wider spectrum of risk-aversion profiles.

The computational bottleneck of Algorithm 1 lies in the validity check: for each criterion $\mathbf{g} \in \mathcal{G}$, one must verify that no terminal act is $\mathbf{g}$-disposable. This takes $|\mathcal{G}| \cdot |\mathcal{A}| \cdot |A|$ operations per act $A \in \mathcal{A}$. In the typical case where $\mathcal{A} = \{\{\omega\}\}_{\omega\in\Omega}$, the total cost is $|\mathcal{G}| \cdot |\mathcal{A}|^2 = |\mathcal{G}| \cdot |\Omega|^2$. For the worst case, when $\mathcal{A} \approx 2^\Omega$, the bound is $|\mathcal{G}| \cdot |\mathcal{A}|^2 \cdot |\Omega|/2 = |\mathcal{G}| \cdot 2^{2|\Omega|} \cdot |\Omega|/2$. We intentionally imposed minimal structure on u. Further structural hypotheses can reduce this burden and yield sharper criteria for specific applications.

Several generalizations are within reach. Allowing $\mathcal{F}$ to contain probability distributions, possibility distributions Dubois & Prade (1988), mass of belief functions Shafer (1976), or imprecise probabilities Walley (1991) does not alter the core problem. With suitable adaptations to the definitions of strategies, criteria, and non-disposability, analogues of Lemma 5.3 and Theorem 7.2 should hold, and Algorithm 1 would remain applicable, though new compatibility results akin to Theorems 8.1–8.4 would be required. Extending the framework to infinite Ω is another interesting direction, and we believe similar adaptations can carry the main results over.

We assumed throughout that $\mathcal{A} \subseteq \mathcal{F}$, i.e., every terminal act is a priori an available epistemic act. This assumption is central to Lemma 5.3. But when it is violated our method can still be applied: Appendix B sketches a procedure that augments the acts and then restricts the resulting reward.

Although we motivated the setup by casting the DM as human and the IA as machine, this is certainly not a requirement. The roles can be swapped or assigned in any combination. This is not merely a conceptual point: in practice, human preferences and behaviors shift in response to AI decisions Bansal et al. (2019). A cautious AI may, for instance, make the DM less cautious, and well-documented cognitive biases exist Green & Chen (2019); Recchia et al. (2026). We also leave aside the dynamic dimension of repeated interactions, which carries non-trivial implications Marklund et al. (2026); Sutton & Barto (1998). Related to this, however, our framework addresses an ethical implication overlooked in prior work. If a reward has a disposable act A, then the IA, when it knows that A is the case, will always report something else to the DM in the name of maximizing their utility. This is a paternalistic behaviour, with consequences that are both moral Tolmeijer et al. (2021) and practical, since the DM's trust in the IA is a prerequisite of successful cooperation. By eliminating disposable acts, our aligned rewards remove this incentive to "deceive".

Finally, this framework fills a gap between utility theory and reward design for human-AI cooperation. By making explicit the DM's behavior under uncertainty, we avoid both the hidden assumptions of existing aggregation-based methods and the rigidity of purely axiomatic approaches. The framework thus offers a principled foundation for building aligned trustworthy AI systems that communicate uncertainty while faithfully respecting human preferences.

AI Disclosure

In this paper, we used LLM-based AI tools for minor text polishing and proofreading. We did not rely on such assistance for the literature survey or for formulating or proving mathematical claims. The implementation of the practical experiments was also aided by generative tools. We take full responsibility for the final content of this work.

References

Gagan Bansal, Besmira Nushi, Ece Kamar, Walter S. Lasecki, Daniel S. Weld, and Eric Horvitz. Beyond Accuracy: The Role of Mental Models in Human-AI Team Performance. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, 7:2–11, October 2019. ISSN 2769-1349. doi: 10.1609/hcomp.v7i1.5285.

Paul Christiano, Jan Leike, Tom B. Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences, February 2023.

Inés Couso and Didier Dubois. Statistical reasoning with set-valued information: Ontic vs. epistemic views. *International Journal of Approximate Reasoning*, 55(7):1502–1518, October 2014. ISSN 0888613X. doi: 10.1016/j.ijar.2013.07.002.

Victor F. Lopes de Souza, Karima Bakhti, Sofiane Ramdani, Denis Mottet, and Abdelhak Imoussaten. Explainable Evidential Clustering, July 2025.

Juan José Del Coz, Jorge Díez, and Antonio Bahamonde. Learning Nondeterministic Classifiers. *Journal of Machine Learning Research*, 10(10), 2009.

Didler Dubois and Henri Prade. Representation and combination of uncertainty with belief functions and possibility measures. *Computational Intelligence*, 4(3):244–264, 1988. ISSN 1467-8640. doi: 10.1111/j.1467-8640.1988.tb00279.x.

Ben Green and Yiling Chen. The Principles and Limits of Algorithm-in-the-Loop Decision Making. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–24, November 2019. ISSN 2573-0142. doi: 10.1145/3359152.

Abdelhak Imoussaten and Lucie Jacquin. Cautious classification based on belief functions theory and imprecise relabelling. *International Journal of Approximate Reasoning*, 142:130–146, March 2022. ISSN 0888-613X. doi: 10.1016/j.ijar.2021.11.009.

Lucie Kunitomo-Jacquin, Abdelhak Imoussaten, Sebastien Destercke, Christophe Marsala, and Ken Fukuda. Towards evaluating set-valued predictions with partial observations. In *2025 IEEE International Conference on Fuzzy Systems (FUZZ)*, pp. 1–4, July 2025. doi: 10.1109/FUZZ62266.2025.11152051.

Liyao Ma and Thierry Denœux. Partial classification in the belief function framework. *Knowledge-Based Systems*, 214:106742, February 2021. ISSN 09507051. doi: 10.1016/j.knosys.2021.106742.

Henrik Marklund, Alex Infanger, and Benjamin Van Roy. Misalignment from Treating Means as Ends. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(44):37719–37727, March 2026. ISSN 2374-3468. doi: 10.1609/aaai.v40i44.41107.

Gabriel Recchia, Chatrik Singh Mangat, Jinu Nyachhyon, Mridul Sharma, Callum Canavan, Dylan Epstein-Gross, and Muhammed Abdulbari. Confirmation Bias: A Challenge for Scalable Oversight. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(44):37877–37886, March 2026. ISSN 2374-3468. doi: 10.1609/aaai.v40i44.41124.

Leonard J. Savage. *The Foundations of Statistics*. John Wiley and Sons, New York, 1954. ISBN 978-0-486-62349-8 978-0-486-13710-0.

Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, April 1976. ISBN 978-0-691-10042-5.

Richard S. Sutton. The Quest for a Common Model of the Intelligent Decision Maker, June 2022.

Richard S Sutton and Andrew G Barto. Reinforcement Learning: An Introduction. 1998.

Suzanne Tolmeijer, Markus Kneer, Cristina Sarasua, Markus Christen, and Abraham Bernstein. Implementations in Machine Ethics: A Survey. *ACM Computing Surveys*, 53(6):1–38, November 2021. ISSN 0360-0300, 1557-7341. doi: 10.1145/3419633.

Peter Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman & Hall, 1991.

Gen Yang, Sebastien Destercke, and Marie-Helene Masson. The Costs of Indeterminacy: How to Determine Them? *IEEE Transactions on Cybernetics*, 47(12):4316–4327, December 2017. ISSN 2168-2267, 2168-2275. doi: 10.1109/TCYB.2016.2607237.

Marco Zaffalon, Giorgio Corani, and Denis Mauá. Evaluating credal classifiers by utility-discounted predictive accuracy. *International Journal of Approximate Reasoning*, 53(8):1282–1301, November 2012. ISSN 0888-613X. doi: 10.1016/j.ijar.2012.06.022.

A PROOFS

Lemma 5.3. *A criterion **g** has no **g**-disposable acts if and only if the induced strategy $\tau_{\mathbf{g}}$ is validable.*

Proof. Let $\mathbf{g}$ be a criterion with no $\mathbf{g}$-disposable acts. Then, as ${}^{\mathbf{g}}\succ^{u}_{A}$ is a weak order, for all $A, B \in \mathcal{A}$, we have $A \,{}^{\mathbf{g}}\succeq^{u}_{A} B$. By the definition of $\tau_{\mathbf{g}}$, this implies that $\mathbf{g}^{u}_{A}(A) \geq \mathbf{g}^{u}_{A}(B)$ for all $B \in \mathcal{A}$, and thus $A \in \arg\max_{C \in \mathcal{A}} \mathbf{g}^{u}_{A}(C) = \tau_{\mathbf{g}}(A)$.

Conversely, let $\tau_{\mathbf{g}}$ be a validable strategy. Then, for all $A \in \mathcal{A}$, we have $A \in \tau_{\mathbf{g}}(A) = \arg\max_{C \in \mathcal{A}} \mathbf{g}^{u}_{A}(C)$, which implies that $\mathbf{g}^{u}_{A}(A) \geq \mathbf{g}^{u}_{A}(B)$ for all $B \in \mathcal{A}$. Thus, $A \,{}^{\mathbf{g}}\succeq^{u}_{A} B$ for all $B \in \mathcal{A}$, and so there are no $\mathbf{g}$-disposable acts. □

Theorem 7.2. *Let **g** valid under u. If $\mathcal{U}$ is **g**-aligned with u under $\tau^{*}_{\mathbf{g}}$, then there are no **g**-disposable acts under $\mathcal{U}$.*

Proof. Fix two arbitrary $A^t, B^t \in \mathcal{A}$ and two arbitrary $A^c, B^c \in \mathcal{A}^c$. Let $\mathbf{g}$ be a criterion valid under $u : \mathcal{A} \times \Omega \to [0, 1]$.

From Lemma 5.3, we know that it is impossible that $B^t \,{}^{\mathbf{g}}\succ^{u}_{A^t} A^t$. As ${}^{\mathbf{g}}\succ^{u}_{A^t}$ is a weak order, this is equivalent to stating that we must have

$$A^t \,{}^{\mathbf{g}}\succeq^{u}_{A^t} B^t \tag{7}$$

and

$$B^t \,{}^{\mathbf{g}}\succeq^{u}_{B^t} A^t. \tag{8}$$

We also assume a $\mathcal{U} : 2^{\Omega} \times \Omega \to [0, 1]$ that is $\mathbf{g}$-aligned with u under the strategy $\tau^{*}_{\mathbf{g}}$. This means that

$$B^t \,{}^{\mathbf{g}}\succeq^{u}_{C} \tau^{*}_{\mathbf{g}}(A^c) \,{}^{\mathbf{g}}\succeq^{u}_{C} D^t \;\Rightarrow\; B^t \,{}^{\mathbf{g}}\succeq^{\mathcal{U}}_{C} A^c \,{}^{\mathbf{g}}\succeq^{\mathcal{U}}_{C} D^t, \tag{9}$$

for all $B^t, D^t \in \mathcal{A}$ and $C \in \mathcal{F}$, where the $\tau^{*}_{\mathbf{g}}$ strategy is defined as

$$\tau^{*}_{\mathbf{g}}(F) = \begin{cases} \{A\}, & \text{if } F = A \in \mathcal{A}, \\ \arg\max_{A \in \mathcal{A}} \mathbf{g}^{u}_{F}(A), & \text{if } F \in \mathcal{A}^c, \end{cases} \tag{10}$$

By Equations (9) and (10), it follows

$$\tau^{*}_{\mathbf{g}}(A^c) \,{}^{\mathbf{g}}\succeq^{u}_{A^c} B^t \Rightarrow A^c \,{}^{\mathbf{g}}\succeq^{\mathcal{U}}_{A^c} B^t. \tag{11}$$

By Equation (8) and considering (10), we have $B^t \,{}^{\mathbf{g}}\succeq^{u}_{B^t} \mathcal{A} \supseteq \tau^{*}_{\mathbf{g}}(A^c)$. Then,

$$B^t \,{}^{\mathbf{g}}\succeq^{u}_{B^t} \tau^{*}_{\mathbf{g}}(A^c) \Rightarrow B^t \,{}^{\mathbf{g}}\succeq^{\mathcal{U}}_{B^t} A^c. \tag{12}$$

Similarly, by Equation (10), we have $\tau^{*}_{\mathbf{g}}(A^c) \,{}^{\mathbf{g}}\succeq^{u}_{A^c} \mathcal{A} \supseteq \tau^{*}_{\mathbf{g}}(B^c)$ and, considering Equation (9) and some $\bar{A} \in \tau^{*}_{\mathbf{g}}(A^c)$,

$$\tau^{*}_{\mathbf{g}}(A^c) \,{}^{\mathbf{g}}\succeq^{u}_{A^c} \bar{A} \,{}^{\mathbf{g}}\succeq^{u}_{A^c} \tau^{*}_{\mathbf{g}}(B^c) \Rightarrow A^c \,{}^{\mathbf{g}}\succeq^{\mathcal{U}}_{A^c} \bar{A} \,{}^{\mathbf{g}}\succeq^{\mathcal{U}}_{A^c} B^c. \tag{13}$$

As $A^t, B^t \in \mathcal{A}$ and $A^c, B^c \in \mathcal{A}^c$ were chosen arbitrarily, Equations (7), (11), (12), and (13) imply that, for all $A, B \in \mathcal{F}$,

$$A \,{}^{\mathbf{g}}\succeq^{\mathcal{U}}_{A} B. \tag{14}$$

Thus, there are no $\mathbf{g}$-disposable acts under $\mathcal{U}$.

□

Theorem 8.1. *If **avg** is valid under u, then $\mathcal{U}^{\tau^{*}_{avg}}_{avg}$ is **avg**-aligned.*

Proof. Let $A \in \mathcal{A}^c$, $B, C \in \mathcal{A}$, and $F \in \mathcal{F}$ and suppose $B \,{}^{\textbf{avg}}\!\succeq^u_F \tau^*_{\textbf{avg}}(A) \,{}^{\textbf{avg}}\!\succeq^u_F C$. By definition of ${}^{\textbf{avg}}\!\succeq^u_F$, this means that, for all $\bar{A} \in \tau^*_{\textbf{avg}}(A)$, we have

$$\frac{1}{|F|}\sum_{\omega\in F} u(B,\omega) \geq \frac{1}{|F|}\sum_{\omega\in F} u(\bar{A},\omega) \geq \frac{1}{|F|}\sum_{\omega\in F} u(C,\omega).$$

As this holds for all $\bar{A} \in \tau^*_{\textbf{avg}}(A)$, we can sum over all $\bar{A} \in \tau^*_{\textbf{avg}}(A)$ to obtain

$$\sum_{\bar{A}\in\tau^*_{\textbf{avg}}(A)} \frac{1}{|F|}\sum_{\omega\in F} u(B,\omega) \geq \sum_{\bar{A}\in\tau^*_{\textbf{avg}}(A)} \frac{1}{|F|}\sum_{\omega\in F} u(\bar{A},\omega) \geq \sum_{\bar{A}\in\tau^*_{\textbf{avg}}(A)} \frac{1}{|F|}\sum_{\omega\in F} u(C,\omega).$$

Then

$$\frac{|\tau^*_{\textbf{avg}}(A)|}{|F|}\sum_{\omega\in F} u(B,\omega) \geq \sum_{\bar{A}\in\tau^*_{\textbf{avg}}(A)} \frac{1}{|F|}\sum_{\omega\in F} u(\bar{A},\omega) \geq \frac{|\tau^*_{\textbf{avg}}(A)|}{|F|}\sum_{\omega\in F} u(C,\omega).$$

Dividing by $|\tau^*_{\textbf{avg}}(A)|$,

$$\frac{1}{|F|}\sum_{\omega\in F} u(B,\omega) \geq \frac{1}{|\tau^*_{\textbf{avg}}(A)|}\sum_{\bar{A}\in\tau^*_{\textbf{avg}}(A)} \frac{1}{|F|}\sum_{\omega\in F} u(\bar{A},\omega) \geq \frac{1}{|F|}\sum_{\omega\in F} u(C,\omega).$$

As F and $\tau^*_{\textbf{avg}}(A)$ are finite, we can exchange the order of summation in the middle term to obtain

$$\frac{1}{|F|}\sum_{\omega\in F} u(B,\omega) \geq \frac{1}{|F|}\sum_{\omega\in F} \frac{1}{|\tau^*_{\textbf{avg}}(A)|}\sum_{\bar{A}\in\tau^*_{\textbf{avg}}(A)} u(\bar{A},\omega) \geq \frac{1}{|F|}\sum_{\omega\in F} u(C,\omega).$$

By definition of $\mathcal{U}^{\tau^*_{\textbf{avg}}}_{\text{avg}}$, this is equivalent to

$$\frac{1}{|F|}\sum_{\omega\in F} \mathcal{U}^{\tau^*_{\textbf{avg}}}_{\text{avg}}(B,\omega) \geq \frac{1}{|F|}\sum_{\omega\in F} \mathcal{U}^{\tau^*_{\textbf{avg}}}_{\text{avg}}(A,\omega) \geq \frac{1}{|F|}\sum_{\omega\in F} \mathcal{U}^{\tau^*_{\textbf{avg}}}_{\text{avg}}(C,\omega),$$

which is equivalent to $B \,{}^{\textbf{avg}}\!\succeq^{\mathcal{U}^{\tau^*_{\textbf{avg}}}_{\text{avg}}}_F A \,{}^{\textbf{avg}}\!\succeq^{\mathcal{U}^{\tau^*_{\textbf{avg}}}_{\text{avg}}}_F C$, as desired. □

Theorem 8.2. *If* ***min*** *is valid under u, then $\mathcal{U}^{\tau^*_{min}}_{min}$ is* ***min****-aligned.*

Proof. Let $A \in \mathcal{A}^c$, $B, C \in \mathcal{A}$, and $F \in \mathcal{F}$ and suppose $B \,{}^{\textbf{min}}\!\succeq^u_F \tau^*_{\textbf{min}}(A) \,{}^{\textbf{min}}\!\succeq^u_F C$. By definition of ${}^{\textbf{min}}\!\succeq^u_F$, this means that, for all $\bar{A} \in \tau^*_{\textbf{min}}(A)$, we have

$$\min_{\omega\in F} u(B,\omega) \geq \min_{\omega\in F} u(\bar{A},\omega) \geq \min_{\omega\in F} u(C,\omega).$$

This implies

$$\min_{\omega\in F} u(B,\omega) \geq \min_{\bar{A}\in\tau^*_{\textbf{min}}(A)} \min_{\omega\in F} u(\bar{A},\omega) \geq \min_{\omega\in F} u(C,\omega).$$

And

$$\min_{\omega\in F} u(B,\omega) \geq \min_{\omega\in F} \min_{\bar{A}\in\tau^*_{\textbf{min}}(A)} u(\bar{A},\omega) \geq \min_{\omega\in F} u(C,\omega).$$

By definition of $\mathcal{U}^{\tau^*_{\textbf{min}}}_{\text{min}}$, this is equivalent to

$$\min_{\omega\in F} \mathcal{U}^{\tau^*_{\textbf{min}}}_{\text{min}}(B,\omega) \geq \min_{\omega\in F} \mathcal{U}^{\tau^*_{\textbf{min}}}_{\text{min}}(A,\omega) \geq \min_{\omega\in F} \mathcal{U}^{\tau^*_{\textbf{min}}}_{\text{min}}(C,\omega),$$

which is equivalent to $B \,{}^{\textbf{min}}\!\succeq^{\mathcal{U}^{\tau^*_{\textbf{min}}}_{\text{min}}}_F A \,{}^{\textbf{min}}\!\succeq^{\mathcal{U}^{\tau^*_{\textbf{min}}}_{\text{min}}}_F C$, as desired. □

Theorem 8.3. *If* ***max*** *is valid under u, then $\mathcal{U}^{\tau^*_{max}}_{max}$ is* ***max****-aligned.*

Proof. The proof is analogous to the proof of Theorem 8.2, replacing $\min$ with $\max$. □

Theorem 8.4. *If* ***reg*** *is valid under u, then $\mathcal{U}_{min}^{\tau_{reg}^*}$ is* ***reg****-aligned.*

Proof. Let $A \in \mathcal{A}^c$, $B, C \in \mathcal{A}$, and $F \in \mathcal{F}$ and suppose $B \overset{\textbf{reg}}{\succeq}_F^u \tau_{\textbf{reg}}^*(A) \overset{\textbf{reg}}{\succeq}_F^u C$. By definition of $\overset{\textbf{reg}}{\succeq}_F^u$, this means that, for all $\bar{A} \in \tau_{\textbf{reg}}^*(A)$, we have

$$\min_{\substack{D\in\mathcal{A}\\ \omega\in F}} u(B,\omega) - u(D,\omega) \geq \min_{\substack{D\in\mathcal{A}\\ \omega\in F}} u(\bar{A},\omega) - u(D,\omega) \geq \min_{\substack{D\in\mathcal{A}\\ \omega\in F}} u(C,\omega) - u(D,\omega).$$

One can take the minimum over $\bar{A} \in \tau_{\textbf{reg}}^*(A)$ in the middle term to obtain

$$\min_{\substack{D\in\mathcal{A}\\ \omega\in F}} u(B,\omega) - u(D,\omega) \geq \min_{\bar{A}\in\tau_{\textbf{reg}}^*(A)} \min_{\substack{D\in\mathcal{A}\\ \omega\in F}} u(\bar{A},\omega) - u(D,\omega) \geq \min_{\substack{D\in\mathcal{A}\\ \omega\in F}} u(C,\omega) - u(D,\omega).$$

Changing the order of the minimum in the middle term, we obtain

$$\min_{\substack{D\in\mathcal{A}\\ \omega\in F}} u(B,\omega) - u(D,\omega) \geq \min_{\substack{D\in\mathcal{A}\\ \omega\in F}} \min_{\bar{A}\in\tau_{\textbf{reg}}^*(A)} u(\bar{A},\omega) - u(D,\omega) \geq \min_{\substack{D\in\mathcal{A}\\ \omega\in F}} u(C,\omega) - u(D,\omega).$$

By definition of $\mathcal{U}_{\min}^{\tau_{\textbf{reg}}^*}$, this is equivalent to

$$\min_{\substack{D\in\mathcal{A}\\ \omega\in F}} \mathcal{U}_{\min}^{\tau_{\textbf{reg}}^*}(B,\omega) - u(D,\omega) \geq \min_{\substack{D\in\mathcal{A}\\ \omega\in F}} \mathcal{U}_{\min}^{\tau_{\textbf{reg}}^*}(A,\omega) - u(D,\omega) \geq \min_{\substack{D\in\mathcal{A}\\ \omega\in F}} \mathcal{U}_{\min}^{\tau_{\textbf{reg}}^*}(C,\omega) - u(D,\omega),$$

which is equivalent to $B \overset{\textbf{reg}}{\succeq}_F^{\mathcal{U}_{\min}^{\tau_{\textbf{reg}}^*}} A \overset{\textbf{reg}}{\succeq}_F^{\mathcal{U}_{\min}^{\tau_{\textbf{reg}}^*}} C$, as desired. □

B When not all available terminal acts are a priori epistemic

Consider $\tilde{u} : \tilde{\mathcal{A}} \times \Omega \to [0,1]$ a utility function defined on a set of terminal acts $\tilde{\mathcal{A}}$ that is not necessarily a subset of the epistemic acts $\mathcal{F}$.

For some $\mathcal{A} \subseteq \mathcal{F}$, there is a canonical mapping $\delta : \mathcal{A} \to \tilde{\mathcal{A}}$ and denote by $\mathcal{A}^* = \tilde{\mathcal{A}} \setminus \delta(\mathcal{A})$ the set of terminal acts that are not a priori epistemic.

Our problem here is to find a reward $\tilde{\mathcal{U}} : \tilde{\mathcal{F}} \times \Omega \to [0,1]$ that extends $\tilde{u}$ to all epistemic acts $\tilde{\mathcal{F}} = \mathcal{F} \cup \mathcal{A}^*$, while preserving the DM's preferences.

Define $\tilde{\Omega} = \Omega \cup \mathcal{A}^*$ an expanded state space, and define $u : \tilde{\mathcal{A}} \times \tilde{\Omega} \to [0,1]$ by

$$u(\delta', \omega) = \begin{cases} \tilde{u}(\delta', \omega), & \text{if } \omega \in \Omega, \\ 1, & \text{if } \delta' = \omega \in \mathcal{A}^*, \\ 0, & \text{otherwise.} \end{cases}$$

For all criteria $\mathbf{g}$ such that $\tilde{u}(\delta', \omega) > \tilde{u}(\delta'', \omega)$ for all $\omega \in F$ implies $\mathbf{g}_F^{\tilde{u}}(\delta') > \mathbf{g}_F^{\tilde{u}}(\delta'')$, we have that $\mathbf{g}$ is valid with respect to $\tilde{u}$ if and only if it is valid with respect to u. Also, if $\tilde{u}$ has no disposable acts, u has no disposable acts either.

We can then apply Algorithm 1 to u to obtain a reward $\mathcal{U} : \tilde{\mathcal{F}} \times \tilde{\Omega} \to [0,1]$ that extends u to all epistemic acts $\tilde{\mathcal{F}}$.

Restricting $\mathcal{U}$ to $\tilde{\mathcal{F}} \times \Omega$ yields a reward $\tilde{\mathcal{U}} : \tilde{\mathcal{F}} \times \Omega \to [0,1]$ that extends $\tilde{u}$ to all epistemic acts $\tilde{\mathcal{F}}$ while preserving the DM's preferences.

C Illustrative Experiments

In this section we present additional results from the illustrative experiment of Section 9. The code for all experiments is available at `https://anonymous.4open.science/r/aligned-reward-under-uncertainty_exps-F0D3/`.

C.1 Other Aggregators

C.1.1 OWA Aggregators

In Ma & Denœux (2021), the authors proposed an OWA-based aggregation method.

An OWA aggregator of dimension n is a mapping $F_w : [0,1]^n \to [0,1]$ associated with a collection of weights $w = (w_1, \ldots, w_n)$ such that $w_k \geq 0$ for all k and $\sum_{k=1}^n w_k = 1$, and defined by

$$F_w(x_1, \ldots, x_n) = \sum_{k=1}^{n} w_k x_{(k)},$$

where $x_{(k)}$ denotes the k-th largest element of $x_1, \ldots, x_n$.

In their approach, the weights are chosen by aiming to reflect the DM's attitude towards imprecision. The aggregator is made single-parameter by fixing the imprecision tolerance degree

$$\sum_{k=1}^{n} \frac{n-k}{n-1} w_k = \gamma$$

and choosing, among the weight vectors with this value of $\gamma \in [0.5, 1]$, the one maximizing the entropy $H(w) = -\sum_{k=1}^n w_k \log w_k$.

C.1.2 Cost of Indeterminacy Aggregators

In Yang et al. (2017), the authors proposed two methods to aggregate the utilities of the selected terminal acts.

For a given $p \in [0, \infty]$, they define the following aggregator:

$$C_p(x_1, \ldots, x_n) = 1 - \left(\frac{1}{n} \sum_{k=1}^{n} (1 - x_k)^p \right)^{1/p}$$

for $p \in (0, \infty)$, and the geometric mean for $p = 0$:

$$C_0(x_1, \ldots, x_n) = \left(\prod_{k=1}^{n} x_k \right)^{1/n}.$$

The first method (Cautiousness Seeking Aggregation) they propose is just applying this aggregator to the utilities of the selected terminal acts for some $p \in [0, 1]$.

The second method (Mistake Averse Aggregation) treats "mistakes" differently. For some epistemic act $F = \{\omega_1, \ldots, \omega_n\}$ and a parameter $r \in [0, 1]$, they define the reward as

$$\mathcal{U}(F, \omega) = \begin{cases} C_{1-r}(u(\omega_1, \omega), \ldots, u(\omega_n, \omega)), & \text{if } \omega \in F, \\ C_{1+r}(u(\omega_1, \omega), \ldots, u(\omega_n, \omega)), & \text{if } \omega \notin F. \end{cases}$$

C.1.3 Experimental Results

Figure 3 reports the ratio of disposable epistemic acts under both the OWA and Cost of Indeterminacy aggregators, for different values of their parameters. The results are qualitatively similar to those of Figure 2, confirming that the naive approach leads to a substantial proportion of disposable epistemic acts under every aggregator and criterion, and that this ratio increases with $|\Omega|$.

C.2 Misuse of Algorithm 1

Algorithm 1 requires the DM to identify their actual criterion, which may be impractical. If the identification is wrong, the resulting reward may not be aligned and disposable acts may reappear. We repeated the experiment while deliberately misattributing the criterion, and report the resulting ratio of disposable epistemic acts in Figure 4. When the criterion is identified correctly, the ratio is

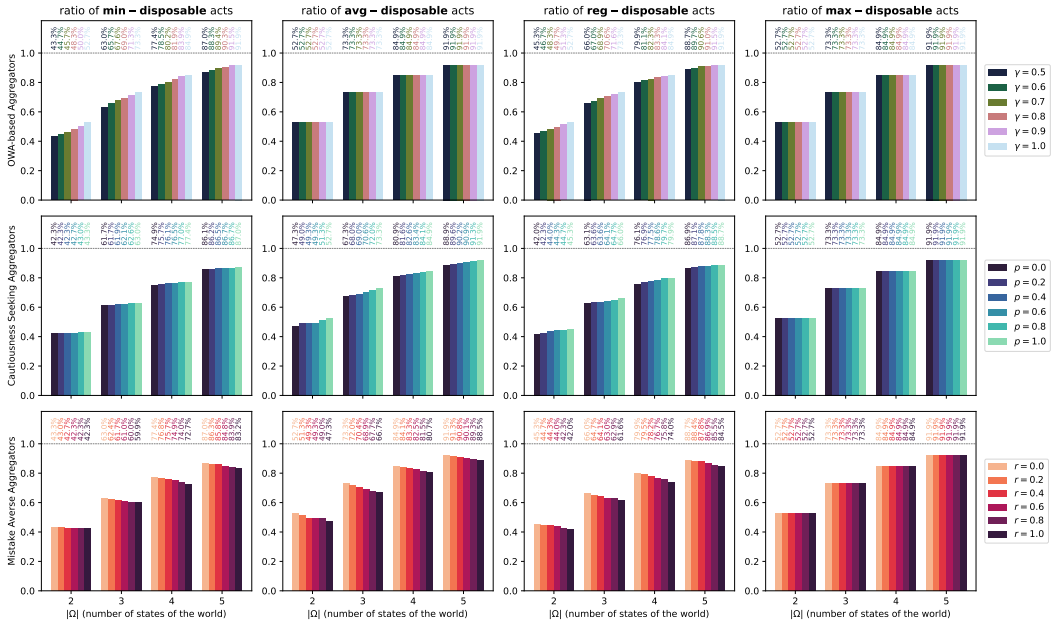


Figure 3: Ratio of disposable epistemic acts when naively applying the OWA and Cost of Indeterminacy aggregators as a function of the number of states of the world. First row: OWA aggregator with different values of the imprecision tolerance degree γ. Second row: Cautiousness Seeking Aggregation with different values of p. Third row: Mistake Averse Aggregation with different values of r. Each bar reports the average fraction of epistemic acts that are **g**-disposable under the corresponding decider criterion, over 100 randomly generated utility functions. Columns correspond to the maximin, Laplace, minimax-regret, and maximax decider criteria, discussed in Section 6. In higher dimensions, the naive approach leads to most epistemic acts being disposable.

zero, as expected. Under a misidentification it remains well below the naive approach, but no longer vanishes, reaching around 40% *for* $|\Omega| = 5$ *in the worst cases. The figure also shows that not all misidentifications are equally harmful: confusing maximin with maximax (and vice versa) yields the highest ratios, whereas Laplace and minimax-regret seem to behave more similarly to each other, so mistaking one for the other costs less. When the DM is uncertain about their criterion, these latter two seem therefore to be the safer choices.*

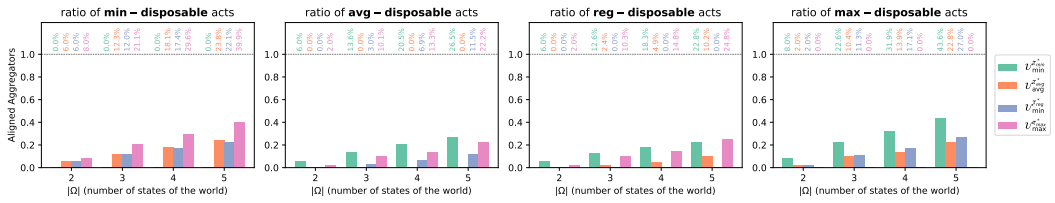


Figure 4: Ratio of disposable epistemic acts under a misuse of Algorithm 1, in which the DM does not identify their actual criterion. Each bar reports the average fraction of epistemic acts that are disposable under the corresponding decider criterion, over 100 randomly generated utility functions. Columns correspond to the maximin, Laplace, minimax-regret, and maximax decider criteria, discussed in Section 6.