Decision-making with E-admissibility given a finite assessment of choices

Arne Decadt, Alexander Erreygers, Jasper De Bock, and Gert de Cooman

Abstract Given information about which options a decision-maker definitely rejects from given finite sets of options, we study the implications for decision-making with E-admissibility. This means that from any finite set of options, we reject those options that no probability mass function compatible with the given information gives the highest expected utility. We use the mathematical framework of choice functions to specify choices and rejections, and specify the available information in the form of conditions on such functions. We characterise the most conservative extension of the given information to a choice function that makes choices based on E-admissibility, and provide an algorithm that computes this extension by solving linear feasibility problems.

1 Introduction

A decision-maker's uncertainty is typically modelled by a probability measure, and it is often argued that her rational decisions maximise expected utility with respect to this probability measure. However, she may not always have sufficient knowledge to come up with a unique and completely specified probability measure. It is then often assumed, as a work-around, that there is some set of probability measures that describes her uncertainty. In this setting, *E-admissibility* is among the more popular criteria for making choices, as indicated by Troffaes (2007).

In this paper, we study and propose an algorithm for decision-making based on this criterion, starting from a finite uncertainty assessment. As E-admissibility is popular, we are not the first to try and deal with this. Utkin and Augustin (2005) and Kikuti et al. (2005) have gone before us, but their assessments essentially only deal with pairwise comparison of options, while

Arne Decadt, Alexander Erreygers, Jasper De Bock and Gert de Cooman
Foundations Lab, Ghent University, e-mail: arne.decadt@ugent.be

we can handle comparisons between sets of options. Decadt et al. (2020) have also studied more general assessments, but for other decision criteria than E-admissibility. In order to achieve this generality, we will use choice functions as tools to model the decision-making process, because they lead to a very general framework, as argued elsewhere by, for instance, Seidenfeld et al. (2010); De Bock and De Cooman (2019); De Bock (2020).

2 Setting & choice functions

A choice function is a function that, for any given set of options, selects some subset of them. The set $\mathcal{V}$ collects all *options* and $\mathcal{Q}$ is the set of all non-empty finite subsets of $\mathcal{V}$. Formally, a *choice function C* is then a map from $\mathcal{Q}$ to itself such that $C(A) \subseteq A$ for all $A \in \mathcal{Q}$. If $C(A)$ is a singleton consisting of a single option u, this means that u is chosen from A. If $C(A)$ has more than one element, however, we don't take this to mean that all the options in $C(A)$ are chosen, but rather that the options in $A \setminus C(A)$ are rejected and that the model does not contain sufficient information to warrant making a choice between the remaining options in $C(A)$. Depending on the desired behaviour, various axioms can be imposed, leading to different types of choice functions; see for example Seidenfeld et al. (2010); De Bock and De Cooman (2019); De Bock (2020). In this contribution we consider choice functions under E-admissibility, as introduced in Section 3.

We furthermore assume that we have an uncertain experiment with n possible outcomes, and we order the set of all outcomes $\mathcal{X}$ as $\{x_1, \ldots, x_n\}$. We interpret an option u as a function that maps each outcome x in $\mathcal{X}$ to the real-valued utility $u(x)$ that we get when the outcome of the uncertain experiment turns out to be x. So we take the set of all options $\mathcal{V}$ to be the real vector space of all real-valued maps on $\mathcal{X}$.

3 E-admissibility

A decision-maker's uncertainty about an experiment is typically modelled by means of a probability mass function $p \colon \mathcal{X} \to [0, 1]$, which represents the probability of each outcome in $\mathcal{X}$; we will use Σ to denote the set of all such probability mass functions on $\mathcal{X}$. The standard way—see for example (Savage, 1972, Chapter 5)—to choose between options u proceeds by maximising expected utility with respect to p, where the expected utility of an option $u \in \mathcal{V}$ is given by $\mathrm{E}_p(u) \coloneqq \sum_{x \in \mathcal{X}} u(x)p(x)$.

For every probability mass function $p \in \Sigma$, the resulting choice function C_p that maximises expected utility is defined by

$$C_p(A) := \{u \in A \colon (\forall a \in A)\mathrm{E}_p(u) \geq \mathrm{E}_p(a)\} \text{ for all } A \in Q. \qquad (1)$$

It is, however, not always possible to pin down exact probabilities for the outcomes (Walley, 1991, Chapter 1). Yet, the decision-maker might have some knowledge about these probabilities, for example in terms of bounds on the probabilities of some events. Such knowledge gives rise to a set of probability mass functions $\mathcal{P} \subseteq \Sigma$, called a *credal set* (Levi, 1978, Section 1.6.2). In this context, there need no longer be a unique expected utility and so the decision-maker cannot simply maximise it. Several other decision criteria can then be used instead; Troffaes (2007) gives an overview. One criterion that is often favoured is *E-admissibility*: choose those options that maximise expected utility with respect to at least one of the probability mass functions p in $\mathcal{P}$ (Levi, 1978).[1] If $\mathcal{P}$ is non-empty, the corresponding choice function $C_{\mathcal{P}}^{\mathrm{E}}$ is defined by

$$C_{\mathcal{P}}^{\mathrm{E}}(A) := \bigcup_{p \in \mathcal{P}} C_p(A) \text{ for all } A \in Q. \qquad (2)$$

It will prove useful to extend this definition to the case that $\mathcal{P} = \emptyset$. Eq. (2) then yields that $C_{\emptyset}^{\mathrm{E}}(A) = \emptyset$ for all $A \in Q$, so $C_{\emptyset}^{\mathrm{E}}$ is no longer a choice function. In either case, it follows immediately from Eqs. (1) and (2) that

$$C_{\mathcal{P}}^{\mathrm{E}}(A) = \{u \in A \colon (\exists p \in \mathcal{P})(\forall a \in A)\mathrm{E}_p(u) \geq \mathrm{E}_p(a)\} \text{ for all } A \in Q. \qquad (3)$$

The behaviour of choice functions under E-admissibility was first studied for horse lotteries by Seidenfeld et al. (2010), characterised in a very general context by De Cooman (2021), and captured in axioms by De Bock (2020).

4 Assessments and extensions

We assume that there is some choice function C that represents the decision-maker's preferences, but we may not fully know this function. Our partial information about C comes in the form of preferences regarding some—so not necessarily all—option sets. More exactly, for some option sets $A \in Q$, we know that the decision-maker rejects all options in $W \subseteq A$, meaning that $C(A) \subseteq A \setminus W$; this can be also be stated as $C(V \cup W) \subseteq V$, with $V := A \setminus W$. We will represent such information by an *assessment*: a set $\mathcal{A} \subseteq Q^2$ of pairs (V, W) of disjoint option sets with the interpretation that, for all $(V, W) \in \mathcal{A}$, the options in W are definitely rejected from $V \cup W$.

Given such an assessment, it is natural to ask whether there is some choice function $C_{\mathcal{P}}^{\mathrm{E}}$ under E-admissibility that agrees with it, in the sense that $C_{\mathcal{P}}^{\mathrm{E}}(V \cup W) \subseteq V$ for all $(V, W) \in \mathcal{A}$. Whenever this is the case, we call the

---

[1] Levi's original definition considered credal sets that are convex, whereas we do not require this. In fact one of the strengths of our approach is that an assessment can lead to non-convex credal sets; see the example in Section 5 further on.

assessment $\mathcal{A}$ *consistent* with E-admissibility. It follows from Eq. (2) that $C_{\mathcal{P}}^{\mathrm{E}}$ agrees with the assessment $\mathcal{A}$ if and only if

$$\mathcal{P} \subseteq \mathcal{P}(\mathcal{A}) := \left\{ p \in \Sigma \colon (\forall (V, W) \in \mathcal{A}) C_p(V \cup W) \subseteq V \right\}.$$

Hence, $\mathcal{A}$ is consistent if and only if $\mathcal{P}(\mathcal{A}) \neq \emptyset$. To check if $\mathcal{A}$ is consistent, the following alternative characterisation will also be useful: for any $A \in Q$,

$$\mathcal{P}(\mathcal{A}) \neq \emptyset \Leftrightarrow C_{\mathcal{P}(\mathcal{A})}^{\mathrm{E}}(A) \neq \emptyset. \tag{4}$$

If an assessment $\mathcal{A}$ is consistent and there is more than one choice function that agrees with it, the question remains which one we should use. A careful decision-maker would only want to reject options if this is implied by the assessment. So she wants a most conservative agreeing choice function under E-admissibility, one that rejects the fewest number of options. Since larger credal sets lead to more conservative choice functions, this most conservative agreeing choice function under E-admissibility clearly exists, and is equal to $C_{\mathcal{P}(\mathcal{A})}^{\mathrm{E}}$. For this reason, we call $C_{\mathcal{P}(\mathcal{A})}^{\mathrm{E}}$ the *E-admissible extension* of the assessment $\mathcal{A}$.

So we conclude that checking the consistency of an assessment $\mathcal{A}$, as well as finding the E-admissible extension of a consistent assessment $\mathcal{A}$, amounts to evaluating $C_{\mathcal{P}(\mathcal{A})}^{\mathrm{E}}$. In the following sections we provide a method for doing this, which makes use of the following more practical expression for $\mathcal{P}(\mathcal{A})$.

Proposition 1 *Consider an assessment $\mathcal{A}$. Then*

$$\mathcal{P}(\mathcal{A}) = \left\{ p \in \Sigma \colon (\forall (V, W) \in \mathcal{A})(\forall w \in W)(\exists v \in V) \mathrm{E}_p(v) > \mathrm{E}_p(w) \right\}.$$

5 A characterisation of the E-admissible extension

Having defined the E-admissible extension $C_{\mathcal{P}(\mathcal{A})}^{\mathrm{E}}$ of an assessment $\mathcal{A}$, it is only natural to wonder whether we can easily compute it. We now turn to a method for doing so, albeit only for finite assessments. As a first step, we derive a convenient characterisation of $C_{\mathcal{P}(\mathcal{A})}^{\mathrm{E}}$. For any positive integer m, it uses the notations $[1{:}m] := \{1, \ldots, m\}$ and $d_{1:m} := (d_1, \ldots, d_m)$.

Theorem 1 *Consider an option set A, an option $u \in A$ and a non-empty, finite assessment $\mathcal{A}$. Enumerate the set $\{\{v - w \colon v \in V\} \colon (V, W) \in \mathcal{A}, w \in W\}$ as $\{D_1, \ldots, D_m\}$ and the set $\{u - a \colon a \in A \setminus \{u\}\}$ as $\{u_1, \ldots, u_\ell\}$. Then $\mathcal{P}(\mathcal{A}) = \bigcup_{d_{1:m} \in \times_{j=1}^m D_j} \mathcal{P}(d_{1:m})$, where, for each $d_{1:m} \in \times_{j=1}^m D_j$, we let*

$$\mathcal{P}(d_{1:m}) := \left\{ p \in \Sigma \colon (\forall j \in [1{:}m]) \mathrm{E}_p(d_j) > 0 \right\}. \tag{5}$$

Furthermore, $u \in C^{\mathrm{E}}_{\mathcal{P}(\mathcal{A})}(A)$ if and only if there is some $p \in \mathcal{P}(\mathcal{A})$—or equivalently some $d_{1:m} \in \times^{m}_{j=1} D_j$ and $p \in \mathcal{P}(d_{1:m})$—such that $\mathrm{E}_p(u_i) \geq 0$ for all $i \in [1:\ell]$.

Let us illustrate the use of Theorem 1 in determining, for a given option set, the resulting choices under the conservative E-admissible extension of a given assessment. Let $\mathcal{X} := \{1, 2, 3\}$. In order to allow for a graphical representation, we identify options and probability mass functions with vectors in $\mathbb{R}^3$, where for any $x \in \mathcal{X}$, the x-th component corresponds to the value of the option or probability mass function in x; so for example the option $w_1 := (1, -3, 1)$ corresponds to the option that maps 1 to 1, 2 to -3 and 3 to 1. We will choose from the option set $A := \{w_1, w_2, w_3\}$, where we also let $w_2 := (1, 1, -2)$ and $w_3 := (0, 0, 0)$.

For the assessment, we will consider $v_1 := (-1, 2, -2)$, $v_2 := (-2, 2, -1)$, $v_3 := (0, 3, -11)$, $v_4 := (0, -7, -1)$, $v_5 := (2, 5, -9)$ and $v_6 := (0, -2, -1)$. Suppose that we are given the information that v_2, v_3 and v_4 are rejected from $\{v_1, v_2, v_3, v_4\}$ and that v_1 is rejected from $\{v_1, v_5, v_6\}$. This corresponds to the assessment $\mathcal{A} = \{(\{v_1\}, \{v_2, v_3, v_4\}), (\{v_5, v_6\}, \{v_1\})\}$.

Now we will check for every option in A whether it is in $C^{\mathrm{E}}_{\mathcal{P}(\mathcal{A})}(A)$, by applying Theorem 1. For the sake of efficiency, we note that for all options, the assessment $\mathcal{A}$ is the same, so they all have

$$
\begin{aligned}
&\{D_1, \ldots, D_4\} \\
&\quad = \big\{\{v_1 - v_2\}, \{v_1 - v_3\}, \{v_1 - v_4\}, \{v_5 - v_1, v_6 - v_1\}\big\} \\
&\quad = \big\{\{(1, 0, -1)\}, \{(-1, -1, 9)\}, \{(-1, 9, -1)\}, \{(3, 3, -7), (1, -4, 1)\}\big\}.
\end{aligned}
$$

In Fig. 1, we have drawn the credal set $\mathcal{P}(\mathcal{A})$ in blue in a ternary plot, using the characterisation in Theorem 1.

For w_1, the probability mass function $p_1 := (^{12}/_{20}, ^{3}/_{20}, ^{5}/_{20})$ is consistent with the assessment, and we have $\mathrm{E}_{p_1}(u_1) = ^{3}/_{20} \geq 0$ and $\mathrm{E}_{p_1}(u_2) = ^{2}/_{5} \geq 0$, with $u_1 := w_1 - w_2 = (0, -4, 3)$ and $u_2 := w_1 - w_3 = (1, -3, 1)$. Therefore, it follows from Theorem 1 that w_1 is not rejected from A by $C^{\mathrm{E}}_{\mathcal{P}(\mathcal{A})}$. That w_2 is not rejected either can be inferred similarly, for example using $p_2 := (^{3}/_{5}, ^{1}/_{5}, ^{1}/_{5})$. For w_3, we have $u_1 := w_3 - w_1 = (-1, 3, -1)$ and $u_2 := w_3 - w_2 = (-1, -1, 2)$. The set of probability mass functions for which $\mathrm{E}_p(u_1) \geq 0$ and $\mathrm{E}_p(u_2) \geq 0$ corresponds to the green region in Fig. 1, which has no overlap with the blue region. Therefore, w_3 is rejected from A by $C^{\mathrm{E}}_{\mathcal{P}(\mathcal{A})}$. So we conclude that $C^{\mathrm{E}}_{\mathcal{P}(\mathcal{A})}(A) = \{w_1, w_3\}$.

6 An algorithmic approach

For larger problems, when the graphical approach in the example above is no longer feasible, we can translate Theorem 1 into an algorithm. A first way is
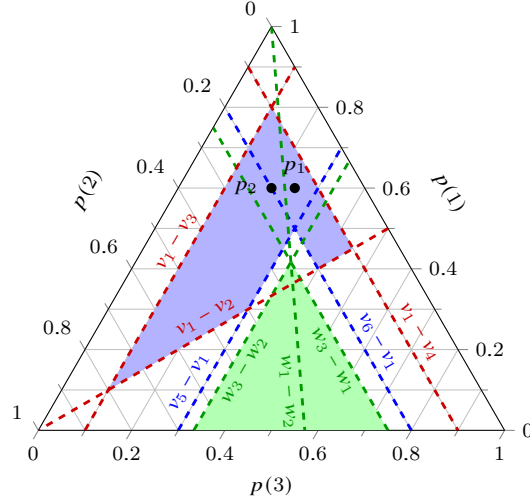
Fig. 1 Ternary plot where the credal set $\mathcal{P}(\mathcal{A})$ consists of those probability mass functions $p \colon \{1, 2, 3\} \to [0, 1]$ that correspond to the blue region. A line labelled with an option v means that $E_p(v) = 0$ for all p on the line. The green region corresponds to the probability mass functions p for which $E_p(w_3 - w_1) \geq 0$ and $E_p(w_3 - w_2) \geq 0$.

to directly search, for each $d_{1:m}$, for a probability mass function p in $\mathcal{P}(d_{1:m})$ such that $u \in C_p(A)$. To this end, we consider the function PRIMAL $\colon \mathcal{V}^\ell \times \mathcal{V}^m \to \{\text{True}, \text{False}\}$ that returns True if the following feasibility problem has a solution for input $(u_{1:\ell}, d_{1:m}) \in \mathcal{V}^\ell \times \mathcal{V}^m$ and False otherwise, where λ_k can be seen as a scaled version of $p(x_k)$:

$$
\begin{aligned}
\text{find} \quad & \lambda_1, \ldots, \lambda_n \in \mathbb{R}, \\
\text{subject to} \quad & \textstyle\sum_{k=1}^n \lambda_k u_i(x_k) \geq 0 && \text{for all } i \in [1{:}\ell], \\
& \textstyle\sum_{k=1}^n \lambda_k d_j(x_k) \geq 1 && \text{for all } j \in [1{:}m], \\
& \lambda_k \geq 0 && \text{for all } k \in [1{:}n].
\end{aligned}
$$

For another way to translate the condition in Theorem 1 into an algorithm, we use duality. That is, we consider the function DUAL $\colon \mathcal{V}^\ell \times \mathcal{V}^m \to \{\text{True}, \text{False}\}$ that returns True if the following feasibility problem has a solution for input $(u_{1:\ell}, d_{1:m}) \in \mathcal{V}^\ell \times \mathcal{V}^m$ and False otherwise:

$$
\begin{aligned}
\text{find} \quad & \lambda_1, \ldots, \lambda_{\ell+m} \in \mathbb{R}, \\
\text{subject to} \quad & \textstyle\sum_{i=1}^\ell \lambda_i u_i(x_k) + \sum_{j=1}^m \lambda_{\ell+j} d_j(x_k) \leq 0 && \text{for all } k \in [1{:}n], \\
& \textstyle\sum_{j=\ell+1}^{\ell+m} \lambda_j \geq 1, \\
& \lambda_i \geq 0 && \text{for all } i \in [1{:}\ell + m].
\end{aligned}
$$

In practice, either of these feasibility problems can be solved by linear programming.[2] Our next result relates these feasibility problems to the conditions in Theorem 1.

Theorem 2 *Consider option sequences $u_{1:\ell} \in \mathcal{V}^\ell$ and $d_{1:m} \in \mathcal{V}^m$ and let $\mathcal{P}(d_{1:m})$ be as defined in Eq. (5). Then the following statements are equivalent:*

(i) *There is some $p \in \mathcal{P}(d_{1:m})$ such that $\mathrm{E}_p(u_i) \geq 0$ for all $i \in [1:\ell]$.*
(ii) $\mathrm{PRIMAL}(u_{1:\ell}, d_{1:m}) = \mathrm{True}$.
(iii) $\mathrm{DUAL}(u_{1:\ell}, d_{1:m}) = \mathrm{False}$.

Theorems 1 and 2 guarantee that we can determine $C_{\mathcal{P}(\mathcal{A})}^{\mathrm{E}}(A)$ for any set $A \in Q$ using Algorithm 1, by checking for each option $u \in A$ whether $u \in C_{\mathcal{P}(\mathcal{A})}^{\mathrm{E}}(A)$. For any single u, this amounts to solving a linear feasibility program for each $d_{1:m}$, using PRIMAL or DUAL, as preferred. Interestingly, consistency is equivalent to $C_{\mathcal{P}(\mathcal{A})}^{\mathrm{E}}(A) \neq \emptyset$, by Eq. (4). In practice, consistency can also be easily verified beforehand, by checking if $0 \in C_{\mathcal{P}(\mathcal{A})}^{\mathrm{E}}(\{0\})$, where '0' is the constant option that is 0 everywhere.

Algorithm 1 Check for an option set $A \in Q$ and a finite assessment $\mathcal{A}$ if an option $u \in A$ is in $C_{\mathcal{P}(\mathcal{A})}^{\mathrm{E}}(A)$.

Precondition: Let $\{D_1, \ldots, D_m\}$ and $\{u_1, \ldots, u_\ell\}$ be as in Theorem 1.

1: **for all** $d_{1:m} \in \times_{j=1}^m D_j$ **do**
2:   **if** $\mathrm{PRIMAL}(u_{1:\ell}, d_{1:m})$ **then**   ▷ Or $\neg\mathrm{DUAL}(u_{1:\ell}, d_{1:m})$.
3:     **return** True   ▷ For one of the $d_{1:m}$ the condition is fulfilled.
4: **return** False   ▷ When all elements of $\times_{j=1}^m D_j$ have been checked.

If the assessment $\mathcal{A}$ consists solely of pairs (V, W) where V is a singleton, then the corresponding set $\times_{j=1}^m D_j$ is a singleton, and the for-loop in Algorithm 1 vanishes. Our algorithm can therefore be seen as repeatedly solving problems that have assessments of that form.

7 Conclusion

We have shown how to make choices using the E-admissibility criterion, starting from a finite assessment, using choice functions. Our main conclusion is that calculating the most conservative E-admissible extension of this assessment reduces to checking linear feasibility multiple times. Our setup is similar

---

[2] It can for instance be considered as a linear programming problem, by adding the trivial objective function that is zero everywhere. Feeding this into a linear programming software package, the software will announce whether the problem is feasible. For a deeper understanding of how software solves such feasibility problems, we refer to the explanation of initial feasible solutions in Matoušek and Gärtner (2006, Section 5.6).

to the one previously studied by Utkin and Augustin (2005); Kikuti et al. (2005), the essential difference being that they have pairwise comparisons in the form of non-strict inequalities on expected utilities, whereas our assessments consider comparisons between sets of options, which leads to strict inequalities and allows for non-convex credal sets. Future work could look into also implementing assessments that incorporate non-strict inequalities. One way to do so would be through infinite assessments, so it might pay to look at which types of infinite assessments can still be handled finitely.

References

Birkhoff G (1940) Lattice theory, 3rd edn. American Mathematical Society

De Bock J (2020) Archimedean choice functions: an axiomatic foundation for imprecise decision making. In: Proceedings of IPMU 2020, Springer, pp 195–209

De Bock J, De Cooman G (2019) Interpreting, axiomatising and representing coherent choice functions in terms of desirability. In: Proceedings of ISIPTA 2019, PMLR, pp 125–134

De Cooman G (2021) Coherent and archimedean choice in general Banach spaces. International Journal of Approximate Reasoning 140:255–281

Decadt A, De Bock J, De Cooman G (2020) Inference with choice functions made practical. In: Proceedings of SUM 2020, Springer, pp 113–127

Kikuti D, Cozman FG, de Campos CP (2005) Partially ordered preferences in decision trees: computing strategies with imprecision in probabilities. In: IJCAI workshop on advances in preference handling, pp 118–123

Levi I (1978) On indeterminate probabilities. In: Foundations and Applications of Decision Theory, Springer, pp 233–261

Matoušek J, Gärtner B (2006) Understanding and Using Linear Programming. Universitext, Springer

Savage LJ (1972) The Foundations of Statistics. Courier Corporation

Seidenfeld T, Schervish MJ, Kadane JB (2010) Coherent choice functions under uncertainty. Synthese 172:157–176

Troffaes MC (2007) Decision making under uncertainty using imprecise probabilities. International Journal of Approximate Reasoning 45(1):17–29

Utkin L, Augustin T (2005) Powerful algorithms for decision making under partial prior information and general ambiguity attitudes. In: Proceedings of ISIPTA 2005, pp 349–358

Walley P (1991) Statistical Reasoning with Imprecise Probabilities. Chapman and Hall

Appendix

Proof of Eq. (4) The reverse implication follows from Eq. (3). To prove the direct implication, we assume that $\mathcal{P}(\mathcal{A}) \neq \emptyset$ and fix any $p \in \mathcal{P}(\mathcal{A})$. Assume *ex absurdo* that $C_p(A)$ is empty. Then for every $u \in A$ there is an $a \in A$ such that $\mathrm{E}_p(a) > \mathrm{E}_p(u)$. But this would contradict (Birkhoff, 1940, I.3 Theorem 3) that says that there is a maximal element in A with respect to the partial order $\succeq_p := \{(x, y) \in \mathcal{V}^2 : \mathrm{E}_p(x) > \mathrm{E}_p(y) \vee x = y\}$. Hence, $C_p^{\mathrm{E}}(A) \neq \emptyset$. Because $p \in \mathcal{P}(\mathcal{A})$, it follows from Eq. (2) that $C_{\mathcal{P}(\mathcal{A})}(A) \neq \emptyset$, as required.

Proof of Proposition 1 By the definition of $\mathcal{P}(\mathcal{A})$, we have to prove that for all $p \in \Sigma$ and $(V, W) \in \mathcal{A}$ the following statements are equivalent:

$$C_p(V \cup W) \subseteq V \tag{6}$$

and

$$(\forall w \in W)(\exists v \in V)\mathrm{E}_p(v) > \mathrm{E}_p(w). \tag{7}$$

Take any $p \in \Sigma$ and $(V, W) \in \mathcal{A}$. First we prove that Eq. (6) implies Eq. (7). From Eq. (6) and the fact that V and W are disjoint, it follows that $w \notin C_p(V \cup W)$ for all $w \in W$. This means by definition that

$$(\forall w \in W)(\exists a \in V \cup W)\mathrm{E}_p(a) > \mathrm{E}_p(w). \tag{8}$$

We will now show that this implies Eq. (7). Take any option $w \in W$. Let $R := \{r \in W : r \succeq_p w\}$, where we define the partial order $\succeq_p := \{(x, y) \in \mathcal{V}^2 : \mathrm{E}_p(x) > \mathrm{E}_p(y) \vee x = y\}$. Then by (Birkhoff, 1940, I.3 Theorem 3), there is some maximal option $w^* \in R$ with respect to $\succeq_p$, since R is non-empty because it contains w and finite as it is a subset of W. Since $w^* \in R \subseteq W$, we know from Eq. (8) that there is some some $a^* \in V \cup W$ such that $\mathrm{E}_p(a^*) > \mathrm{E}_p(w^*)$, and therefore also $a^* \succeq_p w^*$ and $a^* \neq w^*$. Since w^* is maximal in R with respect to $\succeq_p$, this implies that it is impossible that $a^* \in R$. It is also impossible that $a^* \in W \setminus R$ because $a^* \succeq_p w^* \succeq_p w$, where the second preference holds because $w^* \in R$. Hence, it must be that $a^* \in V$. Since W and V are disjoint, and $w \in W$, this implies that $a^* \neq w$. Since $a^* \succeq_p w$, it follows that $\mathrm{E}_p(a^*) > \mathrm{E}_p(w)$. So we have found some a^* in V such that $\mathrm{E}_p(a^*) > \mathrm{E}_p(w)$. As this holds for any option $w \in W$, we have proved Eq. (7).

Next we prove that Eq. (7) implies Eq. (6). Take any option $w \in W$. Since $V \subseteq V \cup W$, we have from Eq. (7) and the definition of C_p that $w \notin C_p(V \cup W)$. Since this holds for any $w \in W$, it follows that $C_p(V \cup W) \subseteq V$, and this is Eq. (6).

Proof of Theorem 1 Let $\mathcal{D} := \times_{j=1}^{m} D_j$. First we prove that

$$\mathcal{P}(\mathcal{A}) = \bigcup_{d_{1:m} \in \mathcal{D}} \mathcal{P}(d_{1:m}). \tag{9}$$

By Proposition 1 and Eq. (5), this is equivalent to proving that for any $p \in \Sigma$ the following statements are equivalent

$$(\forall(V, W) \in \mathcal{A})(\forall w \in W)(\exists v \in V)\mathrm{E}_p(v) > \mathrm{E}_p(w) \tag{10}$$

and

$$(\exists d_{1:m} \in \mathcal{D})(\forall j \in [1:m])\mathrm{E}_p(d_j) > 0. \tag{11}$$

First we prove that Eq. (10) implies Eq. (11). Take any $j \in [1:m]$. By definition of $\{D_1, \ldots, D_m\}$, there is a $(V, W) \in \mathcal{A}$ and a $w \in W$ such that $D_j = \{v - w \colon v \in V\}$. By Eq. (10) there is a some $v^* \in V$ such that $\mathrm{E}_p(v^*) > \mathrm{E}_p(w)$. Let $d_j := v^* - w$. Then $d_j \in D_j$ by definition. From linearity of the expectation operator E_p and $\mathrm{E}_p(v^*) > \mathrm{E}_p(w)$, it follows that $\mathrm{E}_p(d_j) = \mathrm{E}_p(v^* - w) > 0$. Since this holds for all $j \in [1:m]$, Eq. (11) holds.

Next we prove that Eq. (11) implies Eq. (10). Take some $d_{1:m} \in \mathcal{D}$ that satisfies Eq. (11). Take any $(V, W) \in \mathcal{A}$ and $w \in W$. Then by definition of $\{D_1, \ldots, D_m\}$, there is some $j \in [1:m]$ such that $D_j = \{v - w \colon v \in V\}$ and thus, since $d_j \in D_j$, also some $v^* \in V$ such that $d_j = v^* - w$. Then by Eq. (11) we have $\mathrm{E}_p(v^* - w) = \mathrm{E}_p(d_j) > 0$ and from this and the linearity of the expectation operator E_p, it follows that $\mathrm{E}_p(v^*) > \mathrm{E}_p(w)$. Since we can find such a v^* for any $(V, W) \in \mathcal{A}$ and $w \in W$, we have proven Eq. (10).

For the second part of the statement, we rewrite $u \in C^{\mathrm{E}}_{\mathcal{P}(\mathcal{A})}(A)$. By Eq. (3), this is equivalent to the statement that there is some $p \in \mathcal{P}(\mathcal{A})$ such that

$$(\forall a \in A)\mathrm{E}_p(u) \geq \mathrm{E}_p(a). \tag{12}$$

It therefore suffices to prove, for any $p \in \Sigma$, that Eq. (12) is equivalent to

$$(\forall i \in [1:\ell])\mathrm{E}_p(u_i) \geq 0, \tag{13}$$

which is what we now set out to do.

First we prove that Eq. (12) implies Eq. (13). Take any $i \in [1:\ell]$, then by definition of $\{u_1, \ldots, u_\ell\}$ there is some $a \in A$ such that $u_i = u - a$. By Eq. (12), $\mathrm{E}_p(u) \geq \mathrm{E}_p(a)$ and by linearity of the expectation operator E_p, we have $\mathrm{E}_p(u_i) = \mathrm{E}_p(u - a) \geq 0$.

Next we prove that Eq. (13) implies Eq. (12). Take any $a \in A$. If $a = u$, we trivially have that $\mathrm{E}_p(u) \geq \mathrm{E}_p(a)$. If $a \neq u$, then by definition of $\{u_1, \ldots, u_\ell\}$ there is some $i \in [1:\ell]$ such that $u_i = u - a$, or $u = u_i + a$. By Eq. (13), $\mathrm{E}_p(u_i) \geq 0$ and by linearity of expectation we have $\mathrm{E}_p(u) = \mathrm{E}_p(u_i) + \mathrm{E}_p(a) \geq \mathrm{E}_p(a)$.

Proof of Theorem 2 First we prove that (i) implies (ii). By (i), there is a $p \in \mathcal{P}(d_{1:m})$ such that $\mathrm{E}_p(u_i) \geq 0$ for all $i \in [1:\ell]$. By definition of $\mathcal{P}(d_{1:m})$, there is a $p \in \Sigma$ such that $\mathrm{E}_p(d_j) > 0$ for all $j \in [1:m]$ and $\mathrm{E}_p(u_i) \geq 0$ for all $i \in [1:\ell]$. Then $\eta := \min_{j \in [1:m]} \mathrm{E}_p(d_j) > 0$. For all $k \in [1:n]$, let $\lambda_k := p(x_k)/\eta \geq 0$. Then $\sum_{k=1}^n \lambda_k u_i(x_k) = \mathrm{E}_p(u_i)/\eta \geq 0$ for all $i \in [1:\ell]$, because $\mathrm{E}_p(u_i) \geq 0$ and $\eta > 0$, and $\sum_{k=1}^n \lambda_k d_j(x_k) = \mathrm{E}_p(d_j)/\eta \geq 1$ for all $j \in [1:m]$

because $E_p(d_j) \geq \eta > 0$. In other words, the real numbers $\lambda_1, \ldots, \lambda_n$ satisfy the primal linear feasibility problem, so $\text{PRIMAL}(u_{1:\ell}, d_{1:m}) = \text{True}$.

Second we prove that (ii) implies (i). By (ii), there are real numbers $\lambda_1, \ldots, \lambda_n$ such that $\lambda_k \geq 0$ for all $k \in [1\!:\!n]$, $\sum_{k=1}^n \lambda_k u_i(x_k) \geq 0$ for all $i \in [1\!:\!\ell]$ and $\sum_{k=1}^n \lambda_k d_j(x_k) \geq 1$ for all $j \in [1\!:\!m]$. Now let $\eta := \sum_{k=1}^n \lambda_k$. Then $\eta > 0$. To see why, assume *ex absurdo* that $\eta \leq 0$. Then for all $k \in [1\!:\!n]$, $\lambda_k \geq 0$ and $\lambda_k = \eta - \sum_{k' \in [1:n] \setminus \{k\}} \lambda_{k'} \leq 0$, and therefore $\lambda_k = 0$, but this would imply that $\sum_{k=1}^n \lambda_k d_j(x_k) = 0 \not\geq 1$, which is a contradiction. Hence, indeed, $\eta > 0$. Now define $p(x_k) := \lambda_k/\eta$ for all $k \in [1\!:\!n]$. Then $p \in \Sigma$, because $\sum_{k=1}^n p(x_k) = 1$ and $p(x_k) \geq 0$, and $E_p(d_j) = \sum_{k=1}^n \lambda_k d_j(x_k)/\eta \geq \frac{1}{\eta} > 0$ for all $j \in [1\!:\!m]$ because $\sum_{k=1}^n \lambda_k d_j(x_k) \geq 1$ and $\eta > 0$. This means that $p \in \mathcal{P}(d_{1:m})$. Furthermore, we also have that $E_p(u_i) = \sum_{k=1}^n \lambda_k u_i(x_k)/\eta \geq 0$ for all $i \in [1\!:\!\ell]$, because $\sum_{k=1}^n \lambda_k u_i(x_k) \geq 0$ and $\eta > 0$.

Next we prove that (ii) is equivalent to (iii). We will use bold letters for vectors and matrices. Note that $\text{PRIMAL}(u_{1:\ell}, d_{1:m}) = \text{True}$ if and only if there is some $\boldsymbol{\lambda} \in \mathbb{R}^n$ such that $\boldsymbol{\lambda} \geq \mathbf{0}$ and $\mathbf{A}\boldsymbol{\lambda} \leq \mathbf{b}$, with

$$\mathbf{A} = \begin{pmatrix} -u_1(x_1) & \cdots & -u_1(x_n) \\ \vdots & & \vdots \\ -u_\ell(x_1) & \cdots & -u_\ell(x_n) \\ -d_1(x_1) & \cdots & -d_1(x_n) \\ \vdots & & \vdots \\ -d_m(x_1) & \cdots & -d_m(x_n) \end{pmatrix} \text{ and } \mathbf{b} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ -1 \\ \vdots \\ -1 \end{pmatrix}.$$

Farkas's Lemma (Matoušek and Gärtner, 2006, Proposition 6.4.3(ii)) tells us that this is equivalent to the condition that all $\mathbf{y} \in \mathbb{R}^{m+\ell}$ that satisfy $\mathbf{y} \geq \mathbf{0}$ and $\mathbf{y}^T \mathbf{A} \geq \mathbf{0}^T$ also satisfy $\mathbf{y}^T \mathbf{b} \geq 0$. By propositional logic, this is equivalent to the fact that there is no $\mathbf{y} \in \mathbb{R}^{m+\ell}$ that satisfies $\mathbf{y} \geq \mathbf{0}$, $\mathbf{y}^T \mathbf{A} \geq \mathbf{0}^T$ and $\mathbf{y}^T \mathbf{b} < 0$. Since multiplying $\mathbf{y}$ with a positive scalar has no effect on the veracity of these inequalities, this is in turn equivalent to the fact that there is no $\mathbf{y} \in \mathbb{R}^{m+\ell}$ such that $\mathbf{y} \geq \mathbf{0}$, $\mathbf{y}^T \mathbf{A} \geq \mathbf{0}^T$ and $\mathbf{y}^T \mathbf{b} \leq -1$, which holds if and only if $\text{DUAL}(u_{1:\ell}, d_{1:m}) = \text{False}$.