

## On the Identifiability of Hierarchical Decision Models

Roman Bresson<sup>1,3</sup>, Johanne Cohen<sup>3</sup>, Eyke Hüllermeier<sup>4</sup>, Christophe Labreuche<sup>1,2</sup>,  
 Michèle Sebag<sup>3</sup>

<sup>1</sup>Thales Research and Technology, Palaiseau, France

<sup>2</sup>SINCLAIR, Industrial AI lab, Saclay, France

<sup>3</sup>LISN CNRS-INRIA, Université Paris Saclay, France

<sup>4</sup>University of Munich, Germany

{roman.bresson, christophe.labreuche}@thalesgroup.com, {johanne.cohen, michele.sebag}@lri.fr,  
 eyke@ifi.lmu.de

### Abstract

Interpretability is a desirable property for machine learning and decision models, particularly in the context of safety-critical applications. Another most desirable property of the sought model is to be unique or *identifiable* in the considered class of models: the fact that the same functional dependency can be represented by a number of syntactically different models adversely affects the model interpretability, and prevents the expert from easily checking their validity. This paper focuses on the Choquet integral (CI) models and their hierarchical extensions (HCI). HCIs aim to support expert decision making, by gradually aggregating preferences based on criteria; they are widely used in multi-criteria decision aiding and are receiving interest from the Machine Learning community, as they preserve the high readability of CIs while efficiently scaling up w.r.t. the number of criteria. The main contribution is to establish the identifiability property of HCI under mild conditions: two HCIs implementing the same aggregation function on the criteria space necessarily have the same hierarchical structure and aggregation parameters. The identifiability property holds even when the marginal utility functions are learned from the data. This makes the class of HCI models a most appropriate choice in domains where the model interpretability and reliability are of primary concern.

## 1 Introduction

Decision aiding models aim at supporting human decision makers (DM) in selecting the most appropriate alternatives with respect to the current context, based on preference relations among the criteria involved. Multi-criteria decision aid (MCDA) models are instrumental in a wide range of critical applications, where the decision rests upon the DM, who assesses the alternatives sorted by the model based on their full awareness of the context.

MCDA models usually proceed in two steps. In a first step, the value of each attribute (criterion) is turned into a utility score using the so-called *marginal utility functions*. In a second step, the utility scores are combined via an *aggregation function*, defining the aggregated utility of the alternative, and a decision is taken based on this aggregated utility (Fig. 1). For a better scalability, the criteria are usually structured along a hierarchy; thus, the aggregation function has a recursive structure, computing the aggregated utility of each hierarchy node from that of its child nodes.

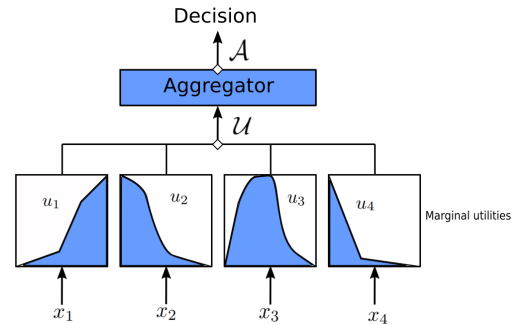


Figure 1: Decision process

**Example 1.** A decision maker (DM) wishes to buy a house. Let us assume that houses are defined by seven criteria: 1. House surface area, 2. Garden surface area, 3. Garage (yes/no), 4. Proximity to a large road, 5. Proximity to public transportation, 6. Proximity to downtown, 7. Price. These criteria are gradually aggregated as illustrated in Fig. 2 to form new compound criteria: 8. Commodities, 9. Building comfort, 10. Accessibility, and 11. Global score. The goal of an MCDA model is to rank the alternatives corresponding to the different ads, allowing the DM to quickly retrieve the best opportunities according to their own preferences.

In many application domains related to decision making and artificial intelligence at large, an essential requirement is to gain the users' trust (O'Neill 2016). To this end, the model must be *interpretable*, that is, the DMs must un-

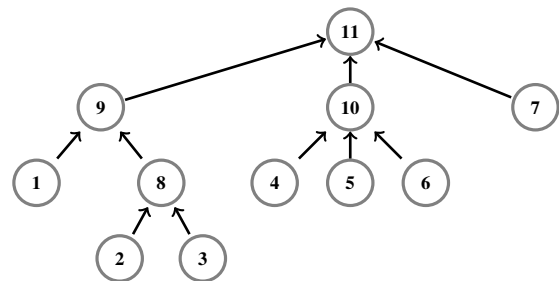


Figure 2: Hierarchy of criteria described in Ex 1.

derstand which criteria influence the decision, how, and to which extent; in other words, they must be able to trace back the assessment of an alternative to the criteria involved. In some cases, syntactic constraints (e.g. monotonicity) might be enforced to facilitate the interpretation of the model; in other cases, specific domain knowledge is available (e.g. implying some preferences w.r.t. some criteria, everything else being equal), and the model must comply with this prior knowledge. Naturally, the trust-worthiness of the model is all the more important in safety-critical contexts. Related with the interpretability and verifiability of the model is its identifiability. Typically, ambiguities arise if the same utility function can be represented in different ways, preventing the DM from understanding the impact of each attribute.

Formally, let us consider a parametric model space  $\mathcal{H}$ , where each model is specified from a parameter vector depending on the attribute space, representing the importance of each criterion and of their interactions. The identifiability property is formalized as follows: letting  $\mathcal{F}$  and  $\mathcal{F}'$  denote two models, parameterized by  $\theta$  (resp.  $\theta'$ ),  $\mathcal{H}$  is identifiable iff  $\mathcal{F} = \mathcal{F}'$  implies  $\theta = \theta'$  (Rothenberg 1971).

Note that the identifiability property is also relevant in the machine learning context: the existence of a single solution brings several benefits to the learning (Paulino and Pereira 1994; Ran and Hu 2017).

This paper focuses on the class of Choquet integral (CI) models (Choquet 1953), widely used in Multi-Criteria decision aiding for their high readability. More specifically, we focus on the hierarchical extension of CIs (HCIs), retaining the CI readability while enabling a better scalability w.r.t. the number of criteria, through a gradual aggregation of the criteria along the considered hierarchy. Recent advances in machine learning have been devoted to the data-driven modeling of both CIs (Fallah Tehrani et al. 2012; Havens and Anderson 2019) and HCIs (Bresson et al. 2020b). The contribution of the paper is to establish the identifiability of HCIs, formally showing that, if two HCIs are equal for all possible alternatives, then they have to have the same parameters, and thus the same interpretation. This result encompasses the identifiability of the marginal utilities from the raw criteria data, of the aggregation hierarchy, and of the aggregation parameters. Moreover, it bridges a gap between both fields of Machine Learning – where powerful black-box models are learned with no general identifiability guarantees – and MCDA, where models are constrained and need to be interpretable. The HCI model space thus offers an interesting trade-off for safety-critical applications, enabling the data-driven learning of interpretable and identifiable models.

The paper is organized as follows. Section 3 describes the formal background and introduces MCDA models. Section 4 defines the identifiability problem, and establishes the HCI (and marginal utilities) identifiability when the hierarchy is fixed. Section 5 finally establishes the identifiability of both the hierarchy and the parameters of the sought model.

## 2 Related Work

Due to their desirable properties, CIs have long been popular decision models (Grabisch and Perny 2003; Grabisch and

Labreuche 2010). A more general, multi-layered form was introduced in (Ovchinnikov 2002), and later refined in (Grabisch et al. 2009; Angilella et al. 2013) in order to support a tree-like hierarchical structure. This hierarchical extension, referred to as HCI, enables to represent some given structure on the criteria, with better interpretability than flat CIs when the number of criteria increases (Labreuche and Fossier 2018). HCIs can also be fitted with marginal attribute rescalings, called marginal utilities, for a better expressivity (Bresson et al. 2020b).

The interest of this class of models has motivated several approaches for eliciting the CI parameters. These methods are mostly either constraint-based (Grabisch, Kojadinovic, and Meyer 2008; Pelissari and Duarte 2020) or machine learning-based (Fallah Tehrani et al. 2012; Havens and Anderson 2019). There was also work on learning HCIs with marginal utilities and a given hierarchy (Bresson et al. 2020b).

These approaches raise the issue of knowing whether the HCI, when fitted with marginal utilities, is identifiable or not in the sense of (Rothenberg 1971) (more in Section 4). In (Bouyssou and Marchant 2010), the authors presents some results on identifiable marginal utilities for ordered classification. Bishop (2006) states that it is not a trivial property for hierarchical models. While some results have been presented for Gaussian-based hierarchical models (Xie and Carlin 2006; Ogle and Barber 2020), they do not extend to HCIs.

As said, this paper tackles the HCI identifiability and presents mild assumptions enforcing the identifiability property. The impact of this result might be appreciated in the view of (Ovchinnikov 2002), showing that CIs organized in a Directed Acyclic Graph (DAG) are not identifiable. The restriction of DAGs to trees is here shown to be sufficient to achieve identifiability.

## 3 Formal Background

### 3.1 General Preference Model

Let us consider a finite set of  $n$  attributes taking values in domains  $X_1, \dots, X_n$ . With no loss of generality, it is assumed that  $X_i \subseteq \mathbb{R} = [-\infty, +\infty]$  for all  $i = 1$  to  $n$ . Let  $N = \{1, \dots, n\}$ ,  $Y := [0, 1]$  and  $U := [0, 1]^n$ . An alternative  $\mathbf{x} = (x_1, \dots, x_n)$  is an element of  $X = X_1 \times \dots \times X_n \subseteq \mathbb{R}^n$ .

In preference modeling problems, the goal is to represent the preferences of a decision maker (DM) over elements of  $X$ . More precisely, we need a *numerical representation*  $\mathcal{F} : X \rightarrow Y$  of the preferences of the DMs, such that  $\mathbf{x}$  is *at least as good as*  $\mathbf{y}$  if and only if  $\mathcal{F}(\mathbf{x}) \geq \mathcal{F}(\mathbf{y})$  (Krantz et al. 1971). Scale  $Y$  is interpreted as a degree of preference where value 1 (resp. 0) represents a complete satisfaction (resp. dissatisfaction).

Many MCDA models are written using the *decomposable form* (Keeney and Raiffa 1976):

$$\mathcal{F}(\mathbf{x}) = \mathcal{A}(u_1(x_1), \dots, u_n(x_n)) \quad \forall \mathbf{x} \in X \quad (1)$$

where the  $u_i : X_i \rightarrow Y$  are called the *marginal utilities*, and  $\mathcal{A} : U \rightarrow Y$  is an *aggregation function*. For  $i \in N$ ,  $u_i$  represents a marginal preference on the values of  $x_i$  (see Ex. 2),

with  $u_i(x_i) \geq u_i(x'_i)$  if value  $x_i$  for attribute  $i$  is more satisfying than  $x'_i$ . A *criterion* is an attribute  $X_i$  together with its marginal utility  $u_i$ .

**Example 2.** [Ex. 1 cont'd] Consider attribute 4: proximity to a large road (in kilometers). The DM prefers to be close enough for easy access, but far enough that they are not bothered by the noise. Assuming 0.1km is too close, 2km is optimal, and 10km is too far. This preference is reflected by  $u_4(2) > u_4(0.1)$  and  $u_4(10) < u_4(2)$ .

The aggregation function  $\mathcal{A}$  combines the normalized scores of the criteria and returns the overall score on the same scale. In order to capture the DM preferences,  $\mathcal{A}$  should satisfy the following monotonicity condition:

$$\forall \mathbf{a}, \mathbf{a}' \in U, \mathbf{a} \geq \mathbf{a}' \Rightarrow \mathcal{A}(\mathbf{a}) \geq \mathcal{A}(\mathbf{a}'), \quad (2)$$

where  $\mathbf{a} \geq \mathbf{a}'$  iff  $a_i \geq a'_i$  for all  $i \in N$ , with  $\mathbf{a} = (a_1, \dots, a_n)$ . All value functions return scores on the same scale. Hence an alternative with same satisfaction degree  $\alpha$  on all criteria yields an overall score of  $\alpha$ :

$$\mathcal{A}(\alpha, \dots, \alpha) = \alpha \quad \forall \alpha \in [0, 1]. \quad (3)$$

This condition is called *idempotency*. E.g., if the alternative is *good* on all criteria, then it remains globally *good*.

### 3.2 Aggregation Model $\mathcal{A}$

The simplest aggregation function is a weighted sum (WS).

$$\text{WS}_{\mathbf{w}}(\mathbf{a}) = \sum_{i \in N} w_i a_i, \quad (4)$$

where the vector  $\mathbf{w} = (w_1, \dots, w_n)$  are the criteria weights. They satisfy  $w_i \geq 0$  and  $\sum_{i \in N} w_i = 1$  in order to fulfill monotonicity (2) and idempotency (3) properties. Even though this model is very easy to understand, it is limited by the fact that it assumes that criteria are independent.

**Example 3** (Ex. 1 cont.). We assume a sub-domain of Ex. 1, where we can evaluate houses only on criteria 4, 5 and 6, focusing on aggregation node 10 and its children in Fig. 2. As we focus on the aggregation, we assume the marginal utilities to have already been computed. Thus, we represent a house as a vector of satisfactions on criteria 4, 5, 6; that is, a house that is totally satisfying on criterion 4 and dissatisfying on criteria 5, 6 is represented as  $(1, 0, 0)$ . Let  $\succ$  and  $\sim$  be a preference relation and an equivalence relation on houses restricted to criteria 4, 5, 6. The DM considers that public transportation will mostly be used for accessing the downtown hub; it is thus equivalently satisfying to be close to downtown, to the public transportation, or to both. This means that, given the satisfaction  $\alpha$  on criterion 4, we have

$$(\alpha, 0, 1) \sim (\alpha, 1, 0) \sim (\alpha, 1, 1). \quad (5)$$

The DM also considers that having only access to the city (5. or 6.) is as satisfying as having only access to the outside world (4.), but is still better than having access to neither. Thus

$$(0, 1, 0) \sim (1, 0, 0) \succ (0, 0, 0). \quad (6)$$

Given a WS, first equation gives us  $w_5 = w_6 = 0$ , and thus  $w_4 = 1$ . On the other hand, the second one gives  $w_4 = w_5 > 0$ ; we can thus not represent the DM's rules with a WS.

A classic way to generalize this model is to consider a piecewise affine model PA( $\mathbf{a}$ ) characterized by a partition  $\mathcal{D}$  of  $U$  composed of a finite set of domains such that PA is a (monotone and idempotent) weighted sum within each domain  $D \in \mathcal{D}$ , and PA is continuous.

This model is very general and can approximate any continuous and smooth (e.g. Lipschitz continuous) function, a phenomenon which is in particular exploited in ReLU-based neural networks (Glorot, Bordes, and Bengio 2011).

The price to pay for such flexible functions is their lack of interpretability. The question thus becomes to find a good trade-off between flexible, piecewise affine functions, and interpretable functions.

The class of Choquet integral models thus is considered.

### 3.3 Choquet Integral

In order to generalize the weighted sum aggregation model, the idea is to assign weights not only to single variables but also to any subset of variables.

**Definition 1.** A fuzzy measure (FM) on a set  $N$  is a set function  $\mu : 2^N \rightarrow [0, 1]$  such that

- $\mu(\emptyset) = 0$ ;  $\mu(N) = 1$
- $B \subseteq A \subseteq N \Rightarrow \mu(B) \leq \mu(A)$

Given a set  $A \subseteq N$ ,  $\mu(A)$  can be seen as the weight of the importance of all the criteria in  $A$ . Therefore, it is at least as large on a set than on one of its subsets.

**Example 4** (Ex. 3 cont.). We can define the following FM  $\mu$  defined on set  $\{4, 5, 6\}$  at aggregation node 10:  $\mu(\emptyset) = 0$ ,  $\mu(\{4\}) = \mu(\{5\}) = \mu(\{6\}) = \mu(\{5, 6\}) = 0.5$ , and  $\mu(\{4, 5\}) = \mu(\{4, 6\}) = \mu(\{4, 5, 6\}) = 1$ .

The Choquet integral is defined as follows.

**Definition 2.** (Choquet 1953) The Choquet integral (CI) of a vector  $\mathbf{a} = (a_1, \dots, a_n)$  parameterized by FM  $\mu$  is defined as:

$$C_{\mu}(\mathbf{a}) = \sum_{i=1}^n a_{\sigma(i)} (\mu(A_i) - \mu(A_{i+1})) \quad (7)$$

with  $\sigma$  the reordering of the indices such that  $a_{\sigma(1)} \leq a_{\sigma(2)} \leq \dots \leq a_{\sigma(n)}$  and  $A_i = \{\sigma(i), \dots, \sigma(n)\}$ .

The conditions on a FM ensure that the CI satisfies monotonicity (2) and idempotency (3) properties.

**Example 5** (Ex. 4 cont.). We wish to show that the FM given in Ex. 4 fulfills the preferences given in Ex. 3 – namely relations (5) and (6). Let us compute

$$C_{\mu}(\alpha, 0, 1) = \alpha(\mu(\{4, 6\}) - \mu(\{6\})) + 1\mu(\{6\}) = \frac{\alpha}{2} + \frac{1}{2}.$$

Likewise,  $C_{\mu}(\alpha, 1, 0) = \frac{\alpha}{2} + \frac{1}{2} = C_{\mu}(\alpha, 1, 1)$ , so that condition (5) is satisfied.

We also check that  $C_{\mu}(0, 1, 0) = \mu(\{5\}) = \frac{1}{2} = C_{\mu}(1, 0, 0) > C_{\mu}(0, 0, 0) = 0$ . Hence (6) is also satisfied.

The CI can represent a wide range of decision behaviours, including interactions among criteria (Grabisch and Labreuche 2010). This model can be interpreted thanks to indicators derived from the Cooperative Game Theory

that can be directly applied to the FM. The Shapley value (Shapley 1953) indicates the mean importance of each criterion, and the interaction indices (Murofushi and Soneda 1993; Grabisch and Roubens 1999) measures the degree to which a subset of criteria interact.

The CI can also be written in the following way:

$$C_\mu(\mathbf{a}) = \sum_{A \subseteq N} m(A) \cdot \min_{i \in A} a_i, \quad (8)$$

where  $m$  is the Möbius representation of  $\mu$ , defined by  $m(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} \mu(B)$ .

In order to reduce the exponential complexity, one usually requires that most Möbius coefficients are zero. We set  $\mathcal{S} = \{A \subseteq N : m(A) \neq 0\}$ . For instance a FM is said to be  $k$ -additive (for  $k \in \{1, \dots, n\}$ ) when  $\mathcal{S} = \{A \subseteq N : |A| \leq k\}$  (Grabisch 1997). Index  $k = 2$  is often considered in practice as it is a good trade-off between the representation power and the interpretability of the model.

### 3.4 Hierarchical Choquet Integral

**Towards a Generalization of the Choquet Integral** By Eq. (8), a CI is a piecewise affine function, involving subdomains where the CI returns a weighted sum, and the domains being separated by hyperplanes of the form  $a_i \geq a_j$ . While these domains make sense in the view of the idempotency property, one would like to generalize them.

It is shown in (Grabisch et al. 2009), based on (Ovchinnikov 2002), that any piecewise affine function can be represented by a network of interconnected CIs. This result can be interpreted as: a set of CIs aggregate all criteria, and their output are then aggregated by a final CI. While this result shows the generality of interconnected CIs, the lack of structure on these CIs might harm the overall interpretability, all the more so as there is no restriction on number of CIs, possibly yielding very large models.

Moreover, as each CI aggregates all inputs, the traceability of the criteria impact and their interactions is lost. The model needs to be inspected in every detail to be understood.

**HCI Model** The hierarchical Choquet Integral (HCI) model retains the generalization from a single CI to a network of CIs. Nonetheless, it restricts their structure to a tree.

Each leaf of the tree corresponds to a single criterion, and each criterion appears as a single leaf. Accordingly, the non-leaf nodes in the tree have a clear meaning to the DM, as illustrated in Fig. 2 in Ex. 1. Note that, whatever the structure of the tree, the number of aggregations and the size of each aggregation is at most  $n - 1$ . Therefore, the number of parameters is lower or equal to that of a CI.

**Example 6** (Ex. 1 cont.). Consider the tree in Fig. 2. With  $2^k$  parameters for an aggregation node with  $k$  children, we have  $4 + 4 + 8 + 8 = 24$  parameters in a model with that hierarchy. On the other hand, a flat model, with a single aggregation, would have  $2^7 = 128$  parameters. The number of parameters is thus greatly reduced by a given hierarchy.

**Definition 3.** A hierarchy on  $N$  consists of a directed rooted tree  $\mathcal{T} = \langle r, M, \text{Ch} \rangle$  where  $M$  is the set of vertices (leaves

included). We write  $V$  the set of non-leaf nodes, and  $N$  the set of leaves (as each leaf corresponds to a native criterion), with thus  $M = V \cup N$ .  $r \in V$  is the root node (the top aggregation node) and  $\text{Ch} : M \rightarrow 2^V$  is the set of children of every node.

For  $k \in V$ , let  $\text{Lf}(k)$  denote the set of leaves in the subtree of  $\mathcal{T}$  rooted at  $k$ . We also write  $\mathbf{x}_k$  (resp.  $X_k$ ) the restriction of  $\mathbf{x}$  (resp.  $X$ ) to the attributes in  $\text{Lf}(k)$ .

For  $k \in V$ , let  $d(k)$  denote the number of children nodes of  $k$ :  $d(k) = |\text{Ch}(k)|$ . We suppose the children to be ordered, and we write  $\text{Ch}(k) = \{k_1, \dots, k_{d(k)}\}$ . Each non-leaf node  $k \in V$  has an associated FM  $\mu_k$  on the set  $\text{Ch}(k)$ . A hierarchical Choquet integral (HCI) on  $\mathcal{T}$  is the set of hierarchical functions  $\mathcal{A} = \{\mathcal{A}_k\}_{k \in M}$  where  $\mathcal{A}_k : Y^{|\text{Lf}(k)|} \rightarrow Y$  is computed recursively as:

$$\mathcal{A}_k(\mathbf{a}) = C_{\mu_k}(\mathcal{A}_{k_1}(\mathbf{a}_{k_1}), \dots, \mathcal{A}_{k_{d(k)}}(\mathbf{a}_{k_{d(k)}}))$$

if  $k \in V$ , and  $\mathcal{A}_k \equiv \text{Id}$  if  $k \in N$ .

**Example 7.** In Figure 2, nodes 1 to 7 are native criteria, and nodes 8 to 12 are aggregation nodes. Their values is computed as a CI of their respective children in the tree, which can be either criteria or aggregation nodes. In particular, Accessibility (node 10) will aggregate the proximity to a large road, to downtown and to public transportation systems (respectively nodes 4, 5, and 6).

HCIs are more general than CIs (a CI being an HCI with a 1-node hierarchy). HCIs, while preserving all CI properties, bring along additional benefits. Essentially, the hierarchical decomposition involves many small, and thus highly interpretable, aggregations (as opposed to a single, big one). This decomposition is most appreciated for large  $n$  values to avoid cognitive fatigue, as humans can hardly keep more than 7 elements in mind at the same time, according to Miller's law (Miller 1956). The reduced number of parameters also helps to learn such models, by limiting overfitting.

This also allows for a local analysis of only the necessary parameters. In Example 1, if two houses have similar evaluations, the DM might wish to look at the direct nodes, and see that one offers more ease of access, while the other is much more comfortable. Nonetheless, if two artificial criteria still have similar values, the DM can "focus" on the associated node, and get a finer explanation, and so on, going as deep as they want into the hierarchy.

A hierarchical version of the Shapley value was also established for analyzing such models (Labreuche and Fossier 2018). Finally, the normalization and monotonicity constraints are still valid, as a composition of CIs.

Due to the compensatoriness property, a CI on a single criterion is the identity function. We thus impose that every non-leaf node  $k$  has at least two children ( $d(k) \geq 2$ ).

**Definition 4.** Let  $\mathcal{T}$  be a hierarchy with root  $r$  on the set of criteria  $N$ . Let

- $\mathcal{U} : X \rightarrow U$  be a function, composed by  $n$  marginal utilities  $\{u_1, \dots, u_n\}$ , one on each native attribute i.e. only on the leaves. With the above notation,  $\mathcal{U}(\mathbf{x}) = (u_1(x_1), \dots, u_n(x_n))$ , thus,  $\mathcal{U}$  maps an alternative  $\mathbf{x}$  to the vector corresponding to  $\mathbf{x}$ 's criteria-wise satisfaction in the utility space.

- $\mathcal{A} : U \rightarrow Y$  be an HCI.

Function  $\mathcal{F} : X \rightarrow Y$  defined by  $\mathcal{F} = \mathcal{A} \circ \mathcal{U}$  is called a Utilitarian Hierarchical Choquet Integral (UHCI).

The utilities are computed recursively from the leaves to the root node. Given  $\mathbf{x}_N \in X$ , we first compute the utilities on the criteria:  $a_i = u_i(x_i)$ , for  $i \in N$  on the leaves. Then the utility at node  $k \in V$  is given by  $a_k = \mathcal{F}_k(\mathbf{x}_k) = \mathcal{A}_k(\mathbf{a}_{\text{Ch}(k)})$ , where it aggregates the utility values of its children. Finally the overall utility is the utility  $a_r$  at the root node:  $a_r = \mathcal{F}_r(\mathbf{x}_N) = \mathcal{F}(\mathbf{x}_N)$ .

**Example 8.** [Ex. 5 cont.] The recursive computation of the utility values in the tree of Fig. 2 is done as follows:

- $a_1 = u_1(x_1), \dots, a_7 = u_7(x_7)$ ,
- $a_8 = C_{\mu_8}(a_2, a_3)$ ,
- $a_9 = C_{\mu_9}(a_1, a_8)$ ,
- $a_{10} = C_{\mu_{10}}(a_4, a_5, a_6)$ ,
- $a_{11} = C_{\mu_{11}}(a_9, a_{10}, a_7)$ .

Note that Ex. 5 gives an illustration of the computation of  $a_{10}$  given the value of its children  $a_4, a_5, a_6$ . Here  $\mu_{10}$  is the FM given in Ex. 4. The overall score is then the score  $a_{11}$  of the root node 11.

## 4 Identifiability of a UHCI Model with a Fixed Hierarchy

As said, the main contribution of the paper is establishing the identifiability of UHCI models. The *identifiability* of a model is the uniqueness of its parameterization. That is, let  $\mathcal{C} = \{\mathcal{F}_\theta, \theta \in \Theta\}$  be a family of functions defined on  $X$ , and with parameters in the parameter space  $\Theta$ . Let  $F_\theta$  (resp.  $F_{\theta'}$ )  $\in \mathcal{C}$  parameterized by  $\theta$  (resp.  $\theta'$ ). Then  $\mathcal{C}$  is identifiable if and only if:  $\forall \mathbf{x} \in X, F_\theta(\mathbf{x}) = F_{\theta'}(\mathbf{x}) \Rightarrow \theta = \theta'$ .

This section focuses on the case where the hierarchy of the HCI is fixed, and establishes the identifiability property under this assumption. Let us consider two UHCIs that are equal everywhere on the input space, and which are assumed to have the same hierarchy denoted by  $\mathcal{T}$ . Proving Identifiability of a UHCI model with fixed hierarchy means that:

- (A) the marginal utilities are equal on both models for each criterion;
- (B) the fuzzy measures of each of the aggregators are equal between both models.

Under some mild assumptions Section 4.1 and 4.2 establishes (A) and (B).

### 4.1 Assumptions

To represent information on the local preferences restricted to single features (attributes), it is usual to impose some assumptions on the marginal utilities. Note that the marginal utilities used in practice are either monotonic or bitonic (either single-peaked or single-valleyed). In Ex. 1, it is increasing w.r.t. criterion 1 and single-peaked w.r.t. criterion 4.

The continuity of  $u_i$  is desirable to avoid a non-stable behavior of the local and hence the global utility. To have meaningful relative importance degrees among criteria, one

also imposes the marginal utilities to be normalized. Moreover, a very common assumption is that the smallest possible utility is 0 (a value suggesting that the corresponding criterion is not met at all), and the largest one is 1 (the criterion is satisfied). For convenience, we impose that these extreme values be reached (possibly asymptotically) for a given value on attribute  $i$ . Overall, the assumptions are summarized as follows:

$$\forall i \in N \quad u_i \text{ is continuous on } \overline{\mathbb{R}} \quad (9)$$

$$\forall i \in N \quad \inf_{x_i \in \overline{\mathbb{R}}} u_i(x_i) = 0 \quad (10)$$

$$\forall i \in N \quad \sup_{x_i \in \overline{\mathbb{R}}} u_i(x_i) = 1 \quad (11)$$

All marginal utilities evoked in the remainder of this paper are assumed to respect (9), (10) and (11).

Assume in Fig. 2 that node 9 has no effect in the aggregation 11. Then one can modify the parameters of aggregations 8 and 9, with no consequence on the global score. This illustrates that identifiability cannot be obtained if there are useless criteria. We formalize this in the following way. We say that no criterion in  $N$  is useless if

$$\forall i \in N \exists x_i, y_i \in X_i, z \in X, \quad \mathcal{F}(x_i, z_{-i}) \neq \mathcal{F}(y_i, z_{-i}) \quad (12)$$

where  $(x_i, z_{-i})$  is the alternative whose value is  $x_i$  on attribute  $i$ , and  $z_j$  on all other attributes  $j \neq i$ . We introduce the following definition on FMs.

**Definition 5.** Consider a FM  $\mu$  on  $N$ . A criterion  $i \in N$  is said to be degenerate if  $\mu(S \cup \{i\}) = \mu(S)$  for every  $S \subseteq N \setminus \{i\}$ . A fuzzy-measure is said to be non-degenerate if there is no degenerate criterion.

A first result (inspired from (Labreuche 2018)) is:

**Lemma 1.** Relation (12) holds for  $\mathcal{F}$  with marginal utility functions if and only all aggregators  $\mathcal{A}_k$  for  $k \in V$  have a non-degenerate measure.

### 4.2 Main Result

The uniqueness of the representation of the HCI model on a fixed hierarchy follows from Lemma 1:

**Theorem 1.** Let  $\mathcal{F}$  and  $\mathcal{F}'$  be two UHCIs with same hierarchy  $\mathcal{T} = \langle r, M, \text{Ch} \rangle$ , and assume that they involve different fuzzy measures and utilities. Assuming that (12) holds for  $\mathcal{F}$  and  $\mathcal{F}'$ , and that the following relation is satisfied

$$\forall \mathbf{x} \in X, \quad \mathcal{F}(\mathbf{x}) = \mathcal{F}'(\mathbf{x}), \quad (13)$$

then  $\mathcal{F}$  and  $\mathcal{F}'$  have the same parameterization, that is:

- $\forall i \in N, \forall x_i \in X_i, u_i(x_i) = u'_i(x_i)$
- $\forall k \in V, \mu_k = \mu'_k$

The proof is organized as follows (see the detailed proof in Section A.1): First,  $\mathcal{F}$  and  $\mathcal{F}'$  can be written:

$$\begin{aligned} \mathcal{F} : \mathbf{x} &\mapsto C_{\mu_r}(\mathcal{F}_{r_1}(\mathbf{x}_{r_1}), \dots, \mathcal{F}_{r_{d(r)}}(\mathbf{x}_{r_{d(r)}})), \\ \mathcal{F}' : \mathbf{x} &\mapsto C_{\mu'_r}(\mathcal{F}'_{r_1}(\mathbf{x}_{r_1}), \dots, \mathcal{F}'_{r_{d(r)}}(\mathbf{x}_{r_{d(r)}})), \end{aligned}$$

with  $\text{Ch}(r) = \{r_1, \dots, r_{d(r)}\}$ . Using this form, it is enough to show these two properties:

(\*)  $\mu_r = \mu'_r$ ,

(\*\*)  $\forall k \in \text{Ch}(r), \forall \mathbf{x} \in X_k, \mathcal{F}_k(\mathbf{x}) = \mathcal{F}'_k(\mathbf{x})$ .

Indeed, if we prove this for the root, then the proof applies on the children as well, as they are all UHCI. We can then propagate the result from the root to the children. We distinguish three cases, represented each by a lemma below.

**Lemma 2.** *If  $\forall k \in \text{Ch}(r), \forall \mathbf{x} \in X_k, \mathcal{F}_k(\mathbf{x}) = \mathcal{F}'_k(\mathbf{x})$ , then we have  $\mu_r = \mu'_r$ .*

Lemma 2 means that (\*\*)  $\Rightarrow$  (\*).

Let  $\overline{B}_k = \{\mathbf{x}_k \in X_k : \mathcal{F}_k(\mathbf{x}) = 1\}$  and  $\underline{B}_k = \{\mathbf{x}_k \in X_k : \mathcal{F}_k(\mathbf{x}) = 0\}$ . The next two steps depend on the intersections of  $\underline{B}_k \cap \underline{B}'_k$  and  $\overline{B}_k \cap \overline{B}'_k$  of all children  $k \in \text{Ch}(r)$ .

**Lemma 3.** *If,  $\forall k \in \text{Ch}(r)$ ,  $[(\underline{B}_k \cap \underline{B}'_k \neq \emptyset) \text{ and } (\overline{B}_k \cap \overline{B}'_k \neq \emptyset)]$ , then we have (\*\*).*

Lemma 3, shows that if all  $\mathcal{F}_k$  and  $\mathcal{F}'_k$  have non-disjoint support for 0 and for 1, then we have (\*\*); by Lemma 2, we also have (\*). We thus have identifiability.

**Lemma 4.** *Assume there exists  $k$  in  $\text{Ch}(r)$  such that  $\underline{B}_k \cap \underline{B}'_k = \emptyset$  or  $\overline{B}_k \cap \overline{B}'_k = \emptyset$ . Then,  $\mathcal{F} \neq \mathcal{F}'$ .*

Lemma 4 shows that, should the assumptions of Lemma 3 be violated, then  $\mathcal{F}$  is necessarily different from  $\mathcal{F}'$ . As a consequence, we have an equivalence between the assumptions of Lemma 3 and that of Theorem 1; we have thus shown that, given two equal UHCIs with the same hierarchy, they have necessarily the same marginal utilities, and the same aggregations.

## 5 Full Identifiability

The aim of this section is to show the identifiability of the UHCI model in the general case, that is, when the hierarchy, the aggregation functions and the marginal utility functions are unknown. Let us start with an example to give the intuition of the approach.

**Example 9.** *Consider three criteria organized as in Fig. 3. We assume the models  $a_4 = \frac{a_1 + \min(a_1, a_2)}{2}$  and  $a_5 = \frac{a_3 + \min(a_3, a_4)}{2}$  and identity marginal utility functions. As we will see in this section,  $a_5$  is piecewise affine with respect to  $a_1, a_2, a_3$  and is composed of the following affine parts  $a_3, \frac{a_1}{2} + \frac{a_3}{2}$  and  $\frac{a_3}{2} + \frac{a_1 + a_2}{4}$  which are separated by the frontiers*

$$a_1 = a_2, a_1 = a_3 \text{ and } \frac{a_1 + a_2}{2} = a_3. \quad (14)$$

*We note that these criteria are partitioned as  $\{1, 2\}$  and  $\{3\}$  in Fig. 3, and this is exactly the structure of the last frontier in (14) (criteria  $\{1, 2\}$  appear in the left hand side while  $\{3\}$  appear in the right hand side).*

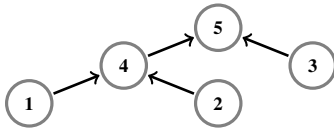


Figure 3: Example of a tree.

As illustrated in this example, the tree structure can be recovered from the expression of the separation frontiers. Theorem 1 can then be applied, and yields the uniqueness of both the aggregator weights and the marginal utility functions. So the major ingredient to show identifiability of the UHCI model is to have a good characterization of the separation frontiers (Section 5.1). In Section 5.2, we present an algorithm  $\text{Alg}_1$  which, given a UHCI under the form of its marginal utilities, hierarchy and set of fuzzy measures on its non-leaf nodes, yields the set of separation frontiers. In Section 5.3, we present an algorithm  $\text{Alg}_2$  which, given a set of separation frontiers of a UHCI model, constructs a tree compatible with these frontiers. Finally, in Section 5.4, we exhibit two conditions on the hierarchy  $\mathcal{T}$  and aggregation of a UHCI  $\mathcal{F}$ . Under those conditions,  $\text{Alg}_2(\text{Alg}_1(\mathcal{F})) = \mathcal{T}$ ; that is,  $\mathcal{T}$  is the only hierarchy that can parameterize  $\mathcal{F}$ .

### 5.1 Structure of the UHCI

From (7), we see that the CI (and thus the HCI) is a continuous, piecewise-linear function w.r.t. its inputs. It is, more precisely, a piecewise-convex combination, as the weights in all regions sum to one and are non-negative.

We consider here a UHCI model  $\mathcal{F}$  characterized by a tree  $\mathcal{T} = \langle r, M, \text{Ch} \rangle$ , a set of FM  $\{\mu_k\}_{k \in V}$  for its aggregations, and a set of marginal utility functions  $\mathcal{U} = \{u_1, \dots, u_n\}$ . In order to propagate the property of piecewise linearity of a HCI model to a UHCI model, we need to make a regularity assumption on the marginal utility functions:

$$\forall i \in N, u_i \text{ is piecewise } C^1. \quad (15)$$

This is a reasonable assumption, as it encompasses most used models such as sigmoids or piecewise affine functions. We also assume (9), (10), (11), as in Section 4.  $\mathcal{F}$  can thus be described by a finite number of  $C^1$  functions  $\mathcal{G} = \{g_1, \dots, g_m\}$ . It is, in particular, a piecewise convex combination of the output of the marginal utilities:

$$\forall g \in \mathcal{G}, g(\mathbf{x}) = \sum_{i=0}^n w_i^g u_i(x_i),$$

with non-negative weights that sum to one.

Space  $X$  can thus be split into a finite number of regions  $R_1, \dots, R_m$ , such that  $\forall \mathbf{x} \in R_i, \mathcal{F}(\mathbf{x}) = g_i(\mathbf{x})$ .  $\mathcal{F}$  is thus  $C^1$  on each of these regions, and we call  $\mathcal{H}$  the set of frontiers between each of these regions.

For a node  $k \in M$ , we call  $\mathcal{G}_k$  the set of  $C^1$  functions that  $\mathcal{F}_k$  can take, where these  $C^1$  functions depend only on  $\mathbf{x}_k$ , and  $\mathcal{H}_k$  the set of separation frontiers between the  $C^1$  regions of  $\mathcal{F}_k$  that are induced by node  $k$ .

Note that  $\mathcal{G} = \mathcal{G}_r$ , but  $\mathcal{H} = \bigcup_{k \in M} \mathcal{H}_k$ .

**Lemma 5.** *For a leaf node  $i \in N$ , the separation frontiers of  $\mathcal{H}_i$  are of the form  $x_i = \theta$  with  $\theta \in X_i$ .*

*For an aggregation node  $k \in V$ , the separation frontiers of  $\mathcal{H}_k$  take the form of linear equations of the marginal utility functions of the attributes:  $\sum_{l \in K} w_l u_l(x_l) = 0$  with  $K \subseteq \text{Lf}(k)$  and  $w_l \neq 0$  for all  $l \in K$ . Then there exist  $k', k'' \in \text{Ch}(k)$  with  $k' \neq k''$  such that we have  $\{l \in K : w_l > 0\} \subseteq \text{Lf}(k')$  and  $\{l \in K : w_l < 0\} \subseteq \text{Lf}(k'')$ .*

$\mathcal{H}$  thus contains separation frontiers of two forms. We call form 1 those of the type  $x_i = \theta$ , i.e. induced by a leaf. We call form 2 those of the type  $\sum_{l \in K} w_l u_l(x_l) = 0$ , i.e. induced by an aggregation node.

## 5.2 Construction of the Set of Separation Frontiers from the UHCI Model

We have just seen that a HCI model is a piecewise linear function. As the marginal utility functions are piecewise  $C^1$ , we conclude that a UHCI model is a piecewise  $C^1$  function.

We present Algorithm 1 in this section. This algorithm takes as an input a UHCI, characterized by its hierarchy  $\mathcal{T} = \langle r, M, \text{Ch} \rangle$ , its set of the fuzzy measures of all of its aggregators and its marginal utility functions. It then computes  $\mathcal{H}$ . Note that this algorithm also computes  $\mathcal{G}$  as an internal variable. The approach is constructive, building the two sets  $\mathcal{G}_k$  and  $\mathcal{H}_k$  for every  $k \in M$  in a bottom up manner from the leaves to the root.

Let  $k \in V$ . For FM  $\mu_k$ , we write  $\mathcal{S}_k$  the set of subsets having a non-zero Möbius coefficient in Eq. (8). As stated  $a_k$  is a piecewise linear function of the  $u_i(\mathbf{x}_i)$  for  $i \in \text{Lf}(k)$ .

**Function**  $\text{Alg}_1(\mathcal{T}, \{\mu_k\}_{k \in M}, \{u_k\}_{k \in N})$

- 1 **While** some nodes remain untreated
- 2   **Let**  $k \in M$  s.t.  $k$  has no untreated children
- 3   **If**  $k$  is a leaf
- 4      $\mathcal{G}_k :=$  set of  $C^1$  expressions of  $u_k$
- 5      $\mathcal{H}_k :=$  set of constraints  $x_k = \theta$  where  $\theta$  is a transition between two  $C^1$  domains of  $u_k$
- 6   **Else**
- 7      $\mathcal{E}_k := \{(l, l') : \exists S \in \mathcal{S}_k \text{ with } \{l, l'\} \subseteq S\}$
- 8     **For all**  $C$  connected components of the graph  $\langle \text{Ch}(k), \mathcal{E}_k \rangle$ :
- 9       Compute the linear model of  $a_k$  as a function of the  $a_{\text{Ch}(k)}$ , by transforming terms  $\min_{i \in A} a_i$  into  $a_i$  for an index  $i \in A$  depending on the order among the variables. This yields  $\sum_{l \in K} w_l a_l$  with  $K \subseteq \text{Ch}(k)$
- 10       **For all** term  $a_l$ , with  $l \in K$
- 11         **For all** expression  $\sum_{j \in \text{Lf}(l)} w_j^l a_j(x_j)$  in  $\mathcal{G}_l$
- 12         Add  $\sum_{l \in K} \sum_{j \in \text{Lf}(l)} w_l w_j^l a_j(x_j)$  to  $\mathcal{G}_k$
- 13       **For all**  $(l, l') \in \text{Ch}(k)^2$  s.t.  $\exists S \in \mathcal{S}_k : \{l, l'\} \subseteq S$
- 14         **For all**  $\sum_{j \in \text{Lf}(l)} w_j^l a_j(x_j) \in \mathcal{G}_l$
- 15         **For all**  $\sum_{j \in \text{Lf}(l')} w_j^{l'} a_j(x_j) \in \mathcal{G}_{l'}$
- 16          $\mathcal{H}_k \leftarrow \mathcal{H}_k \cup \left\{ \mathbf{x} \in X : \sum_{j \in \text{Lf}(l)} w_j^l a_j(x_j) = \sum_{j \in \text{Lf}(l')} w_j^{l'} a_j(x_j) \right\}$
- 17       Add  $\mathcal{H}_k$  to  $\mathcal{H}$
- 18 **Return**  $\mathcal{H}$

**Algorithm 1:** Obtaining  $\mathcal{H}$  (and also  $\mathcal{G}$ ) from a UHCI

We now illustrate  $\text{Alg}_1$  in an example.

**Example 10.** Consider the tree given by Fig. 4 and a UHCI model defined by  $a_6 = \frac{1}{2} \min(a_3, a_5) + \frac{1}{2} \min(a_4, a_5)$ ,  $a_5 = \frac{1}{2} a_1 + \frac{1}{2} \min(a_1, a_2)$ ,  $u_k(x_k) = x_k^{k+1}$  for  $k = 1, 2, 3$ ,

and  $u_4(x_4) = \sqrt{x_4}$  if  $x_4 \leq \frac{1}{4}$  and  $u_4(x_4) = \frac{2x_4+1}{3}$  otherwise, with  $X_1 = \dots = X_4 = [0, 1]$ . Utility function  $u_4$  is continuous and has two  $C^1$  segments.

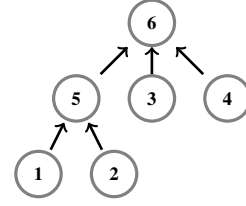


Figure 4: Example of a tree.

The execution of Algorithm 1 gives:

- At  $k = 1, 2, 3$ :  $\mathcal{G}_k = \{x_k^{k+1}\}$  and  $\mathcal{H}_k = \emptyset$ .
- At  $k = 4$ :  $\mathcal{G}_4 = \{\sqrt{x_4}, \frac{2x_4+1}{3}\}$  and  $\mathcal{H}_4 = \{x_4 = \frac{1}{4}\}$ .
- At  $k = 5$ :  $\mathcal{E}_5 = \{(1, 2)\}$  so that the connected component is  $\{1, 2\}$ . Then  $a_5$  can take the expressions  $a_1$  and  $\frac{a_1+a_2}{2}$  depending on whether  $a_1 \leq a_2$ . Hence  $\mathcal{G}_5 = \{x_1^2; \frac{x_1^2+x_2^3}{2}\}$  and  $\mathcal{H}_5 = \{x_1^2 = x_2^3\}$ .
- At  $k = 6$ :  $\mathcal{E}_6 = \{(3, 5), (4, 5)\}$  so that the connected component is  $\{3, 4, 5\}$ . Then  $a_6$  can take the expressions  $a_5$ ,  $\frac{a_3+a_5}{2}$ ,  $\frac{a_5+a_4}{2}$  and  $\frac{a_3+a_4}{2}$  depending on whether  $a_5 \leq a_3$  and  $a_5 \leq a_4$ . Hence  $\mathcal{G}_6$  is composed of
  - $x_1^2$  [for  $a_1 < a_2, a_5 < a_3, a_5 < a_4$ ]
  - $\frac{x_1^2+x_2^3}{2}$  [for  $a_1 > a_2, a_5 < a_3, a_5 < a_4$ ]
  - $\frac{x_1^2}{2} + \frac{x_3^4}{2}$  [for  $a_1 < a_2, a_5 > a_3, a_5 < a_4$ ]
  - $\frac{x_1^2+x_2^3}{4} + \frac{x_3^4}{2}$  [for  $a_1 > a_2, a_5 > a_3, a_5 < a_4$ ]
  - $\frac{x_1^2}{2} + \frac{\sqrt{x_4}}{2}$  [for  $a_1 < a_2, a_5 < a_3, a_5 > a_4, a_4 < \frac{1}{4}$ ]
  - $\frac{x_1^2+x_2^3}{4} + \frac{\sqrt{x_4}}{2}$  [for  $a_1 > a_2, a_5 < a_3, a_5 > a_4, a_4 < \frac{1}{4}$ ]
  - $\frac{x_1^2}{2} + \frac{2x_4+1}{6}$  [for  $a_1 < a_2, a_5 < a_3, a_5 > a_4, a_4 > \frac{1}{4}$ ]
  - $\frac{x_1^2+x_2^3}{4} + \frac{2x_4+1}{6}$  [for  $a_1 > a_2, a_5 < a_3, a_5 > a_4, a_4 > \frac{1}{4}$ ]
  - $\frac{x_3^4}{2} + \frac{\sqrt{x_4}}{2}$  [for  $a_5 > a_3, a_5 > a_4, a_4 < \frac{1}{4}$ ]
  - $\frac{x_3^4}{2} + \frac{2x_4+1}{6}$  [for  $a_5 > a_3, a_5 > a_4, a_4 > \frac{1}{4}$ ]
 and  $\mathcal{H}_6 = \{x_1^2 = x_3^4; \frac{x_1^2+x_2^3}{2} = x_3^4; x_1^2 = \sqrt{x_4}; x_1^2 = \frac{2x_4+1}{3}; \frac{x_1^2+x_2^3}{2} = \sqrt{x_4}; \frac{x_1^2+x_2^3}{2} = \frac{2x_4+1}{3}\}$ .

Finally  $\mathcal{H} = \mathcal{H}_4 \cup \mathcal{H}_5 \cup \mathcal{H}_6$ .  $\square$

The next result shows that Algorithm 1 computes the set  $\mathcal{G}$  of  $C^1$  functions that  $\mathcal{F}$  takes, and returns the set of separation frontiers  $\mathcal{H}$  between these functions.

**Lemma 6.** Algorithm 1 computes the correct values of the set of piecewise  $C^1$  functions  $\mathcal{G}_k$  and returns the correct set of separation frontiers  $\mathcal{H}_k$ , for each  $k \in M$ .

Lemma 6 states, in particular, that the output of  $\text{Alg}_1$  depends only on the values of the UHCI as a piecewise- $C^1$  function. The following corollary is a simple rewriting.

**Corollary 1.** Let  $\mathcal{F}$  (resp.  $\mathcal{F}'$ ) be a UHCI defined on hierarchy  $\mathcal{T}$  (resp.  $\mathcal{T}'$ ), a set of FMs  $\nu$  (resp.  $\nu'$ ) and a set of marginal utility functions  $\mathcal{U}$  (resp.  $\mathcal{U}'$ ). If  $\mathcal{F}(\mathbf{x}) = \mathcal{F}'(\mathbf{x}) \quad \forall \mathbf{x} \in X$ , then both UHCIs are the same piecewise  $C^1$  function. Thus, we have  $\text{Alg}_1(\mathcal{T}, \nu, \mathcal{U}) = \text{Alg}_1(\mathcal{T}', \nu', \mathcal{U}')$ .

### 5.3 Construction of the Hierarchy from the Set of Separation Frontiers

Assume that we are given the set of separation frontiers  $\mathcal{H}$  between the  $C^1$  functions. We would like to be able to recover the hierarchy from this.

The idea is that, from Lemma 5, each term  $\min_{i \in A} a_i$  (with  $A \subseteq \text{Ch}(k)$ ) in the expression of  $\mathcal{A}_k$  yields the separation frontier  $a_l = a_{l'}^{-1}$  for any  $\{l, l'\} \subseteq A$ . In other words, the positive and negative values of the weights correspond to two separate sub-trees in the tree. Given the separation frontiers  $\mathcal{H}$  (e.g.  $\frac{x_1^2 + x_2^3}{2} - \sqrt{x_4} = 0$  in Ex. 10), the idea is that the set of nodes having positive weights (e.g.  $\{1, 2\}$  in the ex.) and the set of nodes having negative weights (e.g.  $\{4\}$  in the ex.) correspond to separate subsets of criteria in the tree (e.g.  $\{1, 2\}$  and  $\{4\}$  belong to separate branches in Fig. 4).

We consider the set of separation frontiers of the 2<sup>nd</sup> form and we look at the indices of the positive and negative weights – see Lemma 5:

$$\mathcal{K} = \left\{ (K^+, K^-) : \{\mathbf{x} \in X : \sum_{l \in K} w_l u_l(x_l) = 0\} \in \mathcal{H}, \right.$$

$$\left. K^+ = \{l \in K, w_l > 0\} \text{ and } K^- = \{l \in K, w_l < 0\} \right\}.$$

Algorithm 2 provides the construction of the tree  $\mathcal{T}$  only from  $\mathcal{K}$ , following the previous idea that  $K^+$  and  $K^-$  shall belong to separate parts of the tree.

**Function**  $\text{Partition}(N, \mathcal{K})$ :

- 1  $\mathcal{R} = \{(i, j) \in N \times N, i \neq j : \exists (K^+, K^-) \in \mathcal{K}, \{i, j\} \subseteq K^+ \text{ or } \{i, j\} \subseteq K^-\}$
- 2  $\{N_1, \dots, N_q\} = \text{connected components of } \langle N, \mathcal{R} \rangle$
- 3 **For**  $l \in \{1, \dots, q\}$
- 4     **If**  $|N_l| > 1$
- 5          $\mathcal{K}_l = \{(K^+, K^-) \in \mathcal{K} : K^+ \cup K^- \subseteq N_l\}$
- 6          $\mathcal{N}_l = \text{Partition}(N_l, \mathcal{K}_l)$
- 7     **Else**
- 8          $\mathcal{N}_l = \{N_l\}$
- 9 **Return**  $(\mathcal{N}_1, \dots, \mathcal{N}_q)$

**Function**  $\text{Alg}_2(\mathcal{H})$ :

- 10 Compute  $\mathcal{K}$  from  $\mathcal{H}$
- 11 **Return**  $\text{Partition}(N, \mathcal{K})$

**Algorithm 2:** Obtaining the hierarchy from  $\mathcal{H}$ .  $N$  is either given by the problem, or trivially retrieved from  $\mathcal{H}$ .

We assume we are given a UHCI model  $\mathcal{F}$ , from which we know the separation frontiers  $\mathcal{H}$ . The following example

<sup>1</sup>Terms  $a_l$  and  $a_{l'}$  can be moved independently so that we can reach  $a_l < a_{l'}$  and  $a_l > a_{l'}$ . For instance, we can have  $a_l = 0, 1$  or  $a_{l'} = 0, 1$ .

illustrates the reconstruction of the hierarchy from  $\mathcal{H}$ . The correctness of  $\text{Alg}_2$  directly derives from Lemma 5.

**Example 11.** Let us consider a UHCI model over four attributes 1, 2, 3, 4, with the following separation frontiers:

$$\mathcal{H} = \left\{ \begin{aligned} x_1^2 + \frac{3x_2^3}{7} &= \frac{10\sqrt{x_4}}{7}; & x_1^2 + \frac{3x_3^4}{7} &= \frac{10\sqrt{x_4}}{7}; \\ x_1^2 + \frac{3x_2^3}{7} &= \frac{20x_4 + 10}{21}; & x_1^2 + \frac{3x_3^4}{7} &= \frac{20x_4 + 10}{21}; \\ x_1^2 &= x_3^4; & x_2^3 &= x_3^4; & x_4 &= \frac{1}{4}; \\ x_1^2 + \frac{x_2^3}{2} + \frac{x_3^4}{6} &= \frac{5\sqrt{x_4}}{3}; & x_1^2 + \frac{2x_3^4}{3} &= \frac{5\sqrt{x_4}}{3}; \\ x_1^2 + \frac{x_2^3}{2} + \frac{x_3^4}{6} &= \frac{10x_4 + 5}{9}; & x_1^2 + \frac{2x_3^4}{3} &= \frac{10x_4 + 5}{9} \end{aligned} \right\}.$$

We apply  $\text{Alg}_2$  to  $\mathcal{H}$ . We compute  $\mathcal{K} = \left\{ (\{1, 2\}, \{4\}), (\{1, 3\}, \{4\}), (\{1\}, \{3\}), (\{1, 2, 3\}, \{4\}), (\{1, 3\}, \{4\}), (\{2\}, \{3\}) \right\}$ . Then we compute the partitions:

- $\text{Partition}(N, \mathcal{K})$ :  $\mathcal{R} = \{(1, 2), (1, 3), (2, 3)\}$ . Hence the connected components of  $\langle N, \mathcal{R} \rangle$  are  $\{1, 2, 3\}$  and  $\{4\}$ . For the first set, we have  $\mathcal{K}_1 = \left\{ (\{1\}, \{3\}), (\{2\}, \{3\}) \right\}$ .
- $\text{Partition}(\{1, 2, 3\}, \mathcal{K}_1)$ :  $\mathcal{R} = \emptyset$ . Hence the connected components of  $\langle \{1, 2, 3\}, \emptyset \rangle$  are  $\{1\}, \{2\}$  and  $\{3\}$ .

The algorithm thus returns the tree of Fig. 5.

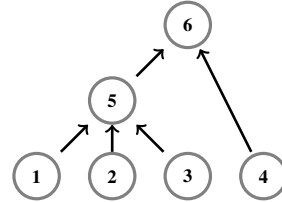


Figure 5: Tree constructed by Algorithm 2 from the separation frontiers from example 11.

### 5.4 Main Result

In order to get uniqueness properties, we need to introduce some assumptions on the HCI model.

**Assumption H1:** At every aggregation node  $k \in V$ ,  $\text{Ch}(k)$  is the only connected component of graph  $(\text{Ch}(k), \{(i, j), i \neq j \text{ s.t. } \exists S \in \mathcal{S}_k : \{i, j\} \subseteq S\})$

H1 forbids to have a model  $C_{\mu_k}$  that is (even only partly) additive. For instance,  $a_5 = C_{\mu_k}(a_1, a_2, a_3, a_4) = \frac{1}{2} \min(a_1, a_2) + \frac{1}{2} \min(a_3, a_4)$  (see Fig. 6-Left) violates H1 as groups  $\{1, 2\}$  and  $\{3, 4\}$  of variables are disconnected. In this example, we could obtain the same function with the tree of Fig. 6-Right having two new aggregation nodes:  $a_6 = \min(a_1, a_2)$ ,  $a_7 = \min(a_3, a_4)$  and thus  $a_5 = \frac{a_6 + a_7}{2}$ . Hence the hierarchy is clearly not unique in this example. On the other hand,  $C_{\mu_k}(a_1, a_2, a_3, a_4) = \frac{1}{3} \min(a_1, a_2) + \frac{1}{3} \min(a_2, a_3) + \frac{1}{3} \min(a_3, a_4)$  satisfies H1 as the four variables are connected, and one cannot decompose  $C_{\mu_k}$  with sub-aggregation nodes.



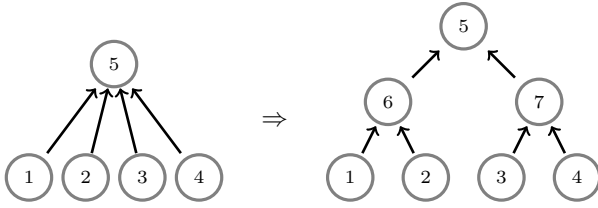


Figure 6: Illustration of H1.

Assumption H2: For all nodes  $k \in V$ :

$$|\mathcal{S}_k| \geq 2. \quad (16)$$

H2 (combined with H1) forbids from having a simple min between two variables. Intuitively, a simple min between two variables can be collapsed at the higher level. Consider the example of Fig. 7-left, where  $a_4 = \min(a_1, a_2)$  (violating H2) and  $a_5 = \frac{a_3}{2} + \frac{\min(a_3, a_4)}{2}$ . Then we can remove node 4 and directly write  $a_5$  in terms on  $a_1, a_2, a_3$ :  $a_5 = \frac{a_3}{2} + \frac{\min(a_1, a_2, a_3)}{2}$ , which is a valid Choquet integral. Hence we can also represent this model with the tree of Fig. 7-right.

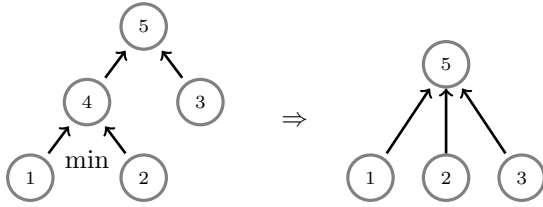


Figure 7: Illustration of H2.

The following lemma shows that under H1 and H2, applying successively Algorithm 2 on the set of separation frontiers produced by Algorithm 1 yields the same tree.

**Lemma 7.** *Let  $\mathcal{F}$  a UHCI on tree  $\mathcal{T} = \langle r, M, \text{Ch} \rangle$ , with a set of FM  $\nu = \{\mu_k, k \in V\}$  and marginal utility functions  $\mathcal{U} = \{u_1, \dots, u_n\}$ . If  $\mathcal{F}$  satisfies (12),  $\mathcal{T}$  and  $\nu$  satisfy H1 and H2, and  $\mathcal{U}$  satisfies (9), (10), (11) and (15), then  $\text{Alg}_2(\text{Alg}_1(\mathcal{T}, \nu, \mathcal{U})) = \mathcal{T}$ .*

The next result shows that the hierarchy of a UHCI model can be uniquely constructed from  $\mathcal{F}$ .

**Theorem 2.** *Under assumptions H1, H2, (9), (10), (11), (12) and (15), there is a single tree that represents a UHCI model.*

Combining Theorems 1 and 2, we obtain our main result.

**Theorem 3.** *Let  $\mathcal{F}$  and  $\mathcal{F}'$  be two UHCI satisfying (13), with potentially different hierarchies, fuzzy measures and marginal utility functions. Assume that both models fulfill H1, H2, (9), (10), (11), (12), (13) and (15). Then, both models have the same hierarchy, fuzzy measures and marginal utilities.*

It is interesting to notice that the necessary conditions are very easily understandable, and make sense to a human DM; they hence do not conflict with readability.

## 6 Conclusion and Perspectives

This paper studies a class of multi-criteria decision aid (MCDA) models. Such models are expected to satisfy certain requirements, notably interpretability and reliability. These requirements are key to establishing and maintaining the user's trust in the model, as they are aimed at supporting human decisions in safety-critical domains. This paper focuses on the class of hierarchical Choquet integral models, retaining the desirable properties of Choquet integrals while being more scalable w.r.t. the number of criteria involved.

Traditionally, such models are built by hand through an interaction with a domain expert. It would be suitable to free the experts from such a long and tedious manual crafting of the models, and to learn them from suitable data whenever available. A first step was made in this direction, by learning the marginal utilities and weights of a UHCI with an expert-given hierarchy. An empirical approach was made in (Bresson et al. 2020a), which showed that there seems to be a single attractor in this model search-space. Further theoretical work could extend the present contribution to reinforce these results, and guarantee that an expert only needs to inspect and interpret the unique HCI model fitting the data in order to validate it.

The second contribution of the paper, namely, the identifiability of the hierarchy along with the other parameters, is to our knowledge a new result in the field of decision support. This result opens two perspectives for further research. A first algorithmic perspective is to learn the hierarchy itself. The key question then concerns the sample complexity of the hierarchy learned. Another theoretical perspective concerns the HCI approximation. The question is to which extent two close models (in the functional space) are close in the parametric space. This could nicely extend similar work assuming a fixed hierarchy.

## A Proofs

### A.1 Proof of Identifiability - Fixed Hierarchy

**Proof of Lemma 2:** Let  $S \subseteq M$  a set of nodes, we write  $\text{Lf}(S) = \bigcup_{k \in S} \text{Lf}(k)$ . We use the following compound

notation. Let  $\mathbf{x}, \mathbf{x}'$  be two vectors in  $X$ , and let  $S, S' \subseteq M$  so that  $(\text{Lf}(S), \text{Lf}(S'))$  is a partition of  $N$ . Then we write

$\mathbf{z} = (\mathbf{x}_S, \mathbf{x}'_{S'})$  the vector s.t.  $\mathbf{z}_i = \begin{cases} \mathbf{x}_i & \text{if } i \in \text{Lf}(S) \\ \mathbf{x}'_i & \text{if } i \in \text{Lf}(S') \end{cases}$

Let  $\mathbf{Z} \in \prod_{k \in \text{Ch}(r)} \underline{B}_k$ , and  $\mathbf{O} \in \prod_{k \in \text{Ch}(r)} \overline{B}_k$ .

Let  $S$  be an arbitrary subset of  $\text{Ch}(r)$ . By definition, we have  $\mathcal{F}(\mathbf{O}_S, \mathbf{Z}_{\text{Ch}(r) \setminus S}) = C_{\mu_r}(1_S, 0_{\text{Ch}(r) \setminus S})$ . Thus  $\mathcal{F}(\mathbf{O}_S, \mathbf{Z}_{\text{Ch}(r) \setminus S}) = \mu_r(S)$ . In the same way, we obtain  $\mathcal{F}'(\mathbf{O}_S, \mathbf{Z}_{\text{Ch}(r) \setminus S}) = \mu'_r(S)$ .

As the values of  $\mathcal{F}$  and  $\mathcal{F}'$  are equal for any  $\mathbf{x} \in X$ , we have  $\mu_r(S) = \mu'_r(S)$ . This applies for all  $S \subseteq \text{Ch}(r)$ , thus Property (\*) holds.  $\square$

**Proof of Lemma 3:** Let  $k \in \text{Ch}(r)$ . By the assumption of the lemma, we can construct  $\mathbf{Z} \in \prod_{k \in \text{Ch}(r)} \underline{B}_k \cap \underline{B}'_k$ ,

and  $\mathbf{O} \in \prod_{k \in \text{Ch}(r)} \overline{B}_k \cap \overline{B}'_k$ . Since  $\mu_r$  is by hypothesis non-degenerate, we have that  $\exists S_k \subseteq \text{Ch}(r) \setminus \{k\}$  s.t.  $\mu_r(S_k \cup \{k\}) - \mu_r(S_k) > 0$ . Thus, we have:  $\forall \mathbf{x}_k \in X_k$

$$\begin{aligned} & \mathcal{F}(\mathbf{O}_{S_k}, \mathbf{Z}_{\text{Ch}(r) \setminus (S_k \cup \{k\})}, \mathbf{x}_k) \\ &= C_{\mu_r}(1_{S_k}, 0_{\text{Ch}(r) \setminus (S_k \cup \{k\})}, \mathcal{F}_k(\mathbf{x}_k)) \\ &= \mathcal{F}_k(\mathbf{x}_k)(\mu_r(S_k \cup \{k\}) - \mu_r(S_k)) + \mu_r(S_k) \end{aligned}$$

Likewise for  $\mathcal{F}'$ , thus, we have the following equality:

$$\begin{aligned} & \mathcal{F}_k(\mathbf{x}_k)(\mu_r(S_k \cup \{k\}) - \mu_r(S_k)) + \mu_r(S_k) \\ &= \mathcal{F}'_k(\mathbf{x}_k)(\mu'_r(S_k \cup \{k\}) - \mu'_r(S_k)) + \mu'_r(S_k) \end{aligned}$$

Since  $\mu_r(S_k \cup \{k\}) - \mu_r(S_k) > 0$  by definition of  $S_k$ , we obtain:

$$\exists \alpha \in \mathbb{R}_+, \beta \in \mathbb{R}, \forall \mathbf{x}_k \in X_k, \mathcal{F}_k(\mathbf{x}_k) = \alpha \mathcal{F}'_k(\mathbf{x}_k) + \beta$$

Applying this equality in  $\mathbf{Z}_k$  and  $\mathbf{O}_k$ , we get  $\alpha = 1$  and  $\beta = 0$ .

$$\mathcal{F}_k(\mathbf{Z}_j) = \alpha \mathcal{F}'_k(\mathbf{Z}_j) + \beta \Rightarrow 0 = 0 + \beta \Rightarrow \beta = 0$$

$$\mathcal{F}_k(\mathbf{O}_j) = \alpha \mathcal{F}'_k(\mathbf{O}_j) + \beta \Rightarrow 1 = 1\alpha \Rightarrow \alpha = 1$$

We have thus shown that the assumption on this case implies that  $\forall k \in \text{Ch}(r), \mathcal{F}_k = \mathcal{F}'_k$ .  $\square$

We define  $\underline{u}_i$  and  $\overline{u}_i$  to the marginal utility. For  $i \in N$ , we write  $\underline{u}_i = \{x_i \in X_i : u_i(x_i) = 0\}$  and  $\overline{u}_i = \{x_i \in X_i : u_i(x_i) = 1\}$ . Note that, in some cases, these values can be reached asymptotically. In these cases,  $\underline{u}_i$  (resp  $\overline{u}_i$ ) might only contain  $-\infty$  or  $+\infty$ .

**Property 1.** Let  $k \in \text{Ch}(r)$ , such that  $\overline{B}_k \cap \overline{B}'_k = \emptyset$  (resp.  $\underline{B}_k \cap \underline{B}'_k = \emptyset$ ). Then there exists a leaf  $i \in \text{Lf}(k)$  such that  $\overline{u}_i \cap \overline{u}'_i = \emptyset$  (resp.  $\underline{u}_i \cap \underline{u}'_i = \emptyset$ ).

**Proof :** We show the first result by contradiction. Let  $k \in \text{Ch}(r)$ , such that  $\overline{B}_k \cap \overline{B}'_k = \emptyset$ . We assume that  $\forall i \in \text{Lf}(k), \exists \overline{x}_i \in \overline{u}_i \cap \overline{u}'_i$ . Then there exists  $\overline{\mathbf{x}}_k = (\overline{x}_1, \dots, \overline{x}_{|\text{Lf}(k)|})$ . Thus  $\mathcal{F}_k(\overline{\mathbf{x}}_k) = \mathcal{A}_k(1, \dots, 1) = 1$ . Using the same argument, we can obtain :  $\mathcal{F}'_k(\overline{\mathbf{x}}_k) = 1$ . Thus,  $\overline{\mathbf{x}}_k \in \overline{B}_k \cap \overline{B}'_k$ , hence the contradiction.

The very same reasoning can be applied for the case  $\underline{B}_k \cap \underline{B}'_k = \emptyset$ .  $\square$

**Proof of Lemma 4:** We prove this lemma by contradiction, by assuming  $\mathcal{F} = \mathcal{F}'$ . Then, let  $i \in \text{Lf}(k)$  such that  $\overline{u}_i \cap \overline{u}'_i = \emptyset$  or  $\underline{u}_i \cap \underline{u}'_i = \emptyset$ , which exists by Property 1. There is thus an interval  $I = [\alpha_i, \beta_i] \subseteq X_i$  such that  $u_i$  and  $u'_i$  have opposite monotonicity on  $I$ . This can easily be verified by enumerating the cases. WLOG, we assume that  $u_i$  is decreasing on  $I$  while  $u'_i$  is non-decreasing.

Let  $S \subseteq N \setminus \{i\}$ . Let  $\mathbf{z}^S \in \prod_{j \in S} \overline{u}_j \times \prod_{i \in N \setminus S} \underline{u}_j$ . Then, by monotonicity of HCIs,  $u_i(\alpha_i) < u_i(\beta_i)$  implies:

$$\begin{aligned} & \mathcal{F}_k(\alpha_i, \mathbf{z}_{\text{Lf}(k) \setminus \{i\}}^S) \leq \mathcal{F}_k(\beta_i, \mathbf{z}_{\text{Lf}(k) \setminus \{i\}}^S) \\ & \Rightarrow \mathcal{F}(\alpha_i, \mathbf{z}_{N \setminus \{i\}}^S) \leq \mathcal{F}(\beta_i, \mathbf{z}_{N \setminus \{i\}}^S) \end{aligned}$$

Likewise, we have:  $\mathcal{F}'(\alpha_i, \mathbf{z}_{N \setminus \{i\}}^S) \geq \mathcal{F}'(\beta_i, \mathbf{z}_{N \setminus \{i\}}^S)$  as  $u'_i$  is non-decreasing on  $[\alpha_i, \beta_i]$ . Since  $\mathcal{F}$  and  $\mathcal{F}'$  are equal everywhere, we thus have:

$$\begin{aligned} & \forall S \subseteq N \setminus \{i\}, \\ & \mathcal{F}(\alpha_i, \mathbf{z}_{N \setminus \{i\}}^S) = \mathcal{F}(\beta_i, \mathbf{z}_{N \setminus \{i\}}^S) \end{aligned}$$

We show that this equation leads to a contradiction, by building an  $S \subseteq N$  such that we do not have the equality.

We denote by  $\pi = \{\pi_1, \dots, \pi_t\}$  the unique path from root  $r$  to leaf  $i$ , with  $t$  the depth of  $i$  in the tree. We order them so that  $\pi_1 = r$  and  $\forall j \in \{2, \dots, t\}, \pi_j \in \text{Ch}(\pi_{j-1})$ . This means that  $\pi_t = i$ .

For each  $j$  in  $\{1, \dots, t-1\}$ , let  $S_j$  be a subset of nodes in  $\text{Ch}(\pi_j) \setminus \{\pi_{j+1}\}$  such that  $\mu_{\pi_j}(S_j) > \mu_{\pi_j}(S_j \cup \{\pi_{j+1}\})$ . The non-degenerateness property implies that such sets exist. We denote by  $S^k = \bigcup_{j=1}^{t-1} S_j$ .

Now, let  $\mathbf{v}^\alpha$  be a vector in  $X$  such that :

$$\begin{cases} \mathbf{v}^\alpha = \alpha_i \\ \mathbf{v}_i^\alpha \in \overline{u}_i & \text{if } \exists g \in S^k, i \in \text{Lf}(g) \\ \mathbf{v}_i^\alpha \in \underline{u}_i & \text{otherwise} \end{cases}$$

and  $\mathbf{v}_\beta = (\mathbf{v}_{N \setminus k}^\alpha, \beta_i)$ .

We write  $d_j = \text{Ch}(\pi_j) \setminus (S_j \cup \{\pi_{j+1}\})$ . Then, at node  $\pi_j$ , we have:

$$\begin{aligned} \mathcal{F}_{\pi_j}(\mathbf{v}_{\pi_j}^\alpha) &= C_{\mu_{\pi_j}}(\mathcal{F}_{\pi_{j+1}}(\mathbf{v}_{\pi_{j+1}}^\alpha), 1_{S_j}, 0_{d_j}) \\ &= \mathcal{F}_{\pi_{j+1}}(\mathbf{v}_{\pi_{j+1}}^\alpha) \gamma_j + \delta_j \end{aligned}$$

with  $\gamma_j = \mu_{\pi_j}(S_j \cup \{\pi_{j+1}\}) - \mu_{\pi_j}(S_j)$  and  $\delta_j = \mu_{\pi_j}(S_j)$ . when  $j = t-1$ , we have:

$$\mathcal{F}_{\pi_{t-1}}(\mathbf{v}_{\pi_{t-1}}^\alpha) = \gamma_{t-1} u_i(\alpha_i) + \delta_{t-1}$$

$$\begin{aligned} \mathcal{F}_{\pi_{t-2}}(\mathbf{v}_{\pi_{t-2}}^\alpha) &= \gamma_{t-2} \mathcal{F}_{\pi_{t-1}}(\mathbf{v}_{\pi_{t-1}}^\alpha) + \delta_{t-2} \\ &= \gamma_{t-2} (\gamma_{t-1} u_i(\alpha_i) + \delta_{t-1}) + \delta_{t-2} \end{aligned}$$

And so on. Since  $\forall j \in \{1, \dots, t\}, \gamma_j > 0$  by construction of  $S_j$ , and  $\delta_j \geq 0$ , we have  $\mathcal{F}_{\pi_1}(\mathbf{v}_{\pi_1}^\alpha) = \Gamma u_i(\alpha_i) + \Delta$  with  $\Gamma > 0$  and  $\Delta \geq 0$  by composition of affine functions with a strictly increasing coefficient. Likewise, for  $\mathbf{v}^\beta$ , we obtain  $\mathcal{F}_{\pi_1}(\mathbf{v}_{\pi_1}^\beta) = \Gamma u_i(\beta_i) + \Delta$ . Nonetheless:

$$\begin{aligned} u_i(\alpha_i) < u_i(\beta_i) &\Rightarrow \Gamma u_i(\alpha_i) + \Delta < \Gamma u_i(\beta_i) + \Delta \\ &\Rightarrow \mathcal{F}(\mathbf{v}_{\pi_1}^\alpha) < \mathcal{F}(\mathbf{v}_{\pi_1}^\beta) \end{aligned}$$

We have built a set  $S \subseteq N$  such that Eq. (13) does not hold. Thus  $\mathcal{F} \neq \mathcal{F}'$  and this yields to the contradiction.  $\square$

**Proof of Theorem 1:**

$$\mathcal{F} = \mathcal{F}' \Rightarrow \forall k \in \text{Ch}(r),$$

$$\begin{aligned} & [(\underline{B}_k \cap \underline{B}'_k \neq \emptyset) \text{ and } (\overline{B}_k \cap \overline{B}'_k \neq \emptyset)] \text{ by Lemma 4} \\ & \Rightarrow \forall k \in \text{Ch}(r), \mathcal{F}_k = \mathcal{F}'_k \text{ by Lemma 3} \\ & \Rightarrow \mu_r = \mu'_r \text{ by Lemma 2} \end{aligned}$$

The proof is completed.  $\square$

## A.2 Proof of Identifiability - Free Hierarchy

**Lemma 8.**  $\forall \sum_{j \in \mathcal{S}(k)} w_j^k u_j(x_j) \in \mathcal{G}_k$ , where  $\mathcal{S}(k) \subseteq \text{Lf}(k)$ , we have  $w_j^k > 0$  for every  $j \in \mathcal{S}(k)$ .

**Proof:** If there exists  $j \in \mathcal{S}(k)$  with  $w_j^k < 0$ , then the score  $a_k$  would be locally decreasing w.r.t. this variable, which contradicts monotonicity of  $\mathcal{A}$  w.r.t. its inputs.  $\square$

**Proof of Lemma 5:** The result is clear for  $i \in N$ .

For  $k \in V$ , the separation frontiers in  $\mathcal{H}_k$  corresponds to an equation  $g_i(x) = g_j(x)$ . Nodes  $i$  and  $j$  corresponds to two distinct sub-trees with roots  $k'$  and  $k''$  respectively. Hence by Lemma 8, all coefficients of  $g_i$  and  $g_j$  are positive. Hence the result.  $\square$

**Proof of Lemma 6:** The proof is done by induction. For leaves, the separation frontiers are the points of separation of the piecewise  $C^1$  segments of the marginal utility functions. For the other nodes, the recursive construction of  $\mathcal{G}_k$  and  $\mathcal{H}_k$  is clear by Lemma 5. There is no required property on the HCI model to have this result.  $\square$

Notation: Set  $\mathcal{G}_k$  is written as a set of linear models of the marginal utility functions. Each  $g \in \mathcal{G}_k$  has a support, i.e. a set of leaves with non-zero weights; we call  $\mathcal{T}_k$  the set of the supports of all  $g \in \mathcal{G}_k$ .

$$\mathcal{G}_k = \left\{ \sum_{l \in T} w_l^T u_l(x_l), T \in \mathcal{T}_k \right\}. \quad (17)$$

We denote by  $\mathcal{G}'_k$  the set of linear models of the marginal utility functions of  $a_k$  as a function of  $\mathbf{a}_{\text{Ch}(k)}$ .

**Lemma 9.** Under assumptions (12) and H2, for every  $k \in M$ , the graph  $\langle \text{Lf}(k), \{(i, j), i, j \in T \text{ with } T \in \mathcal{T}_k\} \rangle$  is connected, where set  $\mathcal{T}_k$  in  $\mathcal{G}_k$  is given by Eq. (17).

**Proof:** The proof is done by backward induction.

For a leaf  $k \in N$ :  $\mathcal{G}_k$  is the set of  $C^1$  expressions that  $a_k$  can take,  $\mathcal{H}_k$  is the set of  $x_k = \theta$  for all  $\theta$  where function  $u_k$  has a discontinuity of its derivative, and thus the assumption is proven at this node.

For an aggregation node  $k \in V$ : Let  $\ell, \ell' \in \text{Ch}(k)$  with  $\ell \neq \ell'$ . We need to show that  $\ell, \ell'$  are connected in the graph  $\mathcal{G}r := \langle \text{Ch}(k), \{(i, j), i, j \in T \text{ with } T \in \mathcal{T}'_k\} \rangle$  where  $\mathcal{T}'_k$  is the set of coalitions appearing in  $\mathcal{G}'_k$ . By assumption H2 and the non degeneracy of the capacities (see Lemma 1 derived from (12)), we have  $|\mathcal{S}_k| \geq 2$  and there exist  $S_1, S_2 \in \mathcal{S}_k$  such that  $\ell, \ell' \in S_1 \cup S_2$  and  $S_1 \neq S_2$ . The CI  $C_{\mu_k}(\mathbf{a}_{\text{Ch}(k)})$  at node  $k$  takes the form

$$m_k(S_1) \min_{i \in S_1} a_i + m_k(S_2) \min_{i \in S_2} a_i + \dots \quad (18)$$

Two cases (the other cases are obtained by symmetry):

- 1st case:  $\ell \in S_1, \ell' \in S_2$  and  $\ell \notin S_2$ . Hence by an appropriate ordering of the utilities, (18) can take the form  $m_k(S_1) a_\ell + m_k(S_2) a_{\ell'} + \dots$ , so that  $\ell$  and  $\ell'$  are directly connected in  $\mathcal{G}r$ .
- 2nd case:  $\{\ell, \ell'\} \subseteq S_1$  and  $\exists \ell'' \in S_2$  with  $\ell'' \notin S_1$ . Hence by an appropriate ordering of the utilities, (18) can take the two forms  $m_k(S_1) a_\ell + m_k(S_2) a_{\ell''} + \dots$  and

$m_k(S_1) a_{\ell'} + m_k(S_2) a_{\ell''} + \dots$ . Therefore  $\ell$  and  $\ell'$  are directly connected in  $\mathcal{G}r$  through  $\ell''$ .

Replacing in the linear equation w.r.t. marginal utilities of  $\mathcal{G}'_k$ , terms  $a_\ell$  ( $\ell \in \text{Ch}(k)$ ) by any element of  $\mathcal{G}_\ell$ , we obtain  $\mathcal{G}_k$  and we easily see, by the induction assumption, that any pair  $\ell, \ell' \in \text{Lf}(k)$  is connected in  $\mathcal{G}_k$ .  $\square$

**Proof of Lemma 7:** Consider a UHCI model  $\mathcal{F} = \langle \mathcal{T}, \nu, \mathcal{U} \rangle$ , and let us apply Algorithms 1 and 2. Let us show that the hierarchy obtained by  $\text{Alg}_2(\text{Alg}_1(\mathcal{F}))$  is exactly  $\mathcal{T}$ . More precisely, we show by induction starting from root down to the leaves that Algorithm 2 progressively produces hierarchy  $\mathcal{T}$ . We use below the sets  $\mathcal{T}_\ell$  as defined in Eq. (17).

At the root node  $k = r$ ,  $\mathcal{H}_k$  is composed of the separation frontiers  $\sum_{\ell \in T} w_\ell^T a_\ell = \sum_{\ell \in T'} w_\ell^{T'} a_\ell$  for every  $T \in \mathcal{T}_\ell$  and  $T' \in \mathcal{T}_{\ell'}$ , where  $\{\ell, \ell'\} \in \mathcal{S}_k$ . In Algorithm 2, such a pair  $(T, T')$  is an element of  $\mathcal{K}$ . The leaves of two separate subtrees  $\text{Lf}(\ell)$  and  $\text{Lf}(\ell')$  are clearly separated.

Let  $\ell \in \text{Ch}(k)$ . By H1,  $\ell$  is necessarily connected to another node  $\ell'$  through  $\mathcal{S}_k$ . Hence, a separation frontier at node  $k$  takes the form  $a_\ell = a_{\ell'}$ . By Lemma 9, all leaves are connected through  $\langle \text{Lf}(k), \{(i, j), i, j \in T \text{ with } T \in \mathcal{T}_k\} \rangle$ . Therefore all leaves of a child  $\ell \in \text{Ch}(k)$  form a connected component of  $\langle \text{Lf}(k), \mathcal{R} \rangle$ . Hence we obtain the right grouping of the elementary criteria at the top level.

We finally reproduce the previous reasoning recursively on any aggregation node  $k$ .  $\square$

**Proof of Theorem 2:** Consider two UHCI models  $\mathcal{F} = \langle \mathcal{T}, \nu, \mathcal{U} \rangle$  and  $\mathcal{F}' = \langle \mathcal{T}', \nu', \mathcal{U}' \rangle$  that both satisfy assumptions H1, H2, (9), (10), (11), (12), (13) and (15). Assume that these models are different and in particular  $\mathcal{T} \neq \mathcal{T}'$ .

Assume by contradiction that these two UHCI models yield exactly the same overall utility model. By Lemma 6, Algorithm 1 to  $\mathcal{F}$  and  $\mathcal{F}'$  yields the piecewise  $C^1$  functions and separating frontiers underlying  $\mathcal{F}$  and  $\mathcal{F}'$ . By corollary 1 of Lemma 6, set  $\mathcal{H}$  obtained by Algorithm 1 is identical. That is  $\text{Alg}_1(\mathcal{T}, \nu, \mathcal{U}) = \text{Alg}_1(\mathcal{T}', \nu', \mathcal{U}') = \mathcal{H}$ .

This equality yields that  $\text{Alg}_2(\text{Alg}_1(\mathcal{T}, \nu, \mathcal{U})) = \text{Alg}_2(\text{Alg}_1(\mathcal{T}', \nu', \mathcal{U}')) = \mathcal{T}''$ . Lemma 7 says that  $\mathcal{T} = \mathcal{T}''$  and  $\mathcal{T}' = \mathcal{T}''$ , as both  $\mathcal{F}$  and  $\mathcal{F}'$  satisfy H1, H2, (9), (10), (11), (12), (13) and (15). We conclude that  $\mathcal{T}$  and  $\mathcal{T}'$  are identical. We raise a contradiction.  $\square$

**Proof of Theorem 3:** Consider two UHCI models  $\mathcal{F} = \langle \mathcal{T}, \nu, \mathcal{U} \rangle$  and  $\mathcal{F}' = \langle \mathcal{T}', \nu', \mathcal{U}' \rangle$  that both satisfy assumptions H1, H2, (9), (10), (11), (12), (13) and (15).

By Theorem 2, we have  $\mathcal{T} = \mathcal{T}'$ . As we are now in a setting where both UHCIs have the same hierarchy, we can apply Theorem 1, which gives us that  $\nu = \nu'$  and  $\mathcal{U} = \mathcal{U}'$ , concluding the proof.  $\square$

## References

Angilella, S.; Corrente, S.; Greco, S.; and Slowinski, R. 2013. Multiple criteria hierarchy process for the Choquet integral. In *7th International Conference on Evolutionary Multi-Criterion Optimization*, 475–489.

- Bishop, C. M. 2006. *Pattern recognition and machine learning*. Information science and statistics. Springer.
- Bouyssou, D., and Marchant, T. 2010. Additive conjoint measurement with ordered categories. *Eur. J. of Operational Research* 203:194–204.
- Bresson, R.; Cohen, J.; Hüllermeier, E.; Labreuche, C.; and Sebag, M. 2020a. Evaluating the stability of the neur-hci framework. In *CAID 2020 - Second Conference on Artificial Intelligence for Defence*, 120–126.
- Bresson, R.; Cohen, J.; Hüllermeier, E.; Labreuche, C.; and Sebag, M. 2020b. Neural representation and learning of hierarchical 2-additive choquet integrals. In Bessiere, C., ed., *IJCAI 2020*, 1984–1991.
- Choquet, G. 1953. Theory of capacities. *Annales de l'Institut Fourier* 5:131–295.
- Fallah Tehrani, A.; Cheng, W.; Dembczyński, K.; and Hüllermeier, E. 2012. Learning monotone nonlinear models using the choquet integral. 89(1):183–211.
- Glorot, X.; Bordes, A.; and Bengio, Y. 2011. Deep sparse rectifier neural networks. In Gordon, G.; Dunson, D.; and Dudík, M., eds., *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research*, 315–323. Fort Lauderdale, FL, USA: JMLR Workshop and Conference Proceedings.
- Grabisch, M., and Labreuche, C. 2010. A decade of application of the Choquet and Sugeno integrals in multi-criteria decision aid. *Annals of Operation Research* 175:247–286.
- Grabisch, M., and Perny, P. 2003. Agrégation multicritère. In *Logique floue, principes, aide à la décision*, Traité IC2, série Informatique et Systèmes d'Information. Hermès-Lavoisier. 81–120.
- Grabisch, M., and Roubens, M. 1999. An axiomatic approach to the concept of interaction among players in cooperative games. *Int. Journal of Game Theory* 28:547–565.
- Grabisch, M.; Marichal, J.-L.; Mesiar, R.; and Pap, E. 2009. *Aggregation Functions*. Encyclopedia of Mathematics and its Applications. Cambridge University Press.
- Grabisch, M.; Kojadinovic, I.; and Meyer, P. 2008. A review of methods for capacity identification in choquet integral based multi-attribute utility theory. *European Journal of Operational Research* 186(2):766–785.
- Grabisch, M. 1997.  $k$ -order additive discrete fuzzy measures and their representation. *Fuzzy Sets & Systems* 92:167–189.
- Havens, T. C., and Anderson, D. T. 2019. Machine learning of choquet integral regression with respect to a bounded capacity (or non-monotonic fuzzy measure). In *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 1–6.
- Keeney, R. L., and Raiffa, H. 1976. *Decision with Multiple Objectives*. New York: Wiley.
- Krantz, D.; Luce, R.; Suppes, P.; and Tversky, A. 1971. *Foundations of measurement*, volume 1: Additive and Polynomial Representations. Academic Press.
- Labreuche, C., and Fossier, S. 2018. Explaining multi-criteria decision aiding models with an extended shapley value. In *IJCAI 2018*, 331–339.
- Labreuche, C. 2018. An axiomatization of the Choquet integral and its utility functions without any commensurability assumption. *Annals of Operation Research* 271(2):701–735.
- Miller, G. A. 1956. The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychological review* 63 2:81–97.
- Murofushi, T., and Soneda, S. 1993. Techniques for reading fuzzy measures (III): interaction index. In *9th Fuzzy System Symposium*, 693–696.
- Ogle, K., and Barber, J. J. 2020. Ensuring identifiability in hierarchical mixed effects bayesian models. *Ecological Applications* 30(7):e02159.
- O'Neill, C. 2016. *Weapons of Math Destruction*. Crown Books.
- Ovchinnikov, S. 2002. Max-min representation of piecewise linear functions. *Beiträge zur Algebra und Geometrie* 43.
- Paulino, C., and Pereira, C. 1994. On identifiability of parametric statistical models. *Journal of the Italian Statistical Society* 3:125–151.
- Pelissari, R., and Duarte, L. T. 2020. Identification of choquet capacity in multicriteria sorting problems through stochastic inverse analysis. *arXiv:2003.12530 [cs, stat]*. arXiv: 2003.12530.
- Ran, Z.-Y., and Hu, B.-G. 2017. Parameter identifiability in statistical machine learning: A review. *Neural Computation* 29(5):1151–1203.
- Rothenberg, T. J. 1971. Identification in parametric models. *Econometrica* 39(3):pp. 577–591.
- Shapley, L. S. 1953. A value for  $n$ -person games. In Kuhn, H. W., and Tucker, A. W., eds., *Contributions to the Theory of Games, Vol. II*, number 28 in *Annals of Mathematics Studies*. Princeton University Press. 307–317.
- Xie, Y., and Carlin, B. P. 2006. Measures of bayesian learning and identifiability in hierarchical models. *Journal of Statistical Planning and Inference* 136(10):3458–3477.