APPLICATION OF THE CHOQUET INTEGRAL TO SUBJECTIVE MENTAL WORKLOAD EVALUATION

M. Grabisch[1] **H. Prade**[2] **É. Raufaste**[3] **P. Terrier**[3]

[1] *Université Paris I Panthéon-Sorbonne ; LIP6*
[2] *Université Toulouse-III Paul Sabatier, IRIT*
[3] *Université Toulouse-II Le Mirail, LTC*

Abstract: We describe a methodology based on Choquet integral to build general models of subjective evaluation. The model is compared to the NASA-TLX, which is of current use in assessment of workload, in an experiment based on two classical computer games.

Keywords: subjective evaluation, Choquet integral, importance index, interaction, workload assessment, NASA-TLX

1 Introduction: the NASA-TLX method

A very important topic for ergonomics and human factors is measuring the mental workload associated with the situations under study. For example, if one is to compare two human-computer interfaces, the device that produces a lower level of subjective mental workload is generally preferred. Among the most widely used methods are the National Aeronautics and Space Administration-Task Load Index (NASA-TLX, Hart and Staveland, 1988) and the Subjective Workload Assessment Technique (SWAT, Reid and Nygren, 1988). Contrary to the Cooper-Harper scale (Cooper and Harper, 1969), where operators are asked a single workload estimate, the NASA-TLX and SWAT both assume that workload is a multidimensional concept, with six and three workload sources respectively. Supposedly, a multidimensional approach provides a richer and less biased picture of workload. In the general case, a multidimensional approach makes difficult to directly compare various settings (work situations, interfaces, and so on).

The NASA-TLX rating procedure provides an overall workload score based on a weighted average of the ratings on six subscales: Mental Demands, Physical Demands, Temporal Demands, Performance, Effort, and Frustration. Depending on situations, the various sources may differently contribute to the operator's subjective workload. Taking into account the relative weights of the sources first requires obtaining a measure of their relative importance. For example, during the standard NASA-TLX procedure participants provide the 15 possible pair-wise comparisons of the six subscales. In each comparison, subjects select the source that contributed to the workload more than the other. Each source receives one point for each comparison where it was deemed to contribute more. The relative weight of a source is then given by the sum of those points, divided by 15 for normalization purposes. In order to avoid confusion, in this paper we will call "rating" the value provided for each source, and "weight" the relative importance of a source. The "score" will denote the global score provided by an aggregation method.

After information about ratings and weights is collected, the question is to choose the aggregation method. The NASA-TLX makes use of a classical weighted mean, which simply sums the products of ratings by their normalized weights (Σw_i = 1). Thus, noting x_i the rating about the i^{th} source and a_i the relative importance of the same source, the subjective workload SW in the NASA-TLX method is provided by

$$SW = \sum_{i=1}^{6} w_i a_i$$

where w_i and a_i respectively denote the weight and rating associated with the i^{th} workload source.

Although some previous studies applied fuzzy sets theory to workload measurement by means of linguistic terms (e.g., Chen, Jung, and Peacock, 1994; Liou and Wang, 1994), another question is addressed here: what is the best way of aggregating data about the six NASA-TLX workload sources into a single workload value that enables direct comparison of different settings?

Is weighted average the best model of aggregation? Indeed, it is easy to compute and familiar to most researchers. On the other hand, despite its apparent simplicity, the weighted average model is built upon several strong mathematical assumptions that are not necessarily verified in workload assessment. For example, it requires independence between ratings and weights. This condition could be attained, for example, by having operators providing the ratings and external experts providing the weights. Unfortunately, in the standard NASA-TLX procedure, each operator provides both the ratings and weights. Second, weighted average does not allow taking into account interactions between sources. Is it a reasonable choice to neglect dependencies and interactions between sources of workload? By neglecting such interaction effects, a weighted average model might induce measurement biases.

As soon as a measure can be used as a learning criterion, the Choquet integral provides a potential solution to those problems because it enables computing weights from an adjustment criterion in a mathematically sound manner.

2 The Choquet integral for multi-criteria decision making

(For a detailed presentation, see Grabisch, Duchêne, Lino, and Perny, 2002; Grabisch, 2003; Grabisch and Labreuche, 2004). We present here the necessary material for introducing our model based on Choquet integral. Let $N = \{1, \dots, n\}$ be the index set of criteria.

Definition 1 *A* capacity (Choquet, 1953) *or* fuzzy measure (Sugeno, 1974) *μ on N is a function $\mu : \mathcal{P}(N) \longrightarrow [0,1]$, satisfying the following axioms.*

(i) $\mu(\emptyset) = 0$.

(ii) $A \subseteq B \subseteq N$ *implies* $\mu(A) \leq \mu(B)$.

We will assume in addition $\mu(N) = 1$ (normalized capacity). For any $A \subseteq N$, $\mu(A)$ represents the importance of the coalition A of criteria for making decision. The capacity is *additive* if it is a probability measure.

Definition 2 *Let μ be a capacity on N. The* Choquet integral *of a function $f : N \longrightarrow \mathbb{R}_+$ with respect to μ is defined by*

$$\mathcal{C}_\mu(f) := \sum_{i=1}^{n} (f_{(i)} - f_{(i-1)})\mu(A_{(i)}), \quad (1)$$

where we have written for simplicity $f_i := f(i)$, and $\cdot_{(i)}$ indicates that the indices have been permuted so that $0 \leq f_{(1)} \leq \dots \leq f_{(n)}$, and $A_{(i)} := \{(i), \dots, (n)\}$, and $f_{(0)} = 0$.

The Choquet integral of f, considered as an alternative to be evaluated, is the overall score of the alternative considering the importance of coalitions of criteria. An important property of the Choquet integral is that

$$\mathcal{C}_\mu(1_A) = \mu(A)$$

where 1_A is the characteristic function of A, $A \subseteq N$. This gives a clear interpretation of the quantity $\mu(A)$. Another property worth to be mentionned is that when μ is additive, then the Choquet integral reduces to a weighted average $\sum_i w_i f_i$, with $w_i = \mu(\{i\})$.

Since the definition of μ involves 2^n values, which may cause some interpretation problem in terms of the importance of criteria, a convenient concept is the one of Shapley index (Shapley, 1953), coming from cooperative game theory. For any criterion $i \in N$, the Shapley index of i is defined by:

$$\phi_i := \sum_{K \subset N \setminus i} \frac{(n - |K| - 1)!|K|!}{n!} [\mu(K \cup \{i\}) - \mu(K)]. \quad (2)$$

Roughly speaking, the Shapley index ϕ_i computes the average contribution of criterion i in all coalitions, the average being weighted by a coefficient taking into account the cardinality of the coalition. In this sense, it can be taken as definition of the *average importance* or *average contribution* of a single criterion for the decision process. The Shapley index satisfies $\sum_{i=1}^{n} \phi_i = \mu(N) = 1$, so that the sum of importance degrees is a constant. The idea to use the Shapley index for multicriteria decision making is due to Murofushi (1992).

Another important topic is the notion of interaction among two criteria, as proposed originally by Murofushi and Soneda (1993).

$$I_{ij} := \sum_{K \subset N \setminus \{i,j\}} \frac{(n - |K| - 2)!|K|!}{(n-1)!} [\mu(K \cup \{i,j\}) - \mu(K \cup \{i\}) - \mu(K \cup \{j\}) + \mu(K)]. \quad (3)$$

A positive interaction I_{ij} occurs whenever criteria i, j are *complementary*, i.e., the satisfaction of both is necessary to get overall satisfaction (the score of i and j are aggregated conjunctively). On the contrary, if it is sufficient to satisfy only i or j, then i and j are *substitutive*, and $I_{ij} < 0$ (the score of i and j are aggregated disjunctively).

Later, Grabisch has generalized this notion to any number of criteria, leading to the following definition of the interaction index (Grabisch, 1997), defined for all coalitions (including the empty one), which is, $\forall A \subseteq N$:

$$I(A) := \sum_{K \subset N \setminus A} \frac{(n - k - a)!k!}{(n - a + 1)!} \sum_{B \subset A} (-1)^{|A| - |B|} \mu(K \cup B), \quad (4)$$

where $k := |K|$ and $a := |A|$. Note that $I(\{i\}) = \phi_i$, and $I(\{i,j\}) = I_{ij}$. Also, it is easy to show that for an additive measure, $I(A) = 0$ whenever $|A| > 1$, and $\phi_i = \mu(\{i\})$. It is interesting to note that giving $I(A)$ for all $A \subseteq N$ allows for recovery of the capacity μ: I is merely another representation of μ (for details, see Grabisch, 1997).

Definition 3 *A capacity μ is said to be k-*additive *if its interaction transform satisfies $I(A) = 0$ for any A*