

# On a pairwise comparison-based consistent non-numerical ranking

RYSZARD JANICKI, *Department of Computing and Software, McMaster University, Hamilton, Ontario, Canada L8S 4K1.*  
E-mail: [janicki@mcmaster.ca](mailto:janicki@mcmaster.ca)

YUN ZHAI, *Artificial Intelligence in Medicine Inc., Toronto, Ontario, Canada, M5A 2W3.*  
E-mail: [zhaiy@mcmaster.ca](mailto:zhaiy@mcmaster.ca)

## Abstract

We discuss a consistent model of pairwise comparison-based non-numerical ranking. An algorithm that enforces consistency for raw or partially organized ranking data is presented and its properties are analysed. The concept of testing subjective rankings is also discussed.

**Keywords:** Non-numerical ranking, pairwise comparisons, consistency, partial orders.

## 1 Introduction

A *ranking* or *preference* is usually defined as a weakly ordered relationship between a set of items such that, for any two items the first is either ‘less preferred’, ‘more preferred’ or ‘indifferent’ to the second one [10]. The ranking is numerical if numbers are used to measure importance and to create the ranking relation. Numerical rankings are usually totally ordered. Various kinds of *global indexes* are popular examples of numerical rankings.

The *pairwise comparisons* method is based on the observation that it is much easier to rank the importance of *two* objects than it is to rank the importance of *several* objects [3]. The problem is then reduced to constructing a global ranking from the set of partially ordered pairs. The method can be traced to the 1785 Marquis de Condorcet paper [1, 5], and was explicitly mentioned and analysed by Fechner in 1860 [7], made popular by Thurstone in 1927 [23] and was transformed into a kind of semi-formal methodology by Saaty in 1977 (called AHP, Analytic Hierarchy Process, see [6, 10, 21]).

At present, pairwise comparisons are practically identified with Saaty’s controversial AHP. On one hand, AHP has respected practical applications, on the other it is still considered by many as a flawed procedure that produces arbitrary rankings. For more details, the reader is referred to references [6, 11, 16].

Pairwise Comparison-based *non-numerical* solutions were proposed and discussed in references [11, 13–15]. The model presented in this article stems from [11] and [13] and can be applied to improve the quality of decision making in eHealth systems [4, 18].

The model presented below uses no numbers and is entirely based on the concept of partial orders.

Non-numerical rankings should be *weak* orders and sometimes *total* orders, but the initial empirical data may not even be a partial order, in general they are just arbitrary relations. This motivates asking

what is the ‘best’ partial order approximation of an arbitrary relation, and what is the ‘best’ weak order approximation of an arbitrary partial order? The latter problem is discussed in detail in [8, 9], while four different solutions to the former were proposed and analysed in [12, 13]. In this article, we will use the approximation denoted  $(R^+)^\bullet$  (calculate the transitive closure and then remove all cycles), which was first proposed by Schröder in 1895 [22]. Parts of this article are based on Yun Zhai’s PhD thesis [24].

## 2 Relations and partial orders

In this section, we recall some well-known concepts and results that will be used in the following sections (cf. [8, 20]).

Let  $X$  be a finite set, fixed for the rest of this article. For every relation  $R \subseteq X \times X$ , let  $R^+ = \bigcup_{i=1}^{\infty} R^i$ , denote the *transitive closure* of  $R$ .

A relation  $< \in X \times X$  is a (*sharp*) *partial order* if it is irreflexive and transitive, i.e. if  $\neg(a < a)$  and  $a < b < c \implies a < c$ , for all  $a, b, c \in X$ .

We write  $a \sim_{<} b$  if  $\neg(a < b) \wedge \neg(b < a)$ , i.e. if  $a$  and  $b$  are either *distinctly incomparable* (w.r.t.  $<$ ) or *identical* elements.

A partial order is

- *total* or *linear*, if  $\sim_{<}$  is empty, i.e. for all  $a, b \in X$ .  $a \neq b \implies (a < b \vee b < a)$ .
- *weak* or *stratified*, if  $a \sim_{<} b \sim_{<} c \implies a \sim_{<} c$ , i.e. if  $\sim_{<}$  is an equivalence relation (i.e. it is reflexive, symmetric and transitive).

A relation  $R$  is *acyclic* if and only if  $\neg(xR^+x)$  for all  $x \in X$ .

For every relation  $R$ , define the relations  $R^{\text{cyc}}$  and  $R^\bullet$  as

$$aR^{\text{cyc}}b \iff aR^+b \wedge bR^+a \quad (2.1)$$

$$aR^\bullet b \iff aRb \wedge \neg(aR^{\text{cyc}}b) \quad (2.2)$$

We will call  $R^\bullet$  the *acyclic refinement* of  $R$ .

COROLLARY 2.1

If  $R$  is a partial order then  $R = R^+ = R^\bullet$ . ■

The relation  $(R^+)^\bullet$  can be treated as a partial order approximation of  $R$  (often ‘the best’ approximation, see [12, 13] for detailed definitions and proofs).

LEMMA 2.2

- (1)  $(R^+)^\bullet$  is a partial order (Schröder 1895 [22]).
- (2)  $(R^+)^\bullet$  is a partial order approximation of  $R$  (see [12, 13]). ■

Approximations of partial orders by weak orders are just proper extensions. Various methods were proposed and discussed in [8] and specially in [9]. For our purposes, the best seems to be the method based on the concept of a *global score function* [8], which is defined as (for every finite set  $X$ ,  $\|X\|$  denotes its number of elements):

$$g_{<}(x) = \|\{z \mid z < x\}\| - \|\{z \mid x < z\}\| \quad (2.3)$$

Given the global score function  $g_{<}(x)$ , we define the relation  $<^w \subseteq X \times X$  as

$$a <^w b \iff g_{<}(a) < g_{<}(b) \quad (2.4)$$

We will use this technique in our model.

### 3 Consistency-driven non-numerical ranking: the model

A *pairwise comparisons ranking data* [13] is a tuple  $\mathcal{R}' = (X, R'_0, R'_1, \dots, R'_k)$ , where  $X$  is the set of objects to be ranked,  $k \geq 1$ , and  $R'_i$ 's are relations satisfying  $R'_0 \cup R'_1 \cup \dots \cup R'_k \cup R'_1^{-1} \cup \dots \cup R'_k^{-1} = X \times X$  and  $R'_i \cap R'_j = \emptyset$  unless  $i = j$ . The relation  $R'_0$ , interpreted as *indifference*, is symmetric and reflexive, while the relations  $R'_1, \dots, R'_k$ , interpreted as *preferences*, are asymmetric and irreflexive.

The relations  $(R'_0, R'_1, \dots, R'_k)$  are based on empirical data or judgements, so no other specific properties are expected.

A tuple  $\mathcal{R} = (X, R_0, R_1, \dots, R_k)$ , is called a *pairwise comparison consistent ranking system* when some additional *consistency* properties are satisfied, and is called *derived from*  $\mathcal{R}' = (X, R'_0, R'_1, \dots, R'_k)$ , if each  $R_i$  is some *approximation* of  $R'_i$ .

We will devote the rest of this section to this problem. Quite often we will use the same symbol to denote both  $R_i$  and  $R'_i$ , our algorithm presented later will take  $\mathcal{R}'$ , and produce  $\mathcal{R}$ .

In [11], the case  $\mathcal{R} = (X, \approx, \sqsubset, \subset, <, \prec)$ , with the following interpretation  $a \approx b$ :  $a$  and  $b$  are *indifferent*,  $a \sqsubset b$ : *slightly in favour of*  $b$ ,  $a \subset b$ : *in favour of*  $b$ ,  $a < b$ :  $b$  is *strongly better*,  $a \prec b$ :  $b$  is *extremely better*, was proposed and some (incomplete) axioms were proposed. For all practical applications, the list  $\sqsubset, \subset, <, \prec$  may be shorter or longer, but not empty and not much longer (due to limitations of the human mind [2, 19]).

In this article, we consider only  $\mathcal{R} = (X, \approx, \sqsubset, \subset, <, \prec)$ , leaving the generalizations and special cases to the reader (see [13] for more comments on this subject).

Let  $X$  be a finite set of objects to be 'ranked', and let  $\approx, \sqsubset, \subset, <$  and  $\prec$  be a family of disjoint relations on  $X$  such that  $X \times X = \approx \cup \sqsubset \cup \subset \cup < \cup \prec \cup \sqsubset^{-1} \cup \subset^{-1} \cup <^{-1} \cup \prec^{-1}$ .

We define the relations  $\widehat{\sqsubset}, \widehat{\subset}, \widehat{<}$  and  $\widehat{\prec}$  as follows:

$$\begin{aligned} \widehat{<} &= < & \widehat{\prec} &= < \cup \prec \\ \widehat{\subset} &= \subset \cup < \cup \subset & \widehat{\sqsubset} &= \sqsubset \cup < \cup \subset \cup \sqsubset \end{aligned}$$

The relations  $\widehat{\sqsubset}, \widehat{\subset}, \widehat{<}$  and  $\widehat{\prec}$  are interpreted as *combined preferences*, i.e.  $a \widehat{\sqsubset} b$ : *at least slightly in favour of*  $b$ ,  $a \widehat{\subset} b$ : *at least in favour of*  $b$ ,  $a \widehat{<} b$ : *at least strongly in favour of*  $b$  and  $a \widehat{\prec} b$ : *at least  $b$  is far superior than*  $a$ .

#### DEFINITION 3.1

A tuple  $\mathcal{R} = (X, \approx, \sqsubset, \subset, <, \prec)$  is a *pairwise comparison ranking system* if it is a pairwise comparison ranking data and the relations  $\widehat{\sqsubset}, \widehat{\subset}, \widehat{<}, \widehat{\prec}$  are partial orders. ■

#### DEFINITION 3.2

The tuple  $\mathcal{R} = (X, \approx, \sqsubset, \subset, <, \prec)$  is a *pairwise comparison consistent ranking system* if it is a pairwise comparison ranking data and the following rules (called *consistency rules*) are also satisfied:

1.  $(a \approx b \wedge b \approx c) \Rightarrow (a \approx c \vee a \sqsubset c \vee c \sqsubset a)$
- 2.1.  $(a \approx b \wedge b \sqsubset c) \vee (a \sqsubset b \wedge b \approx c) \Rightarrow (a \sqsubset c \vee a \subset c)$
- 2.2.  $(a \approx b \wedge b \subset c) \vee (a \subset b \wedge b \approx c) \Rightarrow (a \subset c \vee a < c)$
- 2.3.  $(a \approx b \wedge b < c) \vee (a < b \wedge b \approx c) \Rightarrow (a < c \vee a \prec c)$
3.  $(a \approx b \wedge b \prec c) \vee (a \prec b \wedge b \approx c) \Rightarrow (a \prec c)$
- 4.1.  $(a \sqsubset b \wedge b \sqsubset c) \Rightarrow (a \sqsubset c \vee a \subset c)$
- 4.2.  $(a \subset b \wedge b \sqsubset c) \vee (a \sqsubset b \wedge b \subset c) \Rightarrow (a \subset c \vee a < c)$
- 4.3.  $(a < b \wedge b \sqsubset c) \vee (a \sqsubset b \wedge b < c) \Rightarrow (a < c \vee a \prec c)$
- 5.1.  $(a \sqsubset b \wedge b < c) \vee (a < b \wedge b \sqsubset c) \Rightarrow (a < c)$
- 5.2.  $(a \subset b \wedge b < c) \vee (a < b \wedge b \subset c) \Rightarrow (a < c)$

## 4 Pairwise comparison-based consistent non-numerical ranking

- 5.3.  $(a < b \wedge b < c) \vee (a < b \wedge b < c) \Rightarrow (a < c)$   
 5.4.  $(a < b \wedge b < c) \vee (a < b \wedge b < c) \Rightarrow (a < c)$   
 6.1.  $(a \sqsubset b \wedge b < c) \vee (a < b \wedge b \sqsubset c) \Rightarrow (a < c)$   
 6.2.  $(a < b \wedge b < c) \Rightarrow (a < c)$   
 7.  $(a \sqsubset b \wedge b \sqsubset c) \Rightarrow (a < c) \vee (a < c)$   
 8.1.  $(a \approx b \wedge b \sqsubset c) \vee (a \sqsubset b \wedge b \approx c) \Rightarrow (a \sqsubset c) \vee (a \sqsupset c)$   
 8.2.  $(a \approx b \wedge b \sqsupset c) \vee (a \sqsupset b \wedge b \approx c) \Rightarrow (a \sqsupset c) \vee (a > c)$   
 8.3.  $(a \approx b \wedge b > c) \vee (a > b \wedge b \approx c) \Rightarrow (a > c) \vee (a > c)$   
 8.4.  $(a \approx b \wedge b > c) \vee (a > b \wedge b \approx c) \Rightarrow (a > c)$   
 9.1.  $(a \sqsupset b \wedge b \sqsubset c) \Rightarrow (a \sqsubset c) \vee (a \sqsupset c)$   
 9.2.  $(a \sqsupset b \wedge b \sqsupset c) \vee (a \sqsupset b \wedge b \sqsubset c) \Rightarrow (a \sqsupset c) \vee (a > c)$   
 9.3.  $(a \sqsupset b \wedge b > c) \vee (a > b \wedge b \sqsubset c) \Rightarrow (a > c) \vee (a > c)$   
 9.4.  $(a \sqsupset b \wedge b > c) \vee (a > b \wedge b \sqsupset c) \Rightarrow (a > c)$   
 10.1.  $(a \sqsupset b \wedge b \sqsupset c) \Rightarrow (a > c) \vee (a > c)$   
 10.2.  $(a \sqsupset b \wedge b > c) \Rightarrow (a > c)$   
 10.3.  $(a \sqsupset b \wedge b > c) \Rightarrow (a > c)$   
 11.1.  $(a > b \wedge b > c) \Rightarrow (a > c)$   
 11.2.  $(a > b \wedge b > c) \vee (a > b \wedge b > c) \Rightarrow (a > c)$   
 12.  $(a > b \wedge b > c) \Rightarrow (a > c)$   
 13.1.  $(a \sqsubset b \wedge b \sqsubset c) \vee (a \sqsubset b \wedge b \sqsupset c) \Rightarrow (a \approx c \vee a \sqsubset c \vee a \sqsupset c)$   
 13.2.  $(a \sqsubset b \wedge b \sqsupset c) \vee (a \sqsubset b \wedge b \sqsupset c) \Rightarrow (a \sqsubset c \vee a \sqsubset c)$   
 13.3.  $(a \sqsubset b \wedge b < c) \vee (a < b \wedge b \sqsupset c) \Rightarrow (a \sqsubset c \vee a < c)$   
 13.4.  $(a \sqsubset b \wedge b < c) \vee (a < b \wedge b \sqsupset c) \Rightarrow (a < c \vee a < c)$   
 14.1.  $(a \sqsupset b \wedge b \sqsubset c) \vee (a \sqsupset b \wedge b \sqsupset c) \Rightarrow (a \sqsupset c \vee a \sqsupset c)$   
 14.2.  $(a \sqsupset b \wedge b \sqsupset c) \vee (a \sqsupset b \wedge b \sqsupset c) \Rightarrow (a \approx c \vee a \sqsubset c \vee a \sqsupset c)$   
 14.3.  $(a \sqsupset b \wedge b < c) \vee (a < b \wedge b \sqsupset c) \Rightarrow (a \sqsupset c \vee a \sqsubset c)$   
 14.4.  $(a \sqsupset b \wedge b < c) \vee (a < b \wedge b \sqsupset c) \Rightarrow (a \sqsubset c \vee a < c \vee a < c)$   
 15.1.  $(a > b \wedge b \sqsubset c) \vee (a \sqsupset b \wedge b > c) \Rightarrow (a \sqsupset c \vee a > c)$   
 15.2.  $(a > b \wedge b \sqsupset c) \vee (a \sqsupset b \wedge b > c) \Rightarrow (a \sqsupset c \vee a \sqsupset c)$   
 15.3.  $(a > b \wedge b < c) \vee (a < b \wedge b > c) \Rightarrow (a \approx c \vee a \sqsubset c \vee a \sqsupset c \vee a \sqsubset c \vee a \sqsupset c)$   
 15.4.  $(a > b \wedge b < c) \vee (a < b \wedge b > c) \Rightarrow (a \sqsubset c \vee a \sqsubset c \vee a < c \vee a < c)$   
 16.1.  $(a > b \wedge b \sqsubset c) \vee (a \sqsupset b \wedge b > c) \Rightarrow (a > c \vee a > c)$   
 16.2.  $(a > b \wedge b \sqsupset c) \vee (a \sqsupset b \wedge b > c) \Rightarrow (a \sqsupset c \vee a > c \vee a > c)$   
 16.3.  $(a > b \wedge b < c) \vee (a < b \wedge b > c) \Rightarrow (a \sqsupset c \vee a \sqsupset c \vee a > c \vee a > c)$   
 16.4.  $(a > b \wedge b < c) \vee (a < b \wedge b > c) \Rightarrow$   
 $(a \approx c \vee a \sqsupset c \vee a \sqsubset c \vee a \sqsupset c \vee a \sqsubset c \vee a > c \vee a < c)$  ■

The rules above implement the idea that composition may not change precedences in a drastic way, in most cases only by ‘one step’, or not at all. Consider for example the rule:

$$2.1. (a \approx b \wedge b \sqsubset c) \vee (a \sqsubset b \wedge b \approx c) \Rightarrow (a \sqsubset c \vee a \sqsubset c).$$

In principle it says that if  $a$  and  $b$  are indifferent and  $c$  is slightly better than  $b$ , then at most  $c$  is in favour of  $a$ , and similarly for a symmetric case. The rules 1–12 can be interpreted as an extension of transitivity, and the rules 13.1–16.4 show what can we expect when composing opposite preferences. The set of rules is complete in the sense that it covers all possible relationships between triples.

We will show that the rules 1–16.4 of Definition 3.2 imply partial ordering of the relations  $\hat{\sqsubset}$ ,  $\hat{\sqsupset}$ ,  $\hat{\approx}$ ,  $\hat{\succ}$ , so any *pairwise comparison consistent non-numerical ranking system* is also a valid *pairwise comparison ranking system* as defined by Definition 3.1.

## PROPOSITION 3.3

If  $\mathcal{R}=(X, \approx, \sqsubset, \sqsubset, <, <)$  is a pairwise comparison *consistent* ranking system, then the relations  $\widehat{\sqsubset}$ ,  $\widehat{\sqsubset}$ ,  $\widehat{\sqsubset}$ ,  $\widehat{\sqsubset}$  are partial orders.

PROOF. To show that  $\widehat{\sqsubset}$  is a partial order we need only to show that it is transitive. Let  $a\widehat{\sqsubset}b$  and  $b\widehat{\sqsubset}c$ . Since  $a\widehat{\sqsubset}b \iff a\sqsubset b \vee a\sqsubset b \vee a < b \vee a < b$  and  $b\widehat{\sqsubset}c \iff b\sqsubset c \vee b\sqsubset c \vee b < c \vee b < c$ , we have to consider  $4 \times 4 = 16$  cases. If  $a\sqsubset b$  and  $b\sqsubset c$ , then by the rule 4.1 we have  $a\sqsubset c$  or  $a\sqsubset c$ , i.e.  $a\widehat{\sqsubset}c$ . For the remaining cases, we proceed similarly. ■

Typically, we need that the finest ranking,  $\sqsubset$  is also a weak order (it is usually unreasonable to require the other orders to have any specific properties).

A ranking  $\mathcal{R}=(X, \approx, \sqsubset, \sqsubset, <, <)$  is *weakly ordered* if the relation  $\widehat{\sqsubset}$  is a weak order.<sup>1</sup>

As we mentioned above, the starting data  $\approx, \sqsubset, \sqsubset, <, <$  may not even be partial orders. The following algorithm has been proposed in [13] to obtain a Pairwise Comparisons Ranking System (that is not necessarily consistent)  $\mathcal{R}=(X, \approx, \sqsubset, \sqsubset, <, <)$  from a Pairwise Comparisons Ranking Data  $\mathcal{R}'=(X, R_0, R_1, R_2, R_3, R_4)$ :

Algorithm 1 (1) Calculate  $\widehat{R}_4=R_4$ ,  $\widehat{R}_3=R_3 \cup \widehat{R}_4$ ,  $\widehat{R}_2=R_2 \cup \widehat{R}_3$ ,  $\widehat{R}_1=R_1 \cup \widehat{R}_2$ .  
 (2) Calculate  $\widehat{\sqsubset}=(\widehat{R}_4)^+ \bullet$ .  
 (3) Calculate  $\widehat{\sqsubset}=(\widehat{R}_3 \cup \widehat{\sqsubset})^+ \bullet$ .  
 (4) Calculate  $\widehat{\sqsubset}=(\widehat{R}_2 \cup \widehat{\sqsubset})^+ \bullet$ .  
 (5) Calculate  $\widehat{\sqsubset}=(\widehat{R}_1 \cup \widehat{\sqsubset})^+ \bullet$ .  
 (6) If we want a *weakly ordered* ranking, then calculate  $\widehat{\sqsubset}^w$  using the global score function  $g_{\widehat{\sqsubset}}(x)$  described in the previous section, and set  $\widehat{\sqsubset}=\widehat{\sqsubset}^w$ . Otherwise there is no change.  
 (7) Calculate  $\approx=(X \times X) \setminus (\widehat{\sqsubset} \cup \widehat{\sqsubset}^{-1})$ .  
 (8) Calculate  $<=\widehat{\sqsubset}$ ,  $<=\widehat{\sqsubset} \setminus \widehat{\sqsubset}$ ,  $\sqsubset=\widehat{\sqsubset} \setminus \widehat{\sqsubset}$ ,  $\sqsubset=\widehat{\sqsubset} \setminus \widehat{\sqsubset}$ . ■

## PROPOSITION 3.4

The time complexity of Algorithm 1 is  $O(n^3)$ , where  $n$  is the size of  $X$ .

PROOF. The operation ‘+’ is  $O(n^3)$ . The others, including calculating global scores, are  $O(n^2)$ . ■

It was also shown in [12, 13] that the technique used satisfies the requirements for computing ‘partial order approximation’. Nevertheless, Algorithm 1 does *not* guarantee consistency.

## 4 Enforcing consistency

Algorithm 1 from the previous section gives a ranking system that may not be consistent. The following algorithm starts *either* with raw pairwise comparison ranking data *or* with a pairwise comparison ranking system and produces a pairwise comparison *consistent* ranking system.

Algorithm 2 (1) Find all pairs that violate consistency rules, if none, proceed to 5.  
 (2) Pick the pairs which violate the biggest number of rules.  
 (3) Revise the relation between the pairs violating a rule, by appropriately *lowering preferences*, e.g. from  $<$  to  $\sqsubset$ , etc. (in general from  $R_i$  to  $R_{i-1}$ , in the other notation)  
 (4) proceed to 1.  
 (5) If we want a *weakly ordered* consistent ranking, calculate  $\widehat{\sqsubset}^w$  using the global score function  $g_{\widehat{\sqsubset}}(x)$ , and set  $\approx=(X \times X) \setminus (\widehat{\sqsubset}^w \cup (\widehat{\sqsubset}^w)^{-1})$ ,  $\sqsubset=\widehat{\sqsubset}^w \setminus (\widehat{\sqsubset}^w \cup \widehat{\sqsubset}^w)$ . Otherwise stop. ■

<sup>1</sup>While extending partial orders to weak orders does make sense for the purpose of ranking, an extension to total orders does *not* (see [13] for details).

PROPOSITION 4.1

- (1) Algorithm 2 always converges.
- (2) Algorithm 2 always produces a pairwise comparison *consistent* ranking system.
- (3) The time complexity of Algorithm 2 is  $O(n^3)$ .
- (4) In the worst case the outcome is  $\approx = X \times X$ .

PROOF. (1) and (4) From step 3 of the algorithm, we always ‘increase the disorder’. In the worst case, we may get that  $\approx = X \times X$ , but the procedure always stops.

(2) We need to show that the step 5 does not introduce new rule violations. But since  $\widehat{\sqsubset}^w$  is an extension of  $\widehat{\sqsubset}$ , i.e.  $a\widehat{\sqsubset}b \implies a\widehat{\sqsubset}^wb$ , only  $\approx$  and  $\sqsubset$  may be changed. Hence, no rule is violated which can be verified by inspection.

(3) This is in principle analysis of triples (step 1 in the algorithm), so we have  $O(n^3)$ . Because of the step 3 of the algorithm, each triple can violate each rule only once and the number of rules to violate is finite, so it remains  $O(n^3)$ . Calculating global score and other operations from step 5 are  $O(n^2)$  so the algorithm is  $O(n^3)$ . ■

Proposition 3.3 might suggest that we do not need Algorithm 1 at all!, but this suggestion is wrong. On contrary, we recommend the following procedure for deriving a pairwise comparison consistent ranking system from a given set of raw pairwise comparison ranking data.

- Procedure 1 (1) Apply Algorithm 1 but without insisting on weak ordering of the relation  $\sqsubset$ .  
 (2) Apply Algorithm 2 in full. ■

Procedure 1 is clearly  $O(n^3)$ . The reasons for *not* using Algorithm 2 on raw pairwise comparison ranking data are the following. Algorithm 1 employs procedures for partial order approximations of arbitrary relations (see [13]). Those procedures both extend and prune the initial relations, but they do not always lower preferences, and may also increase them. Algorithm 2 always lowers preferences when there is any inconsistency. While Algorithm 1 produces partial order *approximations* (in the sense of [13], Algorithm 2 alone, *does not!* If the initial data are ‘almost’ a ranking system or ‘almost’ consistent ranking system, Algorithm 2 alone may produce the same result as Procedure 1. However, if the initial data are ‘far away’ from a ranking system, Algorithm 2 alone will produce a huge  $\approx$  and few higher preferences, and the result may be of no practical value. Using Procedure 1 where raw data are first made into partial orders and later made consistent should give better results. Our tests discussed in next section support this argument.

## 5 Testing

How can we test the results of the algorithms presented in Sections 3 and 4? How do we know if they produce results that make any sense? How can we compare them with algorithms constructed using numerical ranking paradigms?

Testing means that there are some data and results that are known to be correct, and then the technique is applied to the same data. The differences between the correct results and those obtained by a given technique are used to judge the value of the technique. Hence testing models such as the one presented above is problematic since *it is not obvious what should be tested against*. What are the correct results for given data? If the object has measurable attributes and there is a precise algorithm to calculate the value, the whole problem disappears. Nevertheless, we think we have designed a proper test (suggested in [17]) for these kinds of ranking techniques.

A blindfolded person compared the weights of stones. The person put one stone in his left hand and another in his right hand, and then decided which of the relations  $\approx$ ,  $\sqsubset$ ,  $\subset$ ,  $<$  or  $\prec$  (interpreted

TABLE 1. The initial ranking data and its numbers of consistency violations

	A	B	C	D	E	F	G	H
A	$\approx$	$\sqsubset$	$\approx$	$\sqsubset$	$\supset$	$\approx$	$<$	$\sqsubset$
B	$\supset$	$\approx$	$>$	$\approx$	$\supset$	$\sqsubset$	$\approx$	$>$
C	$\approx$	$<$	$\approx$	$\sqsubset$	$\sqsubset$	$<$	$<$	$\sqsubset$
D	$\sqsubset$	$\approx$	$\sqsubset$	$\approx$	$\supset$	$\approx$	$<$	$\sqsubset$
E	$\sqsubset$	$<$	$\sqsubset$	$<$	$\approx$	$<$	$<$	$\approx$
F	$\approx$	$\sqsubset$	$\sqsubset$	$\approx$	$>$	$\approx$	$\sqsubset$	$\supset$
G	$>$	$\approx$	$>$	$\sqsubset$	$\supset$	$\sqsubset$	$\approx$	$>$
H	$\sqsubset$	$<$	$\sqsubset$	$<$	$\approx$	$\sqsubset$	$<$	$\approx$

	A	B	C	D	E	F	G	H
A	0	0	0	0	0	0	0	1
B	0	0	0	0	0	0	0	0
C	0	0	0	0	1	0	0	5
D	0	0	0	0	2	0	0	1
E	0	0	1	2	0	0	0	0
F	0	0	0	0	0	0	0	1
G	0	0	0	0	0	0	0	1
H	1	0	5	1	0	1	1	0

The initial ranking data, not a consistent ranking.

Numbers of violations for the initial ranking data. The grey cells indicate maximum number of violations.

as described in previous section) held. The experiment was repeated for the same set of stones by various people; and then again for different stones and different number of stones; and again for various subsets of  $\{\approx, \sqsubset, \sqsubset, <, <\}$ . Those experiments have most likely been carried out by prehistoric man. Our ancestors probably used this technique to decide which stone was better to kill an enemy or an animal.

*In this experiment, the stones can be weighted using a precise scale, so we have the precise results to test against.*

Both the results of those experiments and their analysis, including the comparison with numerical ranking techniques, can be found in [24]. They support all the claims made in this article. In general, inconsistencies occur quite often, but after using Procedure 1 we always obtained ranking that did not contradict the real weights of stones.

## 6 An example

We will now illustrate Algorithm 2. Examples of applying Algorithm 1 can be found in [13].

The following experiment has been conducted. A blindfolded person compared the weights of the eight different stones, named  $A, B, C, D, E, F, G, H$ . Table 1 (left) presents the results of one such an experiment [11, 24]. The raw ranking data  $(X, \approx, \sqsubset, \sqsubset, <, <)$ , where  $X = \{A, B, C, D, E, F, G, H\}$ , described by Table 1 (left) is not a consistent ranking and its violation numbers are presented in Table 1 (right). From Table 1 (right), we can see that the pair  $(C, H)$  and its symmetric counterpart  $(H, C)$  have five violations (grey cells in Table 1 (right)), and all other pairs have less. In Table 1 (left), we have  $C \sqsubset H$ ; however, according to the consistency rules, the relationship  $(C, H)$  violates the following rules:

- $(C \approx A) \wedge (A \sqsubset H) \Rightarrow (C \sqsubset H) \vee (C \supset H)$  (by rule 8.1)
- $(C \sqsubset D) \wedge (D > H) \Rightarrow (C \sqsubset H) \vee (C \supset H)$  (by rule 15.2)
- $(C \sqsubset E) \wedge (E \approx H) \Rightarrow (C \sqsubset H) \vee (C \supset H)$  (by rule 8.1)
- $(C \sqsubset F) \wedge (F \supset H) \Rightarrow (C \sqsubset H) \vee (C \supset H)$  (by rule 14.1)
- $(C < G) \wedge (G > H) \Rightarrow (C \sqsubset H) \vee (C \supset H) \vee (C > H) \vee (C > H)$  (by rule 16.3)

similarly for a symmetric pair  $H \sqsubset C$ . The total number of violations is 5.

Algorithm 2 revises the pairs  $(C, H)$  and  $(H, C)$  to  $\approx$ , and the new relationship is presented in Table 2 (left). This is still not a consistent ranking as Table 2 (right) indicates. From Table 2 (right), we can see that the pair  $(D, E)$  and its symmetric counterpart  $(E, D)$  have maximum violations (grey cells in Table 2 (right)). In this case we have  $D > E$ , the pair  $(D, E)$  violates the following rules:

- $(D \sqsubset A) \wedge (A \supset E) \Rightarrow (D \supset E) \vee (D > E)$  (by rule 9.2)

TABLE 2. The relationship after revising pairs  $(C, H)$  and  $(H, C)$  to  $\approx$ 

	A	B	C	D	E	F	G	H
A	$\approx$	$\subset$	$\approx$	$\sqsubset$	$\supset$	$\approx$	$<$	$\sqsubset$
B	$\supset$	$\approx$	$>$	$\approx$	$>$	$\sqsubset$	$\approx$	$>$
C	$\approx$	$<$	$\approx$	$\subset$	$\sqsubset$	$\sqsubset$	$<$	$\approx$
D	$\sqsubset$	$\approx$	$\supset$	$\subset$	$>$	$\approx$	$\sqsubset$	$>$
E	$\subset$	$>$	$\sqsubset$	$<$	$\approx$	$<$	$>$	$\approx$
F	$\approx$	$\sqsubset$	$\sqsubset$	$\approx$	$>$	$\approx$	$\subset$	$\supset$
G	$>$	$\approx$	$>$	$\supset$	$>$	$\supset$	$\approx$	$>$
H	$\sqsubset$	$<$	$\approx$	$<$	$\approx$	$\subset$	$<$	$\approx$

	A	B	C	D	E	F	G	H
A	0	0	0	0	0	0	0	0
B	0	0	0	0	0	0	0	0
C	0	0	0	0	0	0	0	0
D	0	0	0	0	2	0	0	0
E	0	0	0	2	0	0	0	0
F	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	0

After the first revision. The grey cells were revised.

Numbers of consistency rules violations. The grey cells have maximum inconsistencies

TABLE 3. The case after revising  $(D, E)$  to  $>$  and  $(E, D)$  to  $<$ . This is a consistent ranking system

	A	B	C	D	E	F	G	H
A	$\approx$	$\subset$	$\approx$	$\sqsubset$	$\supset$	$\approx$	$<$	$\sqsubset$
B	$\supset$	$\approx$	$>$	$\approx$	$>$	$\sqsubset$	$\approx$	$>$
C	$\approx$	$<$	$\approx$	$\subset$	$\sqsubset$	$\sqsubset$	$<$	$\approx$
D	$\sqsubset$	$\approx$	$\supset$	$\subset$	$>$	$\approx$	$\sqsubset$	$>$
E	$\subset$	$>$	$\sqsubset$	$<$	$\approx$	$<$	$>$	$\approx$
F	$\approx$	$\sqsubset$	$\sqsubset$	$\approx$	$>$	$\approx$	$\subset$	$\supset$
G	$>$	$\approx$	$>$	$\supset$	$>$	$\supset$	$\approx$	$>$
H	$\sqsubset$	$<$	$\approx$	$<$	$\approx$	$\subset$	$<$	$\approx$

	A	B	C	D	E	F	G	H
A	0	0	0	0	0	0	0	0
B	0	0	0	0	0	0	0	0
C	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0	0
E	0	0	0	0	0	0	0	0
F	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	0

After the second revision. This is a consistent ranking. The grey cells were revised.

Number of consistency rules violations (none!)

TABLE 4. Global scores for partial orders  $\hat{\sqsubset}$ ,  $\hat{\sqsubset}$ ,  $\hat{\supset}$  and  $\hat{\supset}$ 

Global score	E	H	C	A	F	D	B	G
$\hat{\sqsubset}$	-6	-5	-3	-1	1	3	5	6
$\hat{\sqsubset}$	-5	-4	-3	-1	1	3	4	5
$\hat{\supset}$	-4	-3	-2	-1	1	2	3	4
$\hat{\supset}$	-2	-1	0	0	0	0	1	2

- $(D \supset C) \wedge (C \supset E) \Rightarrow (D \supset E) \vee (D > E)$  (by rule 9.2),

similarly for a symmetric pair  $E < D$ . Here the total number of violations is 2.

Algorithm 2 revises the pair  $(D, E)$  to  $>$  and  $(E, D)$  to  $<$ . The results are presented in Table 3 (left), and this is a consistent ranking. Its numbers of consistency violations are given in Table 3 (right). From it, we can see that none of pair violates the consistency rules.

The relationship presented in Table 3 (left), is now a consistent ranking. However, it is *not* a weakly ordered ranking. Using global score function, we obtain the global scores for the partial order  $\hat{\sqsubset}$ ,  $\hat{\sqsubset}$ ,  $\hat{\supset}$  and  $\hat{\supset}$  presented in Table 4. The weak extensions ranking orders are described in Table 5. Usually we are only interested in finding a weak extension of  $\hat{\sqsubset}$  (see step 5 of Algorithm 2); however, in this case we have found weak extensions of all preferences to illustrate a phenomenon discussed in detail in [13], namely *less precise preferences often work as well, or better, than finer preferences*.

The stones were weighted and their weights created an increasing total order  $E, H, C, A, F, D, B, G$ . Note that this is *the same order* as weak extensions  $\hat{\sqsubset}^w$ ,  $\hat{\sqsubset}^w$ ,  $\hat{\supset}^w$ . The fact that  $\hat{\supset}^w$  correctly describes the real ordering is very interesting since it means that the very rough ranking was sufficient to produce a correct total ordering. This problem is discussed in detail in [13].



TABLE 5. Ranking orders for weak extensions  $\hat{C}^w$ ,  $\hat{C}^w$ ,  $\hat{C}^w$  and  $\hat{C}^w$ 

Weak extension	Ranking order
$\hat{C}^w$	$E \hat{C}^w H \hat{C}^w C \hat{C}^w A \hat{C}^w F \hat{C}^w D \hat{C}^w B \hat{C}^w G$
$\hat{C}^w$	$E \hat{C}^w H \hat{C}^w C \hat{C}^w A \hat{C}^w F \hat{C}^w D \hat{C}^w B \hat{C}^w G$
$\hat{C}^w$	$E \hat{C}^w H \hat{C}^w C \hat{C}^w A \hat{C}^w F \hat{C}^w D \hat{C}^w B \hat{C}^w G$
$\hat{C}^w$	$E \hat{C}^w H \hat{C}^w C \hat{C}^w A \hat{C}^w F \hat{C}^w D \hat{C}^w B \hat{C}^w G$

## 7 Final comment

The concepts of consistent ranking and pairwise comparison ranking data have been defined and analysed in the setting of partial orders. Some algorithms have been presented. No numbers were used whatsoever, which we believe is more fair and objective approach. A method of testing has been proposed. The approach presented in this article is an extension of models proposed in [11, 13]. Implementation of Algorithms 1 and 2, and much more, can be found in [24].

## References

- [1] K. J. Arrow. *Social Choice and Individual Values*. J. Wiley, 1951.
- [2] N. Cowan. The magical number 4 in short-term memory. A reconsideration of mental storage capacity. *Behavioural and Brain Sciences*, **24**, 87–185, 2001.
- [3] H. A. David. *The Method of Paired Comparisons*. Oxford University Press, 1988.
- [4] A. K. A. De Castro, P. R. Pinheiro, and M. C. D. Pinheiro. A Multicriteria Model Applied in the Diagnosis of Alzheimer's Disease. Vol. 5009 of *Lecture Notes in Artificial Intelligence*, pp. 612–619. Springer, 2008.
- [5] M. de Condorcet. Essai sur l'application de l'analyse a la probabilité des décisions rendue a la pluralité des voix. In *Imprimerie Royale*, Paris 1785, reprinted by Chelsea Publishing Co., 1972.
- [6] J. S. Dyer. Remarks on the analytic hierarchy process. *Management Science*, **36**, 244–258, 1990.
- [7] G. T. Fechner. *Elemente der Psychophysik*. Breitkopf und Härtel, 1860.
- [8] P. C. Fishburn. *Interval Orders and Interval Graphs*. J. Wiley, 1985.
- [9] P. C. Fishburn and W. V. Gehrlein. A comparative analysis of methods for constructing weak orders from partial orders. *J. Math. Sociol.* **4**, 93–102, 1975.
- [10] S. French. *Decision Theory*. Ellis Horwood, 1986.
- [11] R. Janicki. Pairwise Comparisons, Incomparability and Partial Orders. In *Proceedings of ICEIS'2007 (International Conference on Enterprise Information Systems)*, Vol. 2 of *Artificial Intelligence and Decision Support System*, pp. 296–302, 2007.
- [12] R. Janicki. Some remarks on approximations of arbitrary binary relations by partial orders. Vol. 5306 of *Lecture Notes in Artificial Intelligence*, pp. 81–91. Springer, 2008.
- [13] R. Janicki. Pairwise comparisons based non-numerical ranking. *Fundamenta Informaticae*, **94**, 197–217, 2009.
- [14] R. Janicki and W. W. Koczkodaj. Weak order approach to group ranking. *Computers Mathematics with Applications*, **32**, 51–59, 1996.
- [15] R. Janicki and W. W. Koczkodaj. A weak order solution to a group ranking and consistency-driven pairwise comparisons. *Applied Mathematics and Computation*, **94**, 227–241, 1998.
- [16] W. W. Koczkodaj. A new definition of consistency of pairwise comparisons. *Mathematical and Computer Modelling*, **18**, 79–84, 1993.
- [17] W. W. Koczkodaj. On testing subjective rankings, Unpublished Suggestion, 2007.

- [18] L. Lima, P. Novais, R. Costa, J. B. Cruz, and J. Neves. Group decision making and Quality-of-Information in e-Health systems. *Logic Journal of IGPL*, 2010.
- [19] G. A. Miller. The magical number seven, plus or minus two. *The Psychological Review*, **63**, 81–97, 1956.
- [20] K. H. Rosen. *Discrete Mathematics and Its Applications*. McGraw-Hill, 1999.
- [21] T. L. Saaty. A scaling methods for priorities in hierarchical structure. *Journal of Mathematical Psychology*, **15**, 234–281, 1977.
- [22] E. Schröder. *Algebra der Logik*, Teuber, 1895.
- [23] L. L. Thurstone. A law of comparative judgements. *Psychological Reviews*, **34**, 273–286, 1927.
- [24] Y. Zhai. *Pairwise Comparisons Based Non-Numerical Ranking*. PhD Thesis. Department of Computing and Software, McMaster University, 2010. Availabe at <http://www.cas.mcmaster.ca/~zhai/>.

Received 12 January 2011