Soft Computing (2019) 23:6827–6845
https://doi.org/10.1007/s00500-018-03724-3

FOUNDATIONS

A study of similarity measures through the paradigm of measurement theory: the classic case

Giulianella Coletti[1] · Bernadette Bouchon-Meunier[2,3]

Published online: 2 January 2019

Abstract
Similarity measures are used in various tasks dealing with the management of data or information, such as decision-making, case-based reasoning, cased-based information retrieval, recommendation systems and user profile analysis, to cite but a few. The paper aims at providing information on similarity measures that can help in choosing "a priori" one of them on the basis of the semantics behind this choice. To this end, we study similarity measures from the point of view of the ranking relation they induce on object pairs. Using a classic method of measurement theory, we establish necessary and sufficient conditions for the existence of a particular class of numerical similarity measures, representing a given binary relation among pairs of objects which express the idea of "no more similar than". The above conditions are all (and only) the rules which are accepted when one decides to evaluate similarity through any element of a specific class of similarity measures. We exemplify the possible application of such conditions and the relevant results on a real-world problem and discuss them in the ambit of cognitive psychology. We consider here a classical context, while the fuzzy context will be studied in a companion paper.

Keywords Comparative similarities · Boundary axioms · Uniformity axioms · Monotonicity axioms · Independence axioms · Representability by similarity measures

1 Introduction

Similarity measures, a review, or even a listing of all the uses of similarity is impossible, since every day new applications appear. Starting from the seminal study of Jaccard (1908), who proposed a similarity measure to classify ecological species, many binary similarity measures have been proposed in various fields. Later on, they have constituted an essential tool to deal with many problems in almost every scientific field, in particular in medicine, law, bioinformatics, retrieval problems, risk forecasting or meteorology. They are currently used in various tasks dealing with the management of data or information, such as cased-based or content-based information retrieval, pattern recognition, text summarisation, time series analysis, recommendation systems, user profile exploitation and decision-making, to cite but a few.

Fuzzy set theory has also developed its own measures of similarity, which find application in analogous areas when data are subject to uncertainty, incompleteness or imprecision.

The degree to which people perceive two objects or situations as similar fundamentally affects their rational thought and behaviour. We recall that the rigorous study of the similarity concept and its measurement was born in the context of cognitive psychology (Tversky 1977), and today experiments and theory in psychology and learning are fascinating fields in which similarity plays a fundamental role. This is especially true in theories of recognition (Hwang et al. 2012; Pelillo 2013) and categorisation of objects (Hahn and Ramscar 2001), in artificial intelligence (Rissland 2006), but also in modelling of preferences or decision models (Gilboa et al. 2006; Gilboa and Schmeidler 1995, 1997; Ha and Haddawy 2003).

Communicated by A. Di Nola.

G. Coletti is a member of INdAM research group GNAMPA.

✉ Giulianella Coletti
giulianella.coletti@unipg.it

Bernadette Bouchon-Meunier
Bernadette.Bouchon-Meunier@lip6.fr

1 Dipartimento di Matematica e Informatica, Università di Perugia, via Vanvitelli, 1, 06023 Perugia, Italy

2 Sorbonne Universités, UPMC Univ Paris 06, UMR 7606, LIP6, 75005 Paris, France

3 CNRS, UMR 7606, LIP6, 75005 Paris, France

Many similarities measures are available in the literature, both in the classical crisp context (see Choi et al. 2010; Lesot et al. 2009, for a survey) and in the fuzzy context (see, for instance, Bhutani and Rosenfeld 2003; Bouchon-Meunier et al. 1996; Li et al. 2014; Couso et al. 2013, for a formal study). The choice of one of them is done each time two images, cases, objects, situations, profiles, texts, data, must be compared. In each task, this choice is based more on an empirical study or a tradition than a thoughtful approach; indeed, a limited number of comparative studies collected the broad variety of binary similarity measures (for recent research on this line, see Boriah et al. 2008; Bouchon-Meunier et al. 2009; Cross and Sudkamp 2002, for the crisp case and Bouchon-Meunier et al. 2010; Couso et al. 2013, for the fuzzy case).

Usually, the comparison among different similarity measures is made for a particular environment and "a posteriori" (on the basis of the obtained results), focusing on one or more specific properties. For instance, in the recent literature, we can cite Zhang et al. (2006), where six similarity measures have been assessed for trajectory clustering in outdoor surveillance scenes and evaluated by taking into account correct clustering rate and time cost. In Penney et al. (1998), a comparison of six similarity measures used in intensity-based multidimensional image recording is considered, with the aim of establishing which of them are able to record accurately and robustly even when soft tissue structures and interventional instruments are present, using differences between the images. In Toussaint (2004), similarity measures are compared on the basis of the insight they provide about the structural interrelationships that exist within families of rhythms. In Filev et al. (2005), twelve different similarity measures are considered with the scope of evaluating the effectiveness of matching the corresponding masses on temporal pairs of current and prior mammograms.

Finally, we want to point out the recent article (Choi et al. 2010), in which 76 different binary similarity and distance measures are collected and analysed and their correlation is revealed through the hierarchical clustering technique.

Nevertheless, the problem of choosing the best similarity measure to handle large classes of problems is in fact not completely solved, and this paper intends to contribute to this discussion.

Since "similarity" and "difference" influence so many other cognitive processes, it is therefore crucial to determine which factors influence the perception of similarity and difference themselves and to investigate if the most used classes of similarity measures capture these instances or, on the contrary, impose too restrictive rules.

The purpose of this paper is to provide data scientists, computer engineers and, more generally, field experts, with conscious reasons to use a particular similarity measure on the basis of the semantics behind this choice.

More precisely, our main aim is to provide information on the similarity measures or better on classes of them that may help in choosing "a priori" one of them on the basis of the semantics behind this choice. In other words, we propose to investigate which conditions related to the primitive idea of similarity are accepted when one chooses a particular measure. We think that this is particularly important in cognitive psychology and in case-based reasoning where a normative framework of reference is necessary.

For this purpose, the best strategy is to use the paradigm of measurement theory, already presented in Tversky's seminal paper (Tversky 1977), based on the method of measurement theory (Krantz et al. 1971; Suppes et al. 1989) and continue a research whose preliminary results have been presented in Bouchon-Meunier et al. (2009) (see also Bouchon-Meunier et al. 1996; Coletti and Di Bacco 1989). More precisely, we consider a binary relation (comparative similarity) defined on a set of pairs of objects described by a set of attributes (in the meaning of Tversky, for instance, "blue eyes"), expressing the primitive idea of "no more similar than", and we look for necessary and sufficient conditions for the representability of such relation by some element of a specific class of similarity measures.

Indeed, it is our conviction that the best way to capture all the implications intrinsic to the choice of one or the other measures of similarity is to remove the numerical superstructure which can either hide the real constraints that the use of a specific measure of similarity induces or suggests only apparent constraints. As discussed in Sect. 4, this paradigm can be used to formalise the idea of similarity of a field expert, which, otherwise would only choose to take into account more or less good performances of one or the other similarity measure in problems or fields close to his.

This approach points out that many differences among similarity measures are only apparent: different measures give different numerical values, but it is important to observe that the order relation in "similarity" among the pairs of objects induced by these similarity measures can be, in whole or in part, the same. For instance, let us consider the class of similarity measures, parametrised by two real (strictly) increasing functions f and g, assuming 0 in 0, and defined by the following equation (where $|A|$ denotes the cardinality of A)

$$S_{f,g}(X,Y)\begin{cases}\dfrac{f(|X\cap Y|)}{f(|X\cap Y|)+g(|X\Delta Y|)} & \text{for } (X,Y)\neq(\underline{0},\underline{0})\\ 0 & \text{for } (X,Y)=(\underline{0},\underline{0})\end{cases} \tag{1}$$

When f and g and f' and g' are such that

$$S_{f',g'}=\theta(S_{f,g})$$

with θ real strictly increasing function, the ordering in similarity among the pairs coincides (for a discussion in this point, see Lesot and Rifqi 2010). But, as we will prove in this paper, many of the equivalence and the strict relations among pairs coincide independently of the choice of the functions f and g in the above class. The relations among the remaining pairs, which in fact depend on the functions f and g, differ, but are controlled by the same rules.

In this paper, the first one of a series of two, we study the problem in a crisp environment, where an attribute can only be (completely) present or absent in any object. (In other words, the objects are represented by subsets of a set of attributes.)

Most similarity measures in the literature, such as those in the class $S_{f,g}$, are functions of the cardinality of the intersection (the set of common present attributes) and the cardinalities of the two differences (the sets of attributes present in only one of the objects). Some measures consider also the cardinality of attributes absent in both the objects. We compare similarities representable by elements of some classes of these measures, and our approach puts in evidence what we must accept, from a semantic point of view, when we choose one or another class of similarity measures.

The paper is organised as follows: In Sect. 2, we introduce the problem and present the axioms, explained on a current example. Moreover, the first results (actually the main ones for applications in cognitive psychology) are proved: we provide a characterisation of the comparative similarities representable by similarity measures with the same properties as Tversky's model, that is to say depending on the cardinality of intersection and differences, increasing with respect to the cardinality of intersection and decreasing with respect to the cardinality of differences. In Sect. 3, theorems of representability of a comparative similarity by the elements of specific classes of similarity measures are presented. Section 4 presents a real-life utilisation of the results, focusing in cognitive psychology and in the elicitation of similarity measures from domain experts.

2 Comparative similarities

Following the ideas of Tversky (1977), we study the concept of similarity, using the framework of measurement theory (Krantz et al. 1971), in a class of objects described by p attributes, i.e. by the set of features from a predefined list. In this section, we consider situations in which the p attributes are binary, which means that the features can only be either present or absent in any object.

We first define the so-called "comparative similarities" as binary relations enabling to compare the similarities of pairs of objects. In order to provide the designer of a similarity-based system with efficient tools to evaluate similarities, we show that such comparative similarities can be represented by similarity measures, producing numerical values of similarities and easier to manipulate. We point out various properties of comparative similarities, and we show that choosing several of them properly enables the designer to tune its system according to the behaviour of comparative similarities he/she prefers.

2.1 Preliminaries

Let $\mathcal{H}$ be a set of p binary attributes a_k $(k = 1, \ldots, p)$ (which can only be present or absent) and fix any permutation of their indices. Then, the data set $\mathcal{X} = \{0, 1\}^p$ consists of all the vectors $X = (x_1, \ldots, x_p)$, $x_i \in \{0, 1\}$, where $x_i = 1$ if the attribute a_i is present in the object represented by X and 0 if it is absent, in particular $\underline{0} = (0, \ldots, 0)$ is in $\mathcal{X}$. In other terms, the elements of $\mathcal{X}$ can be regarded as the characteristic vectors (functions) of the subsets of $\mathcal{H}$. Consider now, for any $X \in \mathcal{X}$, the set $I_X = \{i : x_i = 1\}$ and denote by $|X|$ the cardinality of I_X.

We will use the following notations: $x_i^c = 1 - x_i$, $X_k^c = (x_1, \ldots, x_k^c, \ldots, x_p)$, $X^c = (x_1^c, \ldots, x_p^c)$,

With an abuse of notation, we indicate by $X \cap Y$ and $X \cup Y$ the elements of $\mathcal{X}$ which are the characteristic vectors of the intersection and union of the two subsets (objects) represented by X and Y, respectively. Precisely, the k-th component of $Z = X \cap Y$ is $z_k = \min\{x_k, y_k\}$ and $w_k = \max\{x_k, y_k\}$ that of $W = X \cup Y$. Moreover, we consider the subsets (objects) $X \setminus Y = X \cap Y^c$ and $Y \setminus X = X^c \cap Y$, that is the sets V and V', whose k-th components are $v_k = \min\{x_k, 1-y_k\}$ and $v'_k = \min\{1-x_k, y_k\}$, and indicate by $X \Delta Y = (X \setminus Y) \cup (Y \setminus X) = (X \cap Y^c) \cup (Y \cap X^c)$, the symmetrical difference between X and Y. Finally, $X \subset Y$ means that the subset (object) represented by X is contained on that represented by Y and so for every $k \in \{1, \ldots, p\}$ one has $x_k \leq y_k$. Similarly, $X \supset Y$ means $x_k \geq y_k$, for every $k \in \{1, \ldots, p\}$.

Let us now consider a *comparative similarity* on $\mathcal{X}^2$, defined as a binary relation $\preceq$ on $\mathcal{X}^2$, with the following meaning:

for $X, Y, X', Y' \in \mathcal{X}$, $(X, Y) \preceq (X', Y')$ means that X is similar to Y no more than X' is similar to Y'.

The relations $\sim$ and $\prec$ are then induced by $\preceq$ in the usual way: $(X, Y) \sim (X', Y')$, meaning that X is similar to Y as X' is similar to Y', when $(X, Y) \preceq (X', Y')$ and $(X', Y') \preceq (X, Y)$. Lastly, $(X, Y) \prec (X', Y')$ when $(X, Y) \preceq (X', Y')$ holds, but not $(X', Y') \preceq (X, Y)$.

If $\preceq$ is assumed to be total, then $\sim$ and $\prec$ are, respectively, the symmetrical and the asymmetrical part of $\preceq$.

Now, for such a comparative similarity, we introduce the notion of representability by means of a numerical similarity measure:

Definition 1 Given a comparative similarity $\preceq$, we say that a similarity measure $S : \mathcal{X}^2 \to \mathbb{R}$ *represents* $\preceq$ if and only if:
for every $(X, Y), (X', Y') \in \mathcal{X}^2$

$$\begin{cases} (X, Y) \preceq (X', Y') \Longrightarrow S(X, Y) \leq S(X', Y'), \\ (X, Y) \prec (X', Y') \Longrightarrow S(X, Y) < S(X', Y'). \end{cases}$$

As is well known, if $\preceq$ is a total relation, the above conditions can be summarised as follows:

$$(X, Y) \preceq (X', Y') \Longleftrightarrow S(X, Y) \leq S(X', Y').$$

We are interested in finding comparative similarities satisfying particular sets of axioms, which are necessary and sufficient conditions for their representability with particular similarity measures (or classes of similarity measures).

In the following subsection, we consider some preliminary results and axioms that lead to some relations between similarity measures and comparative degrees of similarity.

2.2 Basic axioms

The first three axioms we introduce describe basic properties that a binary relation must satisfy in order to be a comparative similarity both natural and representable by a similarity measure. The first one only states that the relation must be a weak order. (We recall that it is necessary for the representability of any relation by a real function.)

Axiom S0 [weak order]
$\preceq$ is a weak order, i.e. it is a complete, reflexive and transitive binary relation.

The second axiom expresses the following boundary conditions: it imposes that for any X, whatever Y, X cannot be less similar to Y than it is to its complement. Moreover, X is strictly more similar to itself than it is similar to its complement.

Axiom S1 [boundary conditions]
For every $X, Y \in \mathcal{X}$, with $X \neq Y$

$$(X^c, X) \sim (Y^c, Y) \preceq (X, Y)$$

and, for $X, Y \neq \underline{0}$

$$(X^c, X) \prec (X, X)$$

The following Axiom S2 expresses that, in pairs of objects having no attribute in common, the number of attributes present in only one object does not influence the degree of similarity. Similarly, in pairs with no attribute present in only one object, the degree of similarity is not influenced by the presence of more or less attributes in common.

Axiom S2 [weak boundary conditions]
For every $X, Y \in \mathcal{X}$, with $X \cap Y = \underline{0}$,

$$(X, Y) \sim (X_k^c, Y) \sim (X, Y_h^c)$$

for any $k \in I_{X \cup (X^c \cap Y^c)}$ and $h \in I_{Y \cup (X^c \cap Y^c)}$
For every $X \in \mathcal{X}$, with $\underline{0} \neq X$ and for any k such that $X_k^c \neq \underline{0}$

$$(X, X) \sim (X_k^c, X_k^c)$$

The following example will be used in the sequel to better explain the meaning of the axioms that will be presented. It can represent, for instance, the basis of a case-based reasoning system to estimate the price of an apartment, given the prices of other apartments recently sold.

Example 1 Let us consider a comparative similarity among apartments in New York based on the following attributes:

a)is in a high floor;
b) has a living kitchen;
c) has a bathroom with a window;
d) is equipped with a lift;
e) has a terrace.

Let us consider the pairs (X, Y), (X', Y'), (X'', Y''), as shown in the next table. Axiom S2 requires them to be equally similar, since they have no attribute in common in the three cases:

$\mathcal{H}$	a	b	c	d	e
X	0	0	1	0	0
Y	0	1	0	1	0
X'	1	0	0	0	1
Y'	0	1	0	1	0
X''	1	0	0	0	1
Y''	0	0	0	0	0

So, for instance, an apartment on a low floor with windows in the bathroom, without terrace, without lift and with no living kitchen (X) is similar to an apartment on a low floor with a living kitchen, a lift, without terrace and without windows in the bathroom (Y), as much as an apartment without any of the considered attributes (Y'') is similar to an apartment on a high floor with a terrace without the other attributes (X'').

The third axiom imposes a symmetry condition, natural for the comparison of objects at the same level, but questioned

by Tversky in the quoted paper (Tversky 1977) because, in many cases, one of the objects acts as a reference (i.e. a prototype) to which the second one is compared.

Axiom S3 [symmetry]
$\forall X, Y \in \mathcal{X}$,

$$(X, Y) \sim (Y, X)$$

Let us now consider any $X \in \mathcal{X}$ and denote by $X'_{h,k}$ the element of $\mathcal{X}$ coinciding with X for $i \neq h, k$ and such that $x'_h = x_k$, $x'_k = x_h$. For instance, if $X = (0, 0, 1, 0, 0, 1, 1)$ then $X'_{1,6} = (1, 0, 1, 0, 0, 0, 1)$.

Axiom S4 [partial attribute uniformity]
For every $X, Y \in \mathcal{X}$ and for every $h, k \in \{1, \ldots, p\}$,

$$(X, Y) \sim (X'_{h,k}, Y'_{h,k}).$$

Axiom S4 requires that all the attributes play the same role with respect to the similarity degree: it rules the construction of a first partition of the objects in equivalence classes with respect to the similarity.

Example 2 By using Example 1, we get $(X, Y) \sim (X', Y')$, where X, Y, X', Y' are those in the following table:

$\mathcal{H}$	a	b	c	d	e
X	1	1	0	0	1
Y	1	0	1	0	1
X'	0	1	0	1	1
Y'	0	0	1	1	1

The following proposition, whose proof follows immediately by applying a finite number of times Axiom S4 and Axiom S0, shows that, under these axioms, a pair (X, Y) is equivalent to all the other ones obtained by permuting in the same way the elements of I_X and I_Y. This means that all the attributes play the same role in the similarity perception expressed by the relation $\preceq$.

Proposition 1 *Let $\preceq$ be a comparative similarity on $\mathcal{X}^2$ satisfying Axioms S0 and S4. Let $X, Y, X', Y' \in \mathcal{X}$ with $X = (x_1, \ldots, x_p)$, $Y = (y_1, \ldots, y_p)$, $X' = (x_{i_1}, \ldots, x_{i_p})$, $Y' = (y_{i_1}, \ldots, y_{i_p})$, where $(i_1, \ldots, i_p)$ is any permutation of $(1, \ldots, p)$. Then, $(X, Y) \sim (X', Y')$.*

We are now able to prove the following lemma which claims that, under Axioms S0, S2, S4, all the pairs with the same number of common attributes, the same number of attributes only present in the first object and the same number of attributes only present in the second object are equally similar. Moreover, all the pairs without common attributes are equally similar. The pairs in which the attributes are either present in both or absent in both the objects are equally similar too. Let us remark that the symmetry Axiom S3, which is questionable, is not taken into account at this point.

Lemma 1 *Let $X, Y, X', Y' \in \mathcal{X}$ fulfilling one of the following conditions:*

1. $|X \cap Y| = |X' \cap Y'| > 0$, $|X \setminus Y| = |X' \setminus Y'|$ *and* $|Y \setminus X| = |Y' \setminus X'|$, *with* $|X \Delta Y| > 0$;
2. $|X \cap Y| = |X' \cap Y'| = 0$;
3. $X = Y$ *and* $X' = Y'$

If $\preceq$ is a comparative similarity on $\mathcal{X}^2$ satisfying Axioms S0, S2, S4, then

$$(X, Y) \sim (X', Y').$$

Proof If condition 1 holds, then X' and Y' are obtained by the same permutation from X and Y,, respectively. So the thesis follows by Proposition 1. If either condition 2 or condition 3 holds, then the thesis follows immediately by applying Axioms S2 and S0 a finite number of times. □

The next result follows immediately by Lemma 1:

Corollary 1 *If $\preceq$ is a comparative similarity on $\mathcal{X}^2$ satisfying Axioms S0, S2, S4, then, for any $X, Y \in \mathcal{X}$,*

$$(X^c, X) \sim (\underline{0}, \underline{0}) \sim (Y, Y^c).$$

Axioms S5 and S6 enlarge the sets of equivalent pairs singled out in Lemma 1: S5 introduces a symmetry with respect to the attributes only present in the first object and those only present in the second one; S6 introduces a symmetry with respect to the attributes present in both objects and those absent from both.

Axiom S5 clearly expresses the idea that the comparative degree of similarity of a pair (X, Y) is the same as the one of a pair (X', Y'), which differs from it by the presence and absence of an attribute a_k which is present in X, but not in Y, in the first pair and is absent from Y' but not from X' in the second pair or vice versa.

Axiom S5 [distinctive attribute interchangeability]
For every $X, Y \in \mathcal{X}$ and for every $k \in I_{X \Delta Y}$, we have

$$(X, Y) \sim (X^c_k, Y^c_k).$$

Example 3 Continuing Example 1, we consider the following objects:

$\mathcal{H}$	a	b	c	d	e
X	1	0	1	0	1
Y	1	0	0	1	0
X'	1	0	0	0	0
Y'	1	0	1	1	1

Axiom S5 requires that $(X, Y) \sim (X'', Y'')$, with $(X'', Y'') = (X_3^c, Y_3^c)$ and $(X'', Y'') \sim (X''^c_5, Y''^c_5) = (X', Y')$.

Remark 1 Axiom S5 can be formulated as follows:

For every $(X, Y), (X', Y') \in \mathcal{X}^2$ and for any $k \in \{1, \ldots, p\}$, if for every $x_i \in X, x_i' \in X', y_i \in Y, y_i' \in Y'$, we have $x_i = x_i'$ and $y_i = y_i'$ for $i \neq k$ and $x_k' = y_k$, $y_k' = x_k$, then $(X, Y) \sim (X', Y')$.

Lemma 2 *Let $\preceq$ be a comparative similarity on $\mathcal{X}^2$ satisfying Axioms S0, S4, S5. For every $(X, Y), (X', Y') \in \mathcal{X}^2$ such that $|X \cap Y| = |X' \cap Y'|$ and $|X \Delta Y| = |X' \Delta Y'|$ we have*

$$(X, Y) \sim (X', Y').$$

Proof By using Axioms S4 and S0, starting from (X', Y'), with a finite number of transformations, we can obtain a pair $(X'', Y'') \sim (X', Y')$ and such that $X \cap Y = X'' \cap Y''$ and $X \Delta Y = X'' \Delta Y''$. Nevertheless, (X'', Y'') can be different from (X, Y), since one or more elements of $X \Delta Y$ can belong to $X \setminus Y$ and to $Y'' \setminus X''$ (or vice versa). Then, by applying Axioms S5 and S0, we can obtain from (X'', Y'') a pair $(\overline{X}, \overline{Y}) \sim (X'', Y'')$ and $(\overline{X}, \overline{Y}) = (X, Y)$. □

Corollary 2 *If $\preceq$ is a comparative similarity on $\mathcal{X}^2$ satisfying Axioms S0, S2, S4, and S5, then $\preceq$ satisfies S3.*

Proof The proof immediately follows by Lemmas 1 and 2. □

The following Axiom S6 is in fact a reinforcement of Axiom S5 and implies the second condition in S2. It claims that, with respect to the degree of similarity, an attribute present in both objects of a pair has the same importance as an attribute absent from both.

Axiom S6 [attribute interchangeability]
For every $(X, Y) \in \mathcal{X}^2$ and for any $k \in \{1, \ldots, p\}$, we have

$$(X, Y) \sim (X_k^c, Y_k^c).$$

It is immediate to prove the following result:

Proposition 2 *If $\preceq$ is a comparative similarity on $\mathcal{X}^2$ satisfying Axioms S0, S4, and S6 then $\preceq$ satisfies S3 and, for any $X, Y \in \mathcal{X}$*

$$(X^c, X) \sim (\underline{0}, \underline{0}) \sim (Y, Y^c).$$

Example 4 Continuing Example 1 and considering the apartments in the following table, Axiom S6 requires that we must set $(X, Y) \sim (X', Y') = (X^c, Y^c)$.

$\mathcal{H}$	a	b	c	d	e
X	1	0	0	1	0
Y	1	1	0	1	1
X'	0	1	1	0	1
Y'	0	0	1	0	0

The next lemma detects equivalence classes of pairs larger than those detected in Lemma 1: all the pairs containing the same number of attributes either present or absent in both objects (and so the same number of attributes present in only one object) are equivalently similar.

Lemma 3 *Let $\preceq$ be a comparative similarity on $\mathcal{X}^2$ satisfying Axioms S0, S4, S6. For every $(X, Y), (X', Y')$ such that $|X \Delta Y| = |X' \Delta Y'|$ then*

$$(X, Y) \sim (X', Y').$$

Proof The proof follows the same line of the proof of Lemma 1. Indeed, if $|X \Delta Y| = |X' \Delta Y'|$, then X' and Y' are obtained by the same permutation from X and Y,, respectively. So the thesis follows by Proposition 1. □

2.3 Monotonicity axioms

We introduce now three monotonicity axioms: they are based on the simultaneous comparison of intersection and difference of several pairs of objects. The first one explains that the more present attributes the pairs have in common, the more similar they are, for the same number of attributes in only one of the objects of the pair. The less attributes are in only one of the objects, the more similar the pairs are, for an equal number of common present attributes.

Axiom S7$_1$ [monotonicity]
For every $X, Y, X', Y', X'', Y'' \in \mathcal{X}$, if $X \cap Y \supset X' \cap Y' = X'' \cap Y'' \neq \underline{0}$ and $\underline{0} \neq X \Delta Y = X' \Delta Y' \subset X'' \Delta Y''$, then

$$(X'', Y'') \prec (X', Y') \prec (X, Y).$$

Example 5 Continuing Example 1, we consider the apartments in the following table:

$\mathcal{H}$	a	b	c	d	e
X	1	0	1	1	1
Y	1	1	1	0	1
X'	1	1	0	0	1
Y'	1	0	0	1	1
X''	1	0	1	1	1
$Y'' = X'$	1	1	0	0	1

If we accept Axiom $S7_1$, then our comparative similarity must be:

$$(X'', Y'') \prec (X', Y') \prec (X, Y),$$

which means that an apartment with all the attributes except the living kitchen (X) is more similar to an apartment with all the attributes except the lift (Y) than an apartment without lift and windows in the bathroom (X') is similar to one without living kitchen and windows in the bathroom (Y') and finally both are more similar than an apartment without a living kitchen (X'') is similar to an apartment without windows in the bathroom and lift (Y'').

We now consider differently attributes only present in the first object of a pair or only present in the second one, instead of considering indifferently attributes present in any of the objects of the pair as we did in Axiom $S7_1$.

Axiom $S7_2$ [weak monotonicity]
For every $X, Y, X', Y', X'', Y'' \in \mathcal{X}$, if $X \cap Y \supset X' \cap Y' = X'' \cap Y'' \neq \underline{0}$, $\underline{0} \neq X \setminus Y = X' \setminus Y' \subseteq X'' \setminus Y''$, $\underline{0} \neq Y \setminus X = Y' \setminus X' \subseteq Y'' \setminus X''$, and at least one of the two following conditions is satisfied $X' \setminus Y' \subset X'' \setminus Y''$, $Y' \setminus X' \subset Y'' \setminus X''$, then

$$(X'', Y'') \prec (X', Y') \prec (X, Y).$$

We notice that Axiom $S7_1$ is stronger than Axiom $S7_2$: in fact if the hypotheses of $S7_2$ are satisfied, then the hypotheses (and so the implication) of $S7_1$ are also satisfied.

In the next axiom, we propose a different view of monotonicity, restricted to attributes only present in one object of each pair.

Axiom $S7_3$ [cumulative monotonicity]
For every $X, Y, X', Y', X'', Y'' \in \mathcal{X}$, if $X \Delta Y = X' \Delta Y' \subset X'' \Delta Y''$, then

$$(X'', Y'') \prec (X', Y') \sim (X, Y).$$

Example 6 Continuing Example 1, we consider the apartments described in the following table:

$\mathcal{H}$	a	b	c	d	e
X	1	0	1	1	1
Y	1	1	1	0	1
X'	0	1	0	0	1
Y'	0	0	0	1	1
X''	0	0	1	1	0
Y''	1	1	0	0	0

If we accept Axiom $S7_3$, then our comparative similarity must be:

$$(X'', Y'') \prec (X', Y') \sim (X, Y),$$

apartments (X, Y) or (X', Y') with two differences in the presence of attributes being more similar than apartments (X'', Y'') with three differences.

This example leads us to a more "countable" approach of comparative similarities, where the considered sets of attributes are not supposed to be equal, but to have the same cardinality.

Lemma 4 *Let $\preceq$ be a comparative similarity on $\mathcal{X}^2$. For every $(X, Y), (X', Y'), (X'', Y'') \in \mathcal{X}^2$ such that $|X \cap Y| > |X' \cap Y'| = |X'' \cap Y''| \neq 0$ and $0 \neq |X \Delta Y| = |X' \Delta Y'| < |X'' \Delta Y''|$ we have*

$$(X'', Y'') \prec (X', Y') \prec (X, Y)$$

if and only if $\preceq$ satisfies Axioms S0, S4, S5, $S7_1$.

Proof If $|X \cap Y| > |X' \cap Y'|$ and $|X \Delta Y| = |X' \Delta Y'|$ then, from Proposition 1, there exists a pair (X''', Y''') such that $X''' \cap Y''' \subset X \cap Y$ and $X''' \Delta Y''' = X \Delta Y$. From Axiom S5 there exists a pair $(\tilde{X}, \tilde{Y}) \sim (X''', Y''')$ such that $(\tilde{X}, \tilde{Y}), (X', Y')$ satisfy the hypotheses of Lemma 1. So $(\tilde{X}, \tilde{Y}) \sim (X', Y')$. By Axiom $S7_1$ $(\tilde{X}, \tilde{Y}) \prec (X, Y)$ and so, by transitivity $(X', Y') \prec (X, Y)$. Similarly, for (X'', Y'').

The proof of the necessity of Axioms S0, S4, S5, $S7_1$ is trivial. □

Replacing Axiom $S7_1$ by the weak form $S7_2$, we can prove the following result.

Lemma 5 *Let $\preceq$ be a comparative similarity on $\mathcal{X}^2$. For every $(X, Y), (X', Y'), (X'', Y'') \in \mathcal{X}^2$ such that $|X \cap Y| > |X' \cap Y'| = |X'' \cap Y''| \neq 0$, $0 \neq |X \setminus Y| = |X' \setminus Y'| < |X'' \setminus Y''|$ and $0 \neq |Y \setminus X| = |Y' \setminus X'| < |Y'' \setminus X''|$ we have*

$$(X'', Y'') \prec (X', Y') \prec (X, Y),$$

if and only if $\preceq$ satisfies Axioms S0, S4, $S7_2$.

Proof The proof follows the same line of the proof of Lemma 4, by using Lemma 2 and Axiom $S7_2$ instead of Lemma 1 and Axiom $S7_1$, respectively. The proof of the necessity of Axioms S0, S4, $S7_2$ is trivial. □

Lemma 6 *Let $\preceq$ be a comparative similarity on $\mathcal{X}^2$. For every $(X, Y), (X', Y'), (X'', Y'') \in \mathcal{X}^2$ such that $|X \Delta Y| = |X' \Delta Y'| < |X'' \Delta Y''|$ we have*

$$(X'', Y'') \prec (X', Y') \sim (X, Y),$$

if and only if $\preceq$ satisfies Axioms S0, S4, S6, $S7_3$.

Proof Under these hypotheses, we have $|(X \cap Y) \cup (X^c \cap Y^c)| = |(X' \cap Y') \cup (X'^c \cap Y'^c)| > |(X'' \cap Y'') \cup (X''^c \cap Y''^c)|$. Then, the proof follows the same line as the proof of Lemma 4, by using Lemma 3 and Axiom $S7_3$ instead of Lemma 1 and Axiom $S7_1$, respectively. The proof of the necessity of Axioms S0, S4, S6, $S7_3$ is trivial. □

Lemmas 4, 5, 6 together with Lemma 1 suggest that the introduced axioms are sufficient conditions to determine that any similarity measure, representing the considered comparative similarities, must only take into account the cardinality of the common attributes and the distinctive ones, affirming that an increase of the number of common attributes, as well as a decrease of the number of distinct ones, increases the value of the measure of similarity. Moreover, the same axioms assure that the maximum value of the similarity measure representing $\preceq$ is obtained for the pairs (X, X) and the minimum for (X, X^c), for every $X \neq \underline{0}$.

Theorems 1, 2, 3 will show that the introduced axioms are actually necessary and sufficient conditions for the above statements.

Theorem 1 *Let $\preceq$ be a comparative similarity on $\mathcal{X}^2$. The following statements are equivalent:*

i) $\preceq$ satisfies S0, S1, S2, S4, S5, $S7_1$
ii) There exists a unique (up to increasing transformations) function $S : \mathcal{X}^2 \to [0, 1]$ representing $\preceq$ and such that:
   a) $S(X, Y) = \Phi(|X \cap Y|, |X \Delta Y|)$,
   b) Φ is increasing with respect to $|X \cap Y|$ and decreasing with respect to $|X \Delta Y|$,
   c) $\Phi(0, b) = 0$, and $\Phi(a, 0) = 1$, for every a and for every $b \neq 0$

Proof We first prove that $i) \Rightarrow ii)$. Since $\mathcal{X}^2$ is finite, Axiom S0 is a necessary and sufficient condition for the existence of a function $S : \mathcal{X}^2 \to [0, 1]$ representing $\preceq$ (Krantz et al. 1971). Axioms S0, S4, S5 through Lemma 2 ensure that S only depends on $|X \cup Y|$ and $|X \Delta Y|$. Axioms S0, S4, S5 and $S7_1$ through Lemma 4 ensure that S is increasing with respect to $|X \cup Y|$ and decreasing with respect to $|X \Delta Y|$.

Finally, Axioms S0, S1, S2, S4, taking into account Lemma 1 and Corollary 1, ensure the validity of statement c of condition $ii)$.

We now prove $ii) \Rightarrow i)$. Every binary relation $\preceq$ representable by a real function is a weak order, so S0 is necessarily satisfied. Axiom S1 immediately follows from condition $c)$ and the representability of $\preceq$ by S. To prove Axiom S2, consider that, if $X \cap Y = \underline{0}$, we also have $X_k^c \cap Y = X \cap Y_h^c = \underline{0}$ for every $k \in I_{X \cup (X^c \cap Y^c)}$ and $h \in I_{Y \cup (X^c \cap Y^c)}$. So, since $\Phi(0, \cdot) = 0$, we have $S(X, Y) = S(X_k^c, Y) = S(X, Y_h^c) = 0$. Similarly, if $X = Y$ ($X \Delta Y = \underline{0}$,) and $X \neq \underline{0}$, we have, for every k, $X_k^c = Y_k^c$ and, if $X_k^c \neq \underline{0}$, by condition $c)$ we obtain $S(X, X) = S(X_k^c, X_k^c) = 1$. Then, Axiom S2 follows from the representability of $\preceq$ by S. To prove Axiom S4, consider that the cardinality does not depend on the specific elements, and $S(X, Y)$ is a function of the cardinality of the intersection and the symmetrical difference of X and Y. To prove Axiom S5, it is sufficient to consider that, for every $k \in I_{X \Delta Y}$, (X, Y) and (X_k^c, Y_k^c) are such that $|X \cap Y| = |X_k^c \cap Y_k^c|$ and $|X \Delta Y| = |X_k^c \Delta Y_k^c|$ and so $S(X, Y) = S(X_k^c, Y_k^c)$, which implies $(X, Y) \sim (X_k^c, Y_k^c)$. To prove Axiom $S7_1$,, it is sufficient to consider that the cardinality is increasing with respect to the partial order induced by the inclusion and function Φ is increasing with respect to the first variable $|X \cup Y|$ and decreasing with respect to the second variable $|X \Delta Y|$. Finally, Axiom S3 follows from Corollary 2 □

Theorem 2 *Let $\preceq$ be a comparative similarity on $\mathcal{X}^2$. The following statements are equivalent:*

1) $\preceq$ satisfies S0, S1, S2, S4, $S7_2$,
ii) There exists a unique (up to increasing transformations) function $S : \mathcal{X}^2 \to [0, 1]$ representing $\preceq$ and such that:
   a) $S(X, Y) = \Psi(|X \cap Y|, |X \setminus Y|, |Y \setminus X|)$,
   b) Ψ is increasing with respect to $|X \cap Y|$ and decreasing with respect to $|X \setminus Y|$, and $|Y \setminus X|$,
   c) $\Psi(0, b, b') = 0$, and $\Psi(a, 0, 0) = 1$, for every a and for every b, b' with $b + b' \neq 0$

Proof The proof exactly follows the same line as the proof of Theorem 1, by using Lemma 1 instead of Lemma 2 and Lemma 5 instead of Lemma 4. □

Theorem 3 *Let $\preceq$ be a comparative similarity on $\mathcal{X}^2$. The following statements are equivalent:*

i) $\preceq$ satisfies S0, S1, S3, S4, S6, $S7_3$
ii) There exists a unique (up to increasing transformations) function $S : \mathcal{X}^2 \to [0, 1]$ representing $\preceq$ and such that:
   a) $S(X, Y) = \Xi\big(|(X \cap Y) \cup (X^c \cap Y^c)|, |X \Delta Y|\big)$,
   b) Ξ is increasing with respect to $|(X \cap Y) \cup (X^c \cap Y^c)|$ and decreasing with respect to $|X \Delta Y|$,

c) $\Xi(0, b) = 0$, *and* $\Xi(c, 0) = 1$, *for every* c *and for every* $b \neq 0$.

Proof The proof exactly follows the same line as the proof of Theorem 1, by using Lemma 3 instead of Lemma 2 and Lemma 6 instead of Lemma 4 and Proposition 2. □

Remark 2 We notice that the set of axioms present in condition *i*) of Theorem 1 is exactly the "price we must pay" when we claim (often without attaching importance to the fact) that a similarity measure should have the characteristics described in condition *ii*) of the same theorem. Similarly, for axioms in condition *i*) of Theorems 2 and 3. In fact, since the axioms are necessary and sufficient conditions, if one does not accept some of those axioms, he/she does not measure similarity by any function in the large family of those present in Theorems 1, 2, 3. On the other hand, every person which retains not adequate the Tversky paradigm, must be not agreeing with at least one of the related axioms.

Nevertheless, the function Φ, Ψ or Ξ representing the comparative similarity under the axioms introduced before is completely arbitrary. The next axioms we are going to introduce will represent "the price to pay" to obtain that the function representing $\preceq$ belongs to a particular class of similarity measures.

2.4 Independence axioms

The independence axioms are discussed in every theory, especially in decision-making and social choice: they essentially require that, under specific conditions, the preference relation between two objects is not influenced by some common component of the objects. We introduce now three independence axioms, which are weakened versions of the classical independence axiom, introduced in Tversky (1977).

Axiom WI [weak independence]
For every $(X_1, Y_1), (X_2, Y_2), (Z_1, W_1), (Z_2, W_2) \in \mathcal{X}^2$, if one of the following conditions holds:

1. $X_i \cap Y_i = Z_i \cap W_i \ \ (i = 1, 2)$ and $X_1 \Delta Y_1 = X_2 \Delta Y_2 \neq \underline{0}$, $Z_1 \Delta W_1 = Z_2 \Delta W_2 \neq \underline{0}$
2. $X_i \Delta Y_i = Z_i \Delta W_i \ \ (i = 1, 2)$ and $X_1 \cap Y_1 = X_2 \cap Y_2 \neq \underline{0}$, $Z_1 \cap W_1 = Z_2 \cap W_2 \neq \underline{0}$

then

$$(X_1, Y_1) \preceq (X_2, Y_2) \Leftrightarrow (Z_1, W_1) \preceq (Z_2, W_2).$$

Example 7 Continuing Example 1, we consider the apartments in the following table:

$\mathcal{H}$	a	b	c	d	e
X_1	1	0	1	1	0
Y_1	1	1	1	0	0
X_2	1	1	0	0	1
Y_2	1	0	0	1	1
Z_1	1	0	1	1	0
W_1	1	0	1	0	0
Z_2	1	0	0	1	1
W_2	1	0	0	0	1

Since $X_1 \cap Y_1 = Z_1 \cap W_1 = \{a, c\}$, $X_2 \cap Y_2 = Z_1 \cap W_2 = \{a, e\}$, $X_1 \Delta Y_1 = X_2 \Delta Y_2 = \{b, d\}$ and $Z_1 \Delta W_1 = Z_2 \Delta W_2 = \{d\}$, if a person considers $(X_1, Y_1) \preceq (X_2, Y_2)$, then he/she must consider also $(Z_1, W_1) \preceq (Z_2, W_2)$, vice versa if he/she sets $(X_2, Y_2) \prec (X_1, Y_1)$ then he/she must accept also $(Z_2, W_2) \prec (Z_1, W_1)$.

Axiom CI [cumulative independence]
For every $(X_1, Y_1), (X_2, Y_2), (Z_1, W_1), (Z_2, W_2) \in \mathcal{X}^2$, if one of the following conditions holds:

1. $X_i \cap Y_i = Z_i \cap W_i$, $X_i = Z_i \ (i = 1, 2)$ and $Y_1 = Y_2$, $W_1 = W_2$
2. $X_i \cap Y_i = Z_i \cap W_i$, $Y_i = W_i \ (i = 1, 2)$ and $X_1 = X_2$, $Z_1 = Z_2$
3. $X_i = Z_i$, $Y_i = W_i \ \ (i = 1, 2)$ and $X_1 \cap Y_1 = X_2 \cap Y_2$, $Z_1 \cap W_1 = Z_2 \cap W_2$

then

$$(X_1, Y_1) \preceq (X_2, Y_2) \Leftrightarrow (Z_1, W_1) \preceq (Z_2, W_2).$$

Axiom TWI [total weak independence]
For every $(X_1, Y_1), (X_2, Y_2), (Z_1, W_1), (Z_2, W_2) \in \mathcal{X}^2$, if one of the following conditions holds:

1. $(X_i \cap Y_i) \cup (X_i^c \cap Y_i^c) = (Z_i \cap W_i) \cup (Z_i^c \cap W_i^c)$, $X_i = Z_i$, $(i = 1, 2)$ and $Y_1 = Y_2$, $W_1 = W_2$
2. $(X_i \cap Y_i) \cup (X_i^c \cap Y_i^c) = (Z_i \cap W_i) \cup (Z_i^c \cap W_i^c)$, $Y_i = W_i \ \ (i = 1, 2)$ and $X_1 = X_2$, $Z_1 = Z_2$
3. $X_i = Z_i$, $Y_i = W_i \ \ (i = 1, 2)$ and $(X_1 \cap Y_1) \cup (X_1^c \cap Y_1^c) = (X_2 \cap Y_2) \cup (X_2^c \cap Y_2^c)$, $(Z_1 \cap W_1) \cup (Z_1^c \cap W_1^c) = (Z_2 \cap W_2) \cup (Z_2^c \cap W_2^c)$,

then

$$(X_1, Y_1) \preceq (X_2, Y_2) \Leftrightarrow (Z_1, W_1) \preceq (Z_2, W_2).$$

We now introduce structural axioms reinforcing the above independence conditions, which are the specifications in our environment of the *finite cancellation axiom*, introduced in Krantz et al. (1971). This axiom is related to binary relations $\preceq$ defined on a Cartesian product $\Theta = \bigtimes_{i=1}^{n} \Theta_i$ and requires that, for any k inequalities $H_j \preceq H'_j$ $(j = 1, \ldots, k)$, if for

every component Θ_i and for every $\theta_i \in \Theta_i$ the number of times that θ_i appears in the first elements of the inequalities $\preceq$ is the same as the number of times that θ_i appears in the second ones, this implies that for all the pairs H_j, H'_j $(j = 1, \ldots, k)$, the relation $\preceq$ is symmetrical, that is $\sim$.

The main result related to binary relations satisfying the finite cancellation axiom is the following theorem (see Krantz et al. 1971; Narens 1974):

Theorem 4 *Let $\Theta = \bigtimes_{i=1}^{n} \Theta_i$ be a finite set and $\preceq$ a reflexive binary relation in Θ. The following statements are equivalent:*

i) $\preceq$ satisfies the finite cancellation axiom;
ii) there exist n functions Φ_i (i=1, ...,n) from Θ_i to $\mathbb{R}$ (or to any other totally ordered set $\mathcal{C}$ containing $\mathbb{R}$) such that, if $H = (\theta_1, \ldots, \theta_n)$ and $H' = (\theta'_1, \ldots, \theta'_n)$ then

$$H \preceq H' \Leftrightarrow \sum_{i=1}^{n} \Phi_i(\theta_i) \leq \sum_{i=1}^{n} \Phi_i(\theta'_i).$$

To obtain this result, given a comparative similarity $\preceq$ on $\mathcal{X}^2$, let Γ be a set of ordered pairs (inequalities) $(X, Y) \preceq (Z, W)$ and let us consider the following sets: $\mathcal{A}^\Gamma = \{X \cap Y\}$, $\mathcal{B}^\Gamma = \{X \Delta Y\}$, $\mathcal{C}^\Gamma = \{(X^c \cap Y^c) \cup (X \cap Y)\}$, $\mathcal{D}_1^\Gamma = \{X\}$, $\mathcal{D}_2^\Gamma = \{Y\}$, with (X, Y) any element of $\mathcal{X}^2$ compared in at least one inequality in Γ. Then, we introduce new axioms adapting the cancellation axiom to this context.

Axiom WkC [weak kth cancellation]
Let $\preceq$ be a comparative similarity on $\mathcal{X}^2$. Let us associate with each pair $(X, Y) \in \mathcal{X}^2$ the pair $(X \cap Y, X \Delta Y)$. The relation $\preceq$ satisfies the *weak kth cancellation axiom* if and only if for each $h \leq k$ and for every inequalities $(X_i, Y_i) \preceq (Z_i, W_i)$, $(i = 1, \ldots, h)$ we have the following: if for every $A \in \mathcal{A}^\Gamma$ the number of pairs (X_i, Y_i) with $X_i \cap Y_i = A$ is the same as the number of pairs (Z_i, W_i) with $Z_i \cap W_i = A$, and similarly for every $B \in \mathcal{B}^\Gamma$, then for every $i = 1, \ldots, h$ we have

$$(X_i, Y_i) \sim (Z_i, W_i).$$

Proposition 3 *Let $\preceq$ be a comparative similarity on $\mathcal{X}^2$ satisfying Axiom S0. If $\preceq$ satisfies the weak second cancellation axiom, then $\preceq$ satisfies the weak independence axiom.*

Proof Let us consider $(X_i, Y_i), (Z_i, W_i)$, $(i = 1, 2)$ satisfying the hypotheses in Axiom WI and suppose $(X_1, Y_1) \preceq (X_2, Y_2)$ and $(Z_2, W_2) \preceq (Z_1, W_1)$. Since the hypotheses in Axiom WkC are fulfilled, it implies that the comparison between (Z_2, W_2) and (Z_1, W_1) cannot be strict, so $(Z_1, W_1) \preceq (Z_2, W_2)$. □

Axiom WFC [weak finite cancellation]
A comparative similarity $\preceq$ on $\mathcal{X}^2$ is said to satisfy the weak finite cancellation if and only if it satisfies the weak kth cancellation for each $k \in \mathbb{N}$.

Axiom CkC [cumulative kth cancellation]
Let $\preceq$ be a comparative similarity on $\mathcal{X}^2$ and let us associate with each pair $(X, Y) \in \mathcal{X}^2$ the triple $(X \cap Y, X, Y)$. The relation $\preceq$ satisfies the cumulative kth cancellation axiom if and only if for each $h \leq k$ and for every inequalities $(X_i, Y_i) \preceq (Z_i, W_i)$, $(i = 1, \ldots, h)$ we have the following: if for every $A \in \mathcal{A}^\Gamma$ the number of pairs (X_i, Y_i) with $X_i \cap Y_i = A$ is the same as the number of pairs (Z_i, W_i) with $Z_i \cap W_i = A$, and similarly for every $X \in \mathcal{D}_1^\Gamma$ and for every $Y \in \mathcal{D}_2^\Gamma$, then for every $i = 1, \ldots, h$ we have

$$(X_i, Y_i) \sim (Z_i, W_i).$$

Proposition 4 *Let $\preceq$ be a comparative similarity on $\mathcal{X}^2$ satisfying Axiom S0. If $\preceq$ satisfies the cumulative second cancellation axiom, then $\preceq$ satisfies the cumulative independence axiom.*

Proof The proof follows the same line of the proof of Proposition 3: consider $(X_i, Y_i), (Z_i, W_i)$, $(i = 1, 2)$ satisfying the hypotheses in Axiom CI and suppose $(X_1, Y_1) \preceq (X_2, Y_2)$ and $(Z_2, W_2) \preceq (Z_1, W_1)$. Since the hypotheses in Axiom CkC are fulfilled, it implies that the comparison between (Z_2, W_2) and (Z_1, W_1) cannot be strict, so $(Z_1, W_1) \preceq (Z_2, W_2)$. □

Axiom CFC [cumulative finite cancellation]
A comparative similarity $\preceq$ on $\mathcal{X}^2$ is said to satisfy the cumulative finite cancellation if and only if it satisfies the cumulative kth cancellation for each $k \in \mathbb{N}$.

Axiom TWkC [total weak kth cancellation]
Let $\preceq$ be a comparative similarity on $\mathcal{X}^2$ and let us associate with each pair $(X, Y) \in \mathcal{X}^2$ the triple $\big((X \cap Y) \cup (X^c \cap Y^c), X, Y\big)$. The relation $\preceq$ satisfies the cumulative kth cancellation axiom if and only if, for each $h \leq k$ and for every inequality $(X_i, Y_i) \preceq (Z_i, W_i)$, $(i = 1, \ldots, h)$, we have the following: if for every $C \in \mathcal{C}^\Gamma$ the number of pairs (X_i, Y_i) with $C = (X_i \cap Y_i) \cup (X_i^c \cap Y_i^c)$ is the same as the number of pairs (Z_i, W_i) with $C = (Z_i \cap W_i) \cup (Z_i^c \cap W_i^c)$, and similarly for every $X \in \mathcal{D}_1^\Gamma$ and for every $Y \in \mathcal{D}_2^\Gamma$, then for every $i = 1, \ldots, h$, we have

$$(X_i, Y_i) \sim (Z_i, W_i).$$

Proposition 5 *Let $\preceq$ be a comparative similarity on $\mathcal{X}^2$ satisfying Axiom S0. If $\preceq$ satisfies the total weak second cancellation axiom, then $\preceq$ satisfies the total weak independence axiom.*

Proof The proof follows the same line of the proof of Proposition 3: consider $(X_i, Y_i), (Z_i, W_i), (i = 1, 2)$ satisfying the hypotheses in Axiom TWI and suppose $(X_1, Y_1) \preceq (X_2, Y_2)$ and $(Z_2, W_2) \preceq (Z_1, W_1)$. Since the hypotheses in Axiom TWkC are fulfilled, it implies that the comparison between (Z_2, W_2) and (Z_1, W_1) cannot be strict, so $(Z_1, W_1) \preceq (Z_2, W_2)$. □

Axiom TWFC [total weak finite cancellation]
A comparative similarity $\preceq$ on $\mathcal{X}^2$ is said to satisfy the cumulative finite cancellation if and only if it satisfies the total weak kth cancellation for each $k \in \mathbb{N}$.

The following three propositions have the same role as those in the previous subsection: they extend the independence axioms from pairs with equal intersection, differences, symmetrical difference and so on, to all the pairs in which the above sets are not required to be equal, but only to have the same cardinality.

Let us consider now for any finite $\Gamma = \{(X_i, Y_i) \preceq (Z_i, W_i)\}$ the following sets: $\mathbf{a}^{\mathbf{\Gamma}} = \{|X \cap Y|\}$, $\mathbf{b}^{\mathbf{\Gamma}} = \{|X \Delta Y|\}$, $\mathbf{c}^{\mathbf{\Gamma}} = \{|X^c \cap Y^c| + |X \cap Y|\}$, $\mathbf{d_1^{\Gamma}} = \{|X|\}$, $\mathbf{d_2^{\Gamma}} = \{|Y|\}$, with (X, Y) any element of $\mathcal{X}^2$ compared in at least one inequality in Γ.

Lemma 7 *Let $\preceq$ be a comparative similarity on $\mathcal{X}^2$ satisfying Axioms S0, S4, S5 and WFC and for every $(X, Y) \in \mathcal{X}^2$, let us associate with (X, Y) the pair $(|X \cap Y|, |X \Delta Y|)$. For any number m of inequalities $(X_i, Y_i) \preceq (Z_i, W_i)$, $(i = 1, \ldots, m)$, if, for every $a \in \mathbf{a}^{\mathbf{\Gamma}}$, the number of pairs (X_i, Y_i) such that $a = |X_i \cap Y_i|$ is the same as the number of pairs (Z_i, W_i) with $a = |Z_i \cap W_i|$, and similarly for every $b \in \mathbf{b}^{\mathbf{\Gamma}}$, then for every $i = 1, \ldots, h$, we have*

$$(X_i, Y_i) \sim (Z_i, W_i).$$

Proof The proof follows from Lemma 2, Axiom WFC and by the same considerations as in the proof of Lemma 4. □

Lemma 8 *Let $\preceq$ on $\mathcal{X}^2$ satisfying Axioms S0, S4 and CFC and for every $(X, Y) \in \mathcal{X}^2$, let us associate with (X, Y) the triple $(|X \cap Y|, |X|, |Y|)$. For any number m of inequalities $(X_i, Y_i) \preceq (Z_i, W_i)$, $(i = 1, \ldots, m)$, if, for every $a \in \mathbf{a}^{\mathbf{\Gamma}}$, the number of pairs $(X_i \cap Y_i)$ such that $a = |X_i \cap Y_i|$ is the same as the number of pairs (Z_i, W_i) with $a = |Z_i \cap W_i|$, and similarly for every $d_1 \in \mathbf{d_1^{\Gamma}}$ and $d_2 \in \mathbf{d_2^{\Gamma}}$, then for every $i = 1, \ldots, h$, we have*

$$(X_i, Y_i) \sim (Z_i, W_i).$$

Proof The proof follows from Lemma 1 and Axiom CFC. □

Lemma 9 *Let $\preceq$ be a comparative similarity on $\mathcal{X}^2$ satisfying Axioms S0, S4, S6 and TWFC and for every $(X, Y) \in \mathcal{X}^2$, let us associate with (X, Y) the triple $(|X^c \cap Y^c| + |X \cap Y|, |X|, |Y|)$. For any number m of inequalities $(X_i, Y_i) \preceq (Z_i, W_i)$, $(i = 1, \ldots, m)$, if, for every $c \in \mathbf{c}^{\mathbf{\Gamma}}$, the number of pairs (X_i, Y_i) such that $c = |X_i^c \cap Y_i^c| + |X_i \cap Y_i|$ is the same as the number of pairs (Z_i, W_i) with $c = |Z_i^c \cap W_i^c| + |Z_i \cap W_i|$, and similarly for every $b \in \mathbf{b}^{\mathbf{\Gamma}}$, then for every $i = 1, \ldots, h$, we have*

$$(X_i, Y_i) \sim (Z_i, W_i).$$

Proof The proof follows from Lemma 3 and Axiom TWFC. □

3 Representation of comparative similarities by similarity measures

Among all similarity measures able to represent comparative similarities satisfying specific axioms, we focus on parametric functions having well-known similarity measures as particular cases. In this way, a link is established with existing measures and the designer is able to verify if the functional form he/she chooses to estimate similarities possesses appropriate properties.

We then consider three classes of similarity measures $S_{f,g}$, $T_{f,g,h}$, $H_{f,g}$, whose elements depend on two or three parameters, which are real functions f, g and h with specific characteristics such as the strict monotonicity.

- The class $S_{f,g}$ contains in particular Anderberg similarity (Anderberg 1973) (for $f(x) = 8x$ and $g(x) = x$), Jaccard similarity (Jaccard 1908) (for $f(x) = g(x) = x$), Sorensen similarity (Sorensen 1948) (for $f(x) = 4x$ and $g(x) = x$) and Dice similarity (Dice 1945) (for $f(x) = 2x$ and $g(x) = x$)
- the class $S_{f,g}$ contains in particular Ochiai similarity (Ochiai 1957) for $f(x) = x$, and $g(x) = h(x) = \sqrt{x}$)
- the class $H_{f,g}$ contains in particular Rogers and Tanimoto similarity (Rogers and Tanimoto 1960) (for $f(x) = x$, and $g(x) = 2x$), Sokal and Michener similarity (Sokal and Michener 1958) (for $f(x) = g(x) = x$) and Sokal and Sneath similarity (Sokal and Sneath 1963) (for $f(x) = x$, and $g(x) = \frac{1}{2}x$)

For more examples, refer to Lesot et al. (2009).

We first prove that the introduced axioms are satisfied by ordinal relations induced by different classes of similarity measures, i.e. each of them is necessary for the representability of comparative similarities by some specific class of

measures. This clearly explains that, if in a particular context, a person considers unacceptable some axiom among those necessary for the representation of the similarity relation by a specific class of similarity measures, then he/she can (should) not choose any element of this class to measure the similarity.

Lemma 10 *Any relation $\preceq$ on $\mathcal{X}^2$ induced (or representable) by a similarity measure $S_{f,g}$, defined as*

$$S_{f,g}(X,Y) = \begin{cases} \dfrac{f(|X\cap Y|)}{f(|X\cap Y|)+g(|X\Delta Y|)} & \text{for } (X,Y)\neq(\underline{0},\underline{0}) \\ 0 & \text{for } (X,Y)=(\underline{0},\underline{0}) \end{cases} \tag{2}$$

(where f and g are nonnegative increasing real functions with $f(0)=g(0)=0$), satisfies S0, S1, S2, S3, S4, S5, S7$_1$, WFC.

Proof To prove that $\preceq$ satisfies Axioms S0, S1, S2, S4, S5, S7$_1$, it is sufficient to prove that $S_{f,g}$ is a function satisfying properties of function Φ in Theorem 1. In fact $S_{f,g}(X,Y) = \Phi(|X\cap Y|, |X\Delta Y|)$. The non-negativity of f and g implies that the values taken by $S_{f,g}$ are in [0, 1]. Condition $f(0) = g(0) = 0$ and the definition of $S_{f,g}(\underline{0},\underline{0}) = 0$ ensures that $\Phi(0,b) = 0$ and $\Phi(a,0) = 1$, for every $b \neq 0$. Moreover, taking into account that f and g are increasing functions, it is immediate to see that $S_{f,g}$ is increasing with respect to $|X\cap Y|$ and decreasing with respect to $|X\Delta Y|$.

The proof that Axiom S3 is satisfied is in Corollary 2. We need to prove Axiom WFC. We first consider that, since $S_{f,g}$ represents $\preceq$, for any $(X,Y),(Z,W)\in\mathcal{X}^2$ we have:

$$\begin{aligned} S_{f,g}(X,Y) \le S_{f,g}(Z,W) &\Leftrightarrow f(|X\cap Y|)g(|Z\Delta W|) \\ &\le f(|Z\cap W|)g(|X\Delta Y|). \end{aligned}$$

Then, since ln is a strictly increasing function, we have:

$$\begin{aligned} &S_{f,g}(X,Y) \le S_{f,g}(Z,W) \\ &\quad\Leftrightarrow \ln[f(|X\cap Y|)] - \ln[g(|X\Delta Y|)] \\ &\quad\le \ln[f(|Z\cap W|)] - \ln[(|Z\Delta W|)]. \end{aligned}$$

Now consider the set $\Theta = \Theta_1\times\Theta_2 = \{|X_i\cap Y_i|\}\times\{|X_i\Delta Y_i|\}$, with $(X_i,Y_i)\in\mathcal{X}^2$ and the relation $\preceq'$ in Θ defined by putting for every $\theta_i,\theta_j\in\Theta$:

$$\theta_i \preceq' \theta_j \Leftrightarrow (X_i,Y_i)\preceq(X_j,Y_j).$$

Then, there exist two functions, $\Phi_1 = \ln(f)$, and $\Phi_2 = -\ln(g)$, respectively, defined in Θ_1 and Θ_2, such that the function $\Phi_1+\Phi_2$ represents $\preceq'$. From Theorem 4, we deduce that $\preceq'$ satisfies the finite cancellation axiom and so $\preceq$ satisfies WFC. □

Lemma 11 *Any relation $\preceq$ on $\mathcal{X}^2$ induced (or representable) by a similarity measure $T_{f,g,h}$, defined as*

$$T_{f,g,h}(X,Y) = \begin{cases} \dfrac{f(|X\cap Y|)}{g(|X|)h(|Y|)} & \text{for } X\neq\underline{0},\ Y\neq\underline{0} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

(where f, g, h are nonnegative increasing real functions with $f(0) = g(0) = h(0) = 0$ and $f(x) = g(x)h(x)$ for every real number x) satisfies S0, S1, S2, S4, S7$_2$, CFC.

Proof To prove that $\preceq$ satisfies Axioms S0, S1, S4, S7$_2$, it is sufficient to prove that $T_{f,g,h}$ is a function satisfying properties of function Ψ in Theorem 2. Since $|X\cap Y|+|X\setminus Y| = |X|$ and $|X\cap Y| + |Y\setminus X| = |Y|$. In fact $T_{f,g,h}(X,Y) = \Psi(|X\cap Y|, |X\setminus Y|, |Y\setminus X|)$. By the same consideration in proof of Lemma 10, we can prove conditions b and c of Theorem 2. We need to prove Axiom CFC. We first consider that, since $T_{f,g,h}$ represents $\preceq$, for any $(X,Y),(Z,W)\in\mathcal{X}^2$ we have:

$$\begin{aligned} T_{f,g,h}(X,Y) \le T_{f,g,h}(Z,W) &\Leftrightarrow f(|X\cap Y|)g(|Z|)h(|W|) \\ &\le f(|Z\cap W|)g(|X|)h(|Y|). \end{aligned}$$

Then, since ln is a strictly increasing function, we have:

$$\begin{aligned} &T_{f,g,h}(X,Y) \le T_{f,g,h}(Z,W) \\ &\quad\Leftrightarrow \ln[f(|X\cap Y|)] - (\ln[g(|X|)] + \ln[h(|Y|)]) \\ &\quad\le \ln[f(|Z\cap W|)] - \ln[g(|Z|)] - \ln[h(|W|)]. \end{aligned}$$

Now consider the set $\Theta = \Theta_1\times\Theta_2\times\Theta_3 = \{|X_i\cap Y_i|\}\times\{|X_i|\}\times\{|Y_i|\}$, with $(X_i,Y_i)\in\mathcal{X}^2$ and the relation $\preceq'$ in Θ defined by putting for every $\theta_i,\theta_j\in\Theta$:

$$\theta_i \preceq' \theta_j \Leftrightarrow (X_i,Y_i)\preceq(X_j,Y_j).$$

Then, there exist three functions, $\Psi_1 = \ln(f)$, $\Psi_2 = -\ln(g)$ and $\Psi_3 = -\ln(h)$, respectively, defined in Θ_1, Θ_2 and Θ_3, such that the function $\Psi_1+\Psi_2+\Psi_3$ represents $\preceq'$. From Theorem 4, we deduce that $\preceq'$ satisfies the finite cancellation axiom and so $\preceq$ satisfies WFC. □

Lemma 12 *Any relation $\preceq$ on $\mathcal{X}^2$ induced (or representable) by a similarity measure $H_{f,g}$, defined as*

$$H_{f,g}(X,Y) = \frac{f(|X\cap Y| + |X^c\cap Y^c|)}{f(|X\cap Y| + |X^c\cap Y^c|) + g(|X\Delta Y|)} \tag{4}$$

(where f, g are nonnegative increasing real functions with $f(0) = 0$) satisfies S0, S1, S3, S4, S6, S7$_3$, TWFC.

Proof The proof follows the same line as the proof of Lemmas 10 and 11, referring to the function Ξ in Theorem 3. □

We now prove that the necessary conditions are also sufficient for the representability of $\preceq$ by similarity measures of the above classes.

Theorem 5 *For a comparative similarity $\preceq$ on $\mathcal{X}^2$ the following conditions are equivalent:*

i) $\preceq$ satisfies S0, S1, S2, S3, S4, S5, S7$_1$, WFC;
ii) there exist two nonnegative increasing real functions f and g, with $f(0) = g(0) = 0$ such that the function $S_{f,g} : \mathcal{X}^2 \to [0, 1]$ defined in Eq. (2) represents $\preceq$.

Proof Implication $ii) \Rightarrow i)$ has been proved in Lemma 10, and we prove implication $i) \Rightarrow ii)$. First of all, by Axioms S0, S2, S4 and Corollary 1, we have $(\underline{0}, \underline{0}) \sim (X^c, X)$ for every $X \neq \underline{0}$, so we can consider $\preceq$ restricted to $\mathcal{X}^2 \setminus \{(\underline{0}, \underline{0})\}$ and then extend the function representing $\preceq$, by assigning to $(\underline{0}, \underline{0})$ the same value assigned to the pairs (X^c, X). Let us consider, as in the proof of Lemma 10, the set $\Theta = \Theta_1 \times \Theta_2$, with $\Theta_1 = \{|X_i \cap Y_i|\}$ and $\Theta_2 = \{|X_i \Delta Y_i|\}$, $(X_i, Y_i) \in \mathcal{X}^2$, and the relation $\preceq'$ in Θ defined by putting for every $\theta_i, \theta_j \in \Theta$:

$$\theta_i \preceq' \theta_j \Leftrightarrow (X_i, Y_i) \preceq (X_j, Y_j).$$

This is possible thanks to Lemma 2, which holds since $\preceq'$ satisfies Axioms S0, S4, S5. Then, every function representing $\preceq'$ also represents $\preceq$ and vice versa, moreover $\preceq'$ inherits all the properties of $\preceq$.

Now, since $\preceq'$ satisfies S0, S4, S5 and WFC and moreover Θ is finite, we can apply Lemma 7 and then Theorem 4 for $\mathcal{C} = \mathbb{R}^* = \mathbb{R} \cup \{-\infty, +\infty\}$ (that is the compactification of $\mathbb{R}$), totally ordered by extending the usual order of $\mathbb{R}$ in the natural way.

Therefore, there exist two functions $\Phi_i : \Theta_i \to \mathbb{R}^*$ $(i = 1, 2)$ such that the function

$$\Phi(|X \cap Y|, |X \Delta Y|) = \Phi_1(|X \cap Y|) + \Phi_2(|X \Delta Y|)$$

represents $\preceq'$, and so $G(X, Y) = \Phi(|X \cap Y|, |X \Delta Y|)$ represents $\preceq$. From Lemma 4, function Φ_1 is strictly increasing and function Φ_2 is strictly decreasing. By Axiom S2, $\Phi_2(0) = +\infty$: in fact S1 implies that for every $X \in \mathcal{X}$, $X \neq \underline{0}$ one has $(X, X) \sim (Y, Y)$, and so, if $x = |X \cap X| < |Y \cap Y| = y$, we must have $\Phi_1(x) + \Phi_2(0) = \Phi_1(y) + \Phi_2(0)$. So, since Φ_1 is strictly increasing, necessarily $\Phi_2(0) = +\infty$. With the same considerations, by Axiom S2, we have $\Phi_1(0) = -\infty$. Now we put $\Phi_2' = -\Phi_2$ and we consider the function $\ln : [0, +\infty] \to \mathbb{R}^*$ (defined by putting $\ln(0) = -\infty$ and $\ln(+\infty) = +\infty$) and the function $\exp : \mathbb{R}^* \to [0, +\infty]$. (defined by putting $\exp(-\infty) = 0$ and $\exp(+\infty) = +\infty$). By using these functions, we can write for every $(X, Y) \neq (\underline{0}, \underline{0})$:

$$G(X, Y) = \begin{cases} \ln \dfrac{\exp(\Phi_1(|X \cap Y|))}{\exp(\Phi_2'(|X \Delta Y|))} & \text{for } |X \Delta Y| \neq 0 \\ +\infty & \text{for } |X \Delta Y| = 0 \end{cases} \tag{5}$$

The function $G : \mathcal{X}^2 \setminus \{(\underline{0}, \underline{0})\} \to [-\infty, +\infty]$ represents $\preceq$ on $\mathcal{X}^2 \setminus \{(\underline{0}, \underline{0})\}$; taking into account the previous considerations, it can be extended to $\mathcal{X}^2$ by putting:

$$G'(X, Y) = \begin{cases} \ln \dfrac{\exp(\Phi_1(|X \cap Y|))}{\exp(\Phi_2'(|X \Delta Y|))} & \text{for } |X \Delta Y| \neq 0 \\ +\infty & \text{for } |X \Delta Y| = 0 \text{ and} |X \cap Y| \neq 0 \\ -\infty & \text{for } |X \Delta Y| = 0 \text{ and} |X \cap Y| = 0 \end{cases} \tag{6}$$

Since $G' : \mathcal{X}^2 \to \mathbb{R}^*$ represents $\preceq$ on $\mathcal{X}^2$, then any of its strictly increasing transformations φ also represents $\preceq$ on $\mathcal{X}^2$. In particular, $\varphi(\exp(G'))$ represents $\preceq$, with $\exp : \mathbb{R}^* \to [0, +\infty]$ and $\varphi : [-\infty, +\infty] \to [0, 1]$ defined as follows:

$$\varphi(z) = \begin{cases} \dfrac{z}{z+1} & \text{for } -\infty < z < +\infty \\ 0 & \text{for } z = -\infty \\ 1 & \text{for } z = +\infty \end{cases} \tag{7}$$

By putting $\exp(\Phi_1) = f$ and $\exp(\Phi_2') = g$, we obtain function $S_{f,g}$ as defined in Eq. 2. □

Theorem 6 *For a comparative similarity $\preceq$ on $\mathcal{X}^2$, the following conditions are equivalent:*

i) $\preceq$ satisfies S0, S1, S2, S4, S7$_2$, CFC;
ii) there exist three nonnegative increasing real-valued functions f, g, h, with $f(0) = g(0) = h(0) = 0$ and $f(x) = g(x)h(x)$ for every nonnegative number x, such that function $T_{f,g} : \mathcal{X}^2 \to [0, 1]$, defined in Eq. (3) represents $\preceq$.

Proof Implication $ii) \Rightarrow i)$ has been proved in Lemma 11, we prove implication $i) \Rightarrow ii)$. The proof follows the same line as the proof of Theorem 5, by considering this time $\Theta = \Theta_1 \times \Theta_2 \times \Theta_3$, with $\Theta_1 = \{|X_i \cap Y_i|\}$, $\Theta_2 = \{|X_i|\}$, $\Theta_3 = \{|Y_i|\}$, $(X_i, Y_i) \in \mathcal{X}^2$. Taking into account Lemmas 1, 8 and Theorem 4, we have three functions Ψ_1, Ψ_2, Ψ_3, taking values in $\mathbb{R}^*$, with Ψ_1 increasing and Ψ_2, Ψ_3 decreasing, such that

$$\Psi(|X \cap Y|, |X|, |Y|) = \Psi_1(|X \cap Y|) + \Psi_2(|X|) + \Psi_3(|Y|)$$

represents $\preceq'$, and so $G(X, Y) = \Psi(|X \cap Y|, |X|, |Y|)$ represents $\preceq$. Among the possible triples $\{\Psi_i\}$, we can choose one of them such that $\Psi_i(1) = 0$ for $i = 1, 2, 3$. This is possible since, if $\Psi = \Psi_1 + \Psi_2 + \Psi_3$ represents $\preceq$, $\Psi' = (\Psi_1 - k_1) + (\Psi_2 - k_2) + (\Psi_3 - k_3)$ also represents $\preceq$. From Axiom S1, by

following the same considerations as in the previous theorem, we have $\Psi_1(0) = -\infty$ and $\Psi_2(0) = \Psi_3(0) = +\infty$. Moreover, still from Axiom S1, we have that, for every $x, y > 0$, $\Psi_1(x)+\Psi_2(x)+\Psi_3(x) = \Psi_1(y)+\Psi_2(y)+\Psi_3(y)$. So, taking into account that $\Psi_1(1) = -\Psi_2(1) - \Psi_3(1) = 0$, necessarily $\Psi_1(x) = -\Psi_2(x) - \Psi_3(x)$. By putting $\Psi_i' = -\Psi_i$ $(i = 2, 3)$ and considering the functions exp and ln defined on $\mathbb{R}^*$ as in Theorem 5, we obtain the function $T_{f,g}(X, Y)$ defined by Eq. 3, with $\exp(\Psi_1) = f$, $\exp(\Psi_2') = g$ and $\exp(\Psi_3') = h$, with $\frac{f(x)}{g(x)h(x)} = 1$ for every x. The considerations about the extension to $\mathcal{X}^2$ of the function representing $\preceq$ on $\mathcal{X}^2 \setminus \{(\underline{0}, \underline{0})\}$, made in the proof of Theorem 5, also hold in this context. □

Taking now into account the total weak finite cancellation axiom, we obtain:

Theorem 7 *For a comparative similarity $\preceq$ on $\mathcal{X}^2$, the following conditions are equivalent:*

i) $\preceq$ satisfies S0, S1, S3, S4, S6, S7$_3$, TWCF;
ii) there exist three nonnegative increasing real-valued functions f, g, h with $f(0) = g(0) = h(0) = 0$ and $f(x) = g(x)h(x)$ for every nonnegative number x, such that function $H_{f,g} : \mathcal{X}^2 \to [0, 1]$, defined in Eq. (4) represents $\preceq$.

Proof Implication $ii) \Rightarrow i)$ has been proved in Lemma 12. The proof of implication $i) \Rightarrow ii)$ follows the same line as the proof of Theorem 5, by taking into account Lemmas 3, 6 and 9 instead of Lemma 2, 4 and 7, respectively, since Axiom S6 implies Axiom S5. The proof that Axiom S3 is satisfied is in Proposition 2. □

Remark 3 We notice that the functions f and g depend on the specific relations among the pairs; if we intend to obtain a particular function, then we need to require particular strong conditions for $\preceq$.

4 Utilisation in real-life problems

In this section, we sketch two possible utilisations of the results shown in the previous sections. The first one is a general discussion regarding the cognitive psychology field, where the results of this paper can help in discovering which rules are violated in the real experiments, and in the second one, we propose a real situation where a field expert can actively take part in finding the most suitable measure of similarity through a learning process, only based on the comparison (with respect to similarity) of pairs of profiles.

4.1 Cognitive psychology

Similarity and difference are fundamental concepts in cognition. As observed in Simmons and Estes (2008), they contribute to determine the recognition of known objects and the categorisation of novel objects. Moreover, they rule inferences about the features of an object and its predicted behaviour in a new context.

In the framework of cognitive psychology, similarity is described in terms of three kinds of information: features, structural relations and thematic relations. Tversky's contrast model is based on features and defines the similarity of two objects as a function of their common features weighed against their distinctive features. Similarity increases as the number of commonalities increases or the number of differences decreases.

Many studies and experiments explain that this model only based on features is not sufficient to determine the similarity between objects.

The first part of this paper investigates the rules we accept when we use a certain form of similarity measures, in the class of contrast models:

- for our comparative degree of similarity, we accept the rules represented by Axioms S0, S1, S2, S4, S5, S7$_1$ when we use any similarity measure of the following form:

  $$S(X, Y) = \Phi(|X \cap Y|, |X \Delta Y|),$$

  with Φ increasing with respect to $|X \cap Y|$ and decreasing with respect to $|X \Delta Y|$, such that $\Phi(0, b) = 0$, and $\Phi(a, 0) = 1$, for every a and for every $b \neq 0$;
- for our comparative degree of similarity, we accept the rules represented by Axioms S0, S1, S4, S7$_2$ when we use any similarity measure of the following form:

  $$S(X, Y) = \Psi(|X \cap Y|, |X \setminus Y|, |Y \setminus X|),$$

  with Ψ increasing with respect to $|X \cap Y|$ and decreasing with respect to $|X \setminus Y|$ and $|Y \setminus X|$, such that $\Psi(0, b, b') = 0$, and $\Psi(a, 0, 0) = 1$, for every a and for every b, b' with $b + b' \neq 0$;
- we finally accept the rules expressed by Axioms S0, S1, S4, S6, S7$_3$, when we use any similarity measure of this kind:

  $$S(X, Y) = \Xi(|X \cap Y|, |X^c \cap Y^c|, |X \Delta Y|),$$

  with Ξ increasing with respect to $|X \cap Y|$ and $|X^c \cap Y^c|$ and decreasing with respect to $|X \Delta Y|$, such that $\Xi(0, 0, b) = 0$, and $\Xi(a, c, 0) = 1$, for every a, c and for every $b \neq 0$.

The axioms are sufficient but also necessary for the representability of an ordering relation $\preceq$ among pairs of objects represented by a set of features with one of the above classes of measures. This means that they are equivalent to this representability and so, when the measures of the above classes only based on features are presumed not to be apt to determine the degree of similarity between objects, at least one of the axioms must not be accepted. It would therefore be easy to investigate which axioms fail, or to construct paradoxes (similar to the ones well known in decision theory) consisting in a set of comparisons to be submitted in an experiment.

4.2 A real-life example

As it is possible to read in its website, the World Anti-Doping Agency (WADA) is an international independent agency, composed and funded equally by the sport movement and governments of the world, whose mission is to lead a collaborative worldwide movement for doping-free sport. Its key activities include scientific research, education, development of anti-doping capacities, and monitoring of the World Anti-Doping Code (Code) the document harmonising anti-doping policies in all sports and all countries. A recent important document (https://www.wada-ama.org/en/resources/the-code/world-anti-doping-code), (see also Dvoraki et al. 2014) stressed the importance to do laboratory and physician tests no longer randomly but targeted, setting up an early alert system, based on the global symptoms of an athlete. The Italian Ministry of Health has funded a project (see the Acknowledgements section for details) for developing mathematical models and algorithms for doping contrasts, where one of the themes was their own early alert.

During this project, several "field experts" were interviewed and it was noted that they expressed their opinion more easily by ordering pairs in which an element represents a positive (or negative) case of a doping athlete and the other element is non-tested athlete. On the contrary, they have no idea of which similarity measure is the most appropriate for this task. (Actually, they have no idea of similarity measurement.)

In such a situation, the first problem to be solved is the identification of the most suitable measure to represent the idea of similarity expressed by the expert through a comparative similarity.

As an example, we report here the sketch of a virtual procedure for finding the most appropriate measure of similarity, expressing the idea of "no more similar than" of a sports doctor and the relevant study of the numerical measure agreeing with his comparative similarity. This procedure can be implemented by combining two different methodologies: presenting direct explicit queries on the simplest axioms and preparing simple fictitious pairs of athletes profiles, focusing on some axioms, and requiring to the expert to compare these pairs in "similarity". This last method permits to test which axioms are violated and which are accepted in the expressed similarity ordering.

Let $\{x_1, \ldots, x_n\}$ be a set of characteristic symptoms or acute effects related to a specific substance prohibited in-competition (for instance, stimulants, cannabinoids or glucocorticoids), which are assumed to be only present or absent in an athlete (i.e. their degree of membership can be only either 0 or 1). This assumption can appear to be a drastic simplification of the context to be modelled, but it is actually the best option in the absence of an expert able to elicit the degree of presence of the various symptoms.

Just as an example, suppose that we submit to a sports doctor a set of profiles related to possible users of glucocorticoids, characterised by the following macro symptoms (attributes):

x_1) thin and fragile skin;
x_2) hypertrichosis (enhanced hair growth)
x_3) iatrogenic Cushing's syndrome symptoms (moon face, buffalo hump, central obesity);
x_4) oral candidiasis;
x_5) mood swings;
x_6) euphoria;
x_7) depression.

and let us consider a data base consisting in the following profiles:

$\mathcal{H}$	x_1	x_2	x_3	x_4	x_5	x_6	x_7
X_1	1	0	0	0	0	1	0
X_2	1	1	0	0	1	0	1
X_3	0	1	0	1	0	0	1
X_4	0	0	1	1	1	1	1
X_5	1	1	0	0	1	1	1
X_6	1	0	1	1	1	1	0
X_7	1	0	1	0	1	0	1
X_8	0	0	1	1	0	0	1
X_9	0	1	1	0	0	0	1
X_{10}	1	0	1	1	1	0	1
X_{11}	1	1	1	0	1	0	1
X_{12}	1	1	0	0	0	1	0
X_{13}	1	1	0	0	1	0	0
X_{14}	0	1	0	0	0	0	1
X_{15}	0	1	1	1	0	0	0
X_{16}	1	0	1	1	1	0	0
X_{17}	0	1	0	1	1	1	1

The doctor is required to provide a binary relation among pairs of profiles, expressing his comparative degree of similarity (representing his opinion that the elements of a pair are no more similar than the elements of another pair).

First of all, his complete acceptance of the transitivity of this relation must be tested, explaining that, without transitivity, no real function can represent the comparative similarity.

Axiom S1 can also be directly discussed with the doctor since it is very intuitive and easily accepted by everyone.

To complete the analysis of the axioms necessary for the representability of a comparative similarity by any element of the three classes considered in this paper, i.e. $S_{g,f}$, $T_{g,f}$ and $H_{g,f}$, we need to test the propensity of the doctor to accept Axiom S4. To achieve this, we present to the doctor the pairs (X_1, X_2), (X_{14}, X_{15}), which only differ with respect to a permutation of the indexes, and we ask him to order them in similarity.

If he does not consider the pairs equally similar (violating in this way the thesis of Proposition 1 and so Axiom S4), then there is no similarity measure considered in this paper representing his comparative similarity (see Theorems 5, 6, 7).

On the contrary, in the case where the judgement is $(X_1, X_2) \sim (X_{13}, X_{14})$, we only need to explain that the equivalence must not be casual, but must be based on the awareness of the equal contribution of the symptoms (more generally the attributes) to the similarity.

If he accepts, we can proceed with the process of discovering the most suitable similarity measure representing the doctor's idea of "no more similar than".

At this point, the following pairs

$$\{(X_1, X_2), (X_3, X_2), (X_4, X_5)\}$$

are submitted for a comparative evaluation in similarity.

The pairs are such that, for two of them, the number of symptoms present in only one athlete is the same, while for the third pair it is smaller. More precisely, $|X_2 \Delta X_3| = |X_1 \Delta X_2| < |X_4 \Delta X_5|$ (note that for the common symptoms we have $|X_1 \cap X_2| < |X_2 \cap X_3| < |X_4 \cap X_5|$). If the doctor assigns the following ordering

$$(X_4, X_5) \prec (X_1, X_2) \sim (X_3, X_2) \tag{8}$$

it is very probable that his numerical model of reference belongs to the class of similarity measures defined by Eq. (4). In fact, the comparative similarity in (8) agrees with the result of Lemma 6 and so it does not violate Axioms S0, S4, S5, S6, $S7_3$; indeed, it does not agree with the result of both Lemmas 4 and 5, so that Axioms $S7_1$, $S7_2$ are violated: they are in fact necessary conditions for the representability of $\preceq$ by every element of the class defined by (2) or (3), respectively (see Lemmas 10 and 11).

At this moment, it is necessary to make the doctor aware of the fact that accepting Axioms S0, S4, S5, S6, $S7_3$ is equivalent to accepting that the similarity between two objects is evaluated by a function only depending on the number of attributes present in only one profile and decreasing with respect to this number (see Theorem 3).

It is important to stress that, in this framework, the joint attributes and the attributes absent in both profiles have the same relevance with respect to the similarity. Conversely, to choose as a similarity measure any function of such a class means to accept all of the axioms mentioned above.

If the doctor continues to agree with all the previous axioms, it only remains to test if he also agrees with the total weak finite cancellation axiom which determines the form of the function Ξ in Theorem 3.

To this end, we first ask him if his order between (X_1, X_{17}) and (X_3, X_{17}) is the same as the order between (X_1, X_2), (X_3, X_2), already considered equivalently similar. A negative answer (meaning the negation of the total weak second cancellation axiom, due to the non-validity of the thesis of Lemma 9) precludes the possibility to represent the doctor's comparative similarity by any element of the class in Eq. (4).

We note that a negative answer underlines the idea that the symptoms x_i are considered not independent, but their mutual influences (more precisely, their positive or negative interactions) are taken into account in the similarity evaluation (for a similarity measure capturing these instances, see Baioletti et al. 2012, for the crisp case, and Coletti et al. 2017, for the fuzzy case).

Indeed, if the answer of the doctor is positive, then we can explore the total weak finite cancellation through more complex examples, but this could be in fact unnecessary, since the basic idea and the relevant semantic is intrinsically present in the total second cancellation axiom strictly connected with the total weak independence axiom (see Proposition 5).

Suppose now that the doctor does not provide the ordering in Eq. (8). In this case, his numerical model is surely not a function Ξ as defined in Theorem 3 and so it does not belong to the class defined by Eq. (4). In fact, from Lemma 6, the ordering in (8) necessarily follows from Axioms S0, S4, S5, S6, $S7_3$, necessary conditions in Theorem 3 and so in Lemma 12. Then, we need to know which axioms among S5, S6, $S7_3$ are not accepted by the doctor.

In particular, it is very probable that he does not consider equally important with respect to the similarity the presence and the absence of a symptom in both the elements of a pair (what Axioms S6 and $S7_3$ require). The natural next step is then to ask to the doctor to order in similarity the following pairs: (X_1, X_{12}), (X_2, X_{13}), (X_{10}, X_{11}), (X_{15}, X_{16}), which have $|X_{10} \Delta X_{11}| = |X_1 \Delta X_{12}| < |X_2 \Delta X_{13}| = |X_{15} \Delta X_{16}|$ and $|X_{10} \cap X_{11}| > |X_1 \cap X_{12}| = |X_2 \cap X_{13}| = |X_{15} \cap X_{16}|$. If the doctor agrees with the following ordering

$$(X_2, X_{13}) \prec (X_1, X_{12}) \prec (X_{10}, X_{11}) \sim (X_{15}, X_{16}), \tag{9}$$

it is very probable that his numerical model of reference is the class defined in Eq. (2), in fact the comparative similarity in (9) agrees with the result in both Lemmas 2 and 4. So, it does not violate Axioms S0, S4, S5, $S7_1$. We can proceed as in the previous case to test in two steps the axiom of weak finite cancellation, starting from the weak second cancellation axiom.

In the case where the doctor's comparative similarity is different from the one given in Eq. (9), we can think that, for the doctor, the symptoms present in the first element of the pair are not combinable with those present in the second and absent in the first: for instance, two pairs, having the same symptoms in common, can be considered not equally similar if the first has all the remaining symptoms only present in one element and the second one has half of the symptoms only present in one element and the other half only in the other.

In this last case, we propose the pairs (X_6, X_7), (X_7, X_8), (X_4, X_9) which have $|X_6 \cap X_7| > |X_7 \cap X_8| = |X_4 \cap X_9|$, $|X_6 \setminus X_7| = |X_7 \setminus X_8| < |X_4 \setminus X_9|$ and $|X_7 \setminus X_6| = |X_8 \setminus X_7| < |X_9 \setminus X_8|$. If the doctor agrees with the following ordering

$$(X_4, X_9) \prec (X_7, X_8) \prec (X_6, X_7), \tag{10}$$

then he agrees with Lemma 4 and so in particular with Axiom 7_2.

At this moment, it is necessary to make the doctor aware of the fact that accepting Axioms S0, S4, S5, $S7_2$ is equivalent to accepting that the similarity between two objects is evaluated by a function depending on the number of attributes in common to the two profiles (and increasing with respect to it) and on the number of the attributes only present in the first profile and those only present in the second one, decreasing with respect to each of these numbers, possibly with a different functional low (see Eq. 3).

At this point, we only need to test the axiom of cumulative second cancellation, to arrive to the finite cumulative cancellation axiom and then to the class defined in Eq. (3) as a numerical model agreeing with the doctor's comparative similarity.

In the case, the doctor's comparative similarity is different from the one given in Eq. (10), then no element of the three proposed classes can respect the idea of similarity expressed by the doctor in his ordering relation.

5 Conclusion

In the paper, we have followed the idea of Tversky (1977) of studying similarity in the environment of the theory of measurements (Krantz et al. 1971). In this framework, we introduced a binary relation on a set of pairs of objects, expressing a *comparative degree of similarity*. We studied the representability of this comparative similarity by means of classes of numerical similarity measures, containing as particular elements some of the best known similarity measures. For this purpose, we established axioms, which are necessary and sufficient conditions under which a given comparative similarity is represented by a specific class of similarity measures.

This study stresses the fact that all the measures belonging to a class have the same behaviour, since they comply with the same comparative model and then respect the same structural properties. So the aim of the paper is not to choose one specific measure among those of a class, as made, for instance, in Choi et al. (2010), but instead to consciously choose a class containing all the measures which fix the same set of ordering relations and submit the remaining to the same constraints.

We note that most axioms are devoted to ensuring that the considered similarity measures representing the relation only depend on the cardinality of the intersection and differences of the two objects. Starting from the seminal work of Tversky Tversky (1977), and in more recent literature, see, for instance, Bertoluzza et al. (2004), Bouchon-Meunier et al. (2008), Coletti and Di Bacco (1989), Lesot et al. (2009), this is usually a prior assumption not supported by conditions on the comparative structure. For that the paper aims to provide a contribution to the discussion on the foundations of the measurement of similarity, presenting an easily understandable qualitative framework which characterises Tversky's contrast model (Tversky 1977), which defines the similarity of two objects as a function depending only on their common and distinctive features, increasing as the number of commonalities and increasing on the number of differences.

In a second part of this research, in the companion paper (Coletti and Bouchon-Meunier 2018), we will consider a more realistic model for the same problem. In it the features (or characteristics), often expressed in natural language or by an imprecise or vague sentence, can be present with different degrees of membership, leading to the extension of the results to a fuzzy framework, where the attribute can be present with a degree of membership different from 0 and 1. We will show that, in this context, the equivalent axioms strictly depend on the t-norm and t-conorm used to define the operations between fuzzy sets.

Funding Giulianella Coletti work was partially supported by Perugia University, funding of 2016 Research Projects, under grant: "Decisions under risk, uncertainty and imprecision", by the Italian Ministry of Health under Grant `J521I14001640001` ("Intelligent systems helping in decisions for the early alert and the dissuasion to the use of doping").

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Anderberg MR (1973) Cluster analysis for applications. Academic Press, New York

Baioletti M, Coletti G, Petturiti D (2012) Advances in computational intelligence: 14th international conference on information processing and management of uncertainty in knowledge-based systems, IPMU 2012, Catania, Italy, July 9–13, 2012, Proceedings, Part III, Chapter. Weighted attribute combinations based similarity measures. Springer, Berlin, pp 211–220

Bertoluzza C, Di Bacco M, Doldi V (2004) An axiomatic characterization of the measures of similarity. Sankhya 66:474–486

Bhutani KR, Rosenfeld A (2003) Dissimilarity measures between fuzzy sets or fuzzy structures. Inf Sci 152:313–318

Boriah S, Chandola V, Kumar V (2008) Similarity measures for categorical data: a comparative evaluation. In: Proceedings of the 8th SIAM international conference on data mining, SIAM, pp 243–254

Bouchon-Meunier B, Rifqi M, Bothorel S (1996) Towards general measures of comparison of objects. Fuzzy Sets Syst 84:143–153

Bouchon-Meunier B, Rifqi M, Lesot MJ (2008) Similarities in fuzzy data mining: from a cognitive view to real-world applications. In Zurada J, Yen G, Wang J (eds) Computational intelligence: research frontiers. WCCI 2008, vol 5050. Springer, LNCS, pp 349–367

Bouchon-Meunier B, Coletti G, Lesot MJ, Rifqi M (2009) Towards a conscious choice of a similarity measure: a qualitative point of view. In: Sossai C, Ghemello G (eds) Symbolic and quantitative approaches to reasoning with uncertainty: Ecsqaru 2009 proceedings, vol 5590. Springer, LNAI, pp 542–553

Bouchon-Meunier B, Coletti G, Lesot MJ, Rifqi M (2010) Towards a conscious choice of a fuzzy similarity measure: a qualitative point of view. In: Hllermeier E, Kruse R, Hoffmann F (eds) Computational intelligence for knowledge-based system design: IPMU 2010 proceedings, vol 6178. Springer, LNAI, pp 1–10

Choi S-S, Cha S-H, Tappert CC (2010) A survey of binary similarity and distance measures. J Syst Cybern Inf 8(1):43–48

Coletti G, Bouchon-Meunier B (2018) A study of similarity measures through the paradigm of measurement theory: the fuzzy case. Soft-Computing **(submitted)**

Coletti G, Di Bacco M (1989) Qualitative characterization of a dissimilarity and concentration index. Metron XLVII:121–130

Coletti G, Petturiti D, Vantaggi B (2017) Fuzzy weighted attribute combinations based similarity measures. In: Proceedings of ECSQARU 2017 (Symbolic and quantitative approaches to reasoning with uncertainty), vol 10369. LNCS, pp 364–374

Couso I, Garrido L, Sànchez L (2013) Similarity and dissimilarity measures between fuzzy sets: a formal relational study. Inf Sci 229:122–141

Cross VV, Sudkamp TA (2002) Similarity and compatibility in fuzzy set theory: assessment and applications. Studies in fuzziness and soft computing, vol 93. Springer, Berlin

Dice LR (1945) Measures of the amount of ecological association between species. Ecology 26:297–302

Dvoraki J, Baume N, Botré Broséus J, Budgett R, Frey WO, Geyer H, Harcourt PR, Ho D, Howman D, Isola V, Lundby C, Marclay F, Peytavin A, Pipe A, Pitsiladis YP, Reichel C, Robinson N, Rodchenkov G, Saugy M, Sayegh S, Segura J, Thevis M, Vernec A, Viret M, Vouillamoz M, Zorzoli M (2014) Time for change: a roadmap to guide the implementation of the World Anti-Doping Code 2015. Br J Sports Med: BJSM 48:801–806

Filev P, Hadjiiski L, Sahiner B, Chan HP, Helvie MA (2005) Comparison of similarity measures for the task of template matching of masses on serial mammograms. Med Phys 32(2):515–529

Gilboa I, Schmeidler D (1995) Case-based decision theory. Q J Econ 110:605–639

Gilboa I, Schmeidler D (1997) Act similarity in case-based decision theory. Econ Theory 9:47–61

Gilboa I, Lieberman O, Schmeidler D (2006) A similarity-based approach to prediction. Rev Econ Stat 162(1):124–131

Ha V, Haddawy P (2003) Similarity of personal preferences: theoretical foundations and empirical analysis. Artif Intell 146:149–173

Hahn U, Ramscar M (eds) (2001) Similarity and categorization. Oxford University Press, Oxford

Hwang CM, Yang MS, Hung WL, Lee MG (2012) A similarity measure of intuitionistic fuzzy sets based on the Sugeno integral with its application to pattern recognition. Inf Sci 189:93–109

Jaccard P (1908) Nouvelles recherches sur la distribution florale. Bull Soc Vaud Sci Nat 44:223–270

Krantz D, Luce R, Suppes P, Tversky A (1971) Foundations of measurement, vol I. Academic Press, New York

Lesot MJ, Rifqi M (2010) Order-based equivalence degrees for similarity and distance measures. In: Hllermeier E, Kruse R, Hoffmann F (eds) Computational intelligence for knowledge-based systems design. IPMU 2010, vol 6178. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, pp 19–28

Lesot MJ, Rifqi M, Benhadda H (2009) Similarity measures for binary and numerical data: a survey. Int J Knowl Eng Soft Data Paradig (KESDP) 1:63–84

Li Y, Qin K, He X (2014) Some new approaches to constructing similarity measures. Fuzzy Sets Syst 234:46–60

Narens L (1974) Minimal conditions for additive conjoint measurement and qualitative probability. J Math Psychol 11:404–430

Ochiai A (1957) Zoogeographic studies on the soleoid fishes found in Japan and its neighbouring regions. Bull Jpn Soc Sci Fish 22:526–30

Pelillo M (ed) (2013) Similarity-based pattern analysis and recognition. Advances in computer vision and pattern recognition. Springer, London

Penney GP, Weese J, Little JA, Desmedt P, Hill DLG, Hawkes DJ (1998) A comparison of similarity measures for use in 2-D-3-D medical image registration. In: Proceedings of MICCAI 1998: medical image computing and computer-assisted intervention MICCAI98, vol. 1496. LNCS, pp 1153–1161

Rissland E (2006) AI and similarity. IEEE Intell Syst 21:33–49

Rogers DJ, Tanimoto TT (1960) A computer program for classifying plants. Science 132:1115–1118

Sokal RR, Michener C (1958) A statistical method for evaluating systematic relationships. Univ Kansas Sci Bull 38:1409–1438

Sokal RR, Sneath PHA (1963) Priciples of numerical taxonomy. W.H. Freeman, San Francisco

Sorensen T (1948) A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. K Dan Vidensk Selsk Biol Skr 5:1–34

Simmons S, Estes Z (2008) Individual differences in the perception of similarity and difference. Cognition 106(3):781–795

Suppes P, Krantz D, Luce R, Tversky A (1989) Foundations of measurement, vol II. Academic Press, New York

Toussaint GT (2004) A comparison of rhythmic similarity measures. In: Proceedings 5th international conference on music information retrieval

Tversky A (1977) Features of similarity. Psychol Rev 84:327–352

WADA https://www.wada-ama.org/en/resources/the-code/world-anti-doping-code

Zhang Z, Huang K, Tan T (2006) Comparison of similarity measures for trajectory clustering in outdoor surveillance scenes. In: Proceedings of 18th international conference on pattern recognition (ICPR'06). IEEE. https://doi.org/10.1109/ICPR.2006.392