Analogical proportion-based induction: from classification to creativity

Henri Prade, Gilles Richard

ANALOGICAL PROPORTION-BASED INDUCTION: FROM CLASSIFICATION TO CREATIVITY

HENRI PRADE
IRIT – CNRS, 118, route de Narbonne, 31062 Toulouse Cedex 9, France *
henri.prade@irit.fr

GILLES RICHARD
IRIT – CNRS, 118, route de Narbonne, 31062 Toulouse Cedex 9, France
gilles.richard@irit.fr

Abstract

The first aim of this article is to position analogical inference (or at least a particular form of it) in relation to induction. After a brief reminder on the induction of plausible conclusions in a probabilistic, logical, possibilistic settings, and with J. S. Mill's methods of induction, we turn our attention to analogical inference, based on analogical proportions. Analogical proportions that hold between Boolean vectors are emphasized as a matter of pairs belonging to the same equivalence class. Then the mechanism of analogical proportions-based classification is explained and the main algorithms and results obtained so far are surveyed. After which, steps towards a logic of creativity are presented. The approach starts from the observation that analogical proportions belong to a larger set of quaternary relations called logical proportions. The six logical proportions giving birth to an equivalence relation between pairs are identified. This includes two important cases: i) a logic of conditional events known as being a basis for non monotonic reasoning (which is a form of plausible deduction) ; ii) a logic of ordered pairs preserving positive changes, closely related to analogical proportions. Within this framework, we revisit the creative nature of analogical proportions and introduce a creative inference mechanism that works on the basis of a specific situation and a collection of ordered pairs representing possible changes.

Special thanks are due to two anonymous referees for their careful reading and their suggestions to improve clarity and avoid confusion.

*This research has been supported by the ANR project "Analogies: from theory to tools and applications" (AT2TA), ANR-22-CE23-002.

1 Introduction

As it is well-known, Charles Sanders Peirce [44] distinguished between three main forms of reasoning: deduction (for deriving necessary consequences), induction (for extrapolating and generalizing from facts), and abduction (for finding out an hypothesis explaining a particular situation).

Peirce [45] was viewing analogical reasoning as a composite form of reasoning that combines induction and abduction (or retroduction, using Peirce's word). Roughly speaking, the idea is to see the fact of relating a given situation to another known situation as a kind of abduction, and the fact of projecting / extrapolating what happens in the latter situation onto the given situation (in order to predict something about it) as a form of induction. In fact, Peirce did not detail the way the two modes of reasoning are combined in analogical reasoning; see [41] and [42] for different understandings of Peirce's suggestion that may also involve deduction.

Indeed analogical reasoning involves drawing a plausible inference that a property found (or holding) in one situation is likely to also apply/hold to/in a second situation when there are significant similarities between the two situations. Such a parallel between two situations is also the starting point of Gentner [27]'s structure-mapping theory, where analogy is viewed as a mapping from one source domain to a target domain which conveys that a system of relations which holds among the source entities also holds among the target entities. Such a view was also proposed in [57]. A restricted rendering of this idea, stated in logical terms, making a parallel between two entities x and y, amounts to infer $Q(y)$ from $P(x), Q(x)$, and $P(y)$ (where P and Q are predicates applicable to situations x and y). Such an inference is clearly an "analogical jump", which offers no guarantee on the conclusion. Some authors [14, 54] looked for what could be added to such premises in order to ensure the truth of the conclusion. But this view is too demanding since it amounts to reduce analogical inference to a form of deduction.

This view of the "analogical jump" pattern of inference is not so far from case-based reasoning [25]. Indeed in this pattern, a "case" y under consideration where P is true is related to a known "case" x where P is true as well[1] (i.e., we have a perfect similarity between x and y) as well as another property Q also true for x. So as in case-based reasoning, this suggests that Q may be true as well for y.

In this paper, we use a slightly different view of analogical inference that is based on analogical proportions. Analogical proportions are statements of the form "a is to b as c is to d", as for instance, "a calf is to a cow as a foal is to a mare". Although it involves four items, we can observe that a calf and a cow are bovids, which corresponds to a first situation, while a foal and a mare are equids, which is the second situation. In each sit-

[1] Note that P may be a compound property involving a number of elementary properties.

uation, properties or relation such as "young", "adult", "mother" apply to the entities put in correspondence by the analogical proportion. More generally the pair (a, b) with all the properties and relations attached to items a and b is paralleled with the pair (c, d) with all the properties and relations attached to items c and d. A philosopher, Mary Hesse [30], pointed out and discussed the link between analogy viewed as a parallel between two situations and analogical proportions.

In the following, we shall use an inference based on analogical proportions. Although its link with the "analogical jump" pattern (as a particular instance) has been established [8], this inference is different from the one in case-base reasoning where the known cases in the repertory are considered one by one in isolation [50][2]. In contrast, in analogical proportion-based inference, the case under consideration, say d, is related to a triplet a, b, c of known cases, each case being described by a set of attribute values [36]. .

Analogical inference is known to be useful in a wide range of applications [28], from aiding explanation [5] to aiding creativity, e.g., in mathematics [47]. In this paper, we shall review how analogical proportion-based inference can be applied to classification tasks, and we shall advocate a new view of analogical proportions that will lead us to propose some elements of a new logic that manipulates ordered pairs of vectors, and that may be used in a creativity task.

The paper is organized as follows. Section 2 discusses analogical inference as a form of induction, together with a brief review of some non conventional form of induction. Section 3 provides a formal survey of analogical proportions. It emphasizes a new view of (Boolean) analogical proportions that express an *equivalence* relation between two ordered pairs of vectors describing items in terms of Boolean features. Section 4 recalls analogical inference (based on analogical proportions) and the main results obtained with this inference in classification. Section 5 first briefly restates the general setting of so-called logical proportions, to which analogical proportions belong. We focus on those logical proportions that define equivalence relations. We thus identify two families of such proportions, one which defines equivalences between "conditional objects" and which is at the basis of nonmonotonic reasoning, and the other which is related to analogical proportions and from which we outline a logic of ordered pairs, applicable to creativity issues.

2 Analogical inference and transduction

Induction is classically opposed to deduction. While deduction applies generic knowledge (represented by general rules ("all men are mortal") to factual information ("Socrate is a

[2]Usually in case-based reasoning the case under consideration y is put in similarity relation with several known cases $x_1, \cdots, x_k$ where each x_i leads to a partial conclusion for y (provided that the similarity between x_i and y is judged sufficient), and the partial conclusions have then to be combined in some way.

man")[3], induction works in the opposite direction, attempting to infer general laws from a limited set of observed facts. Mathematical models of induction predominantly stem from probability and statistics theories. Sometimes what is covered by the word 'induction' also includes the deductive step which corresponds to the application of the general laws induced to a factual situation.

But there is another form of induction named *transduction* [26] that directly infers a particular factual conclusion from a set of data. A well-known transduction-based mechanism is the k-nearest neighbor method where the prediction for a new item is only based on the observation of closed neighbors. It is also known as lazy learning. In both cases, induction is a form of inference that provides plausible conclusions, unlike deduction, which provides safe conclusions.

Let us consider a classification problem. Given a set of items or entities (objects, situations, profiles, ...) that are all described in terms of the values of a collection of observable attributes applicable to all of them, and that belong to a known class, the classification problem amounts to assign a class to a new item whose class is unknown. More formally, let S be a set of m items, each one $\vec{a^i}$ is represented by a vector of n attribute values $\vec{a^i} = (a^i_1, a^i_2, \cdots, a^i_n)$ $i = 1, ..., m$, together with its class/label $cl(\vec{a^i}) \in \mathcal{C}$, where $\mathcal{C}$ is a finite set of labels: for instance $\mathcal{C} = \{good, bad\}$, or $\mathcal{C} = \{bovid, equid, canid\}$. The set of attributes used to describe an observable piece of data is fixed: for instance $\{color, age, weight, position, \ldots\}$. Thus, a class gathers items of the same kind in a categorization process. Each class $C \in \mathcal{C}$ divides the set of items into the $\vec{a^i}$'s that are examples of C if $\vec{a^i} \in C$, and those that are counter-examples for C (and examples of other classes). In general, attributes can take into account values of various types, such as integers, real numbers, words, and more. In a context of binary attributes (i.e., attributes with values belonging to $\{0, 1\}$), attributes can be regarded as properties, for instance, representing whether an observable individual is wealthy (1 for 'rich,' 0 for 'not rich').

Given a new item $\vec{a^\star} = (a^\star_1, a^\star_2, \cdots, a^\star_n) \notin S$, the problem is then to predict its class.

When applying Bayesian classification with the assumption that the attributes are statistically independent of each other given the class, we calculate, for each $C \in \mathcal{C}$:

$$Prob(C|\vec{a^\star}) = \frac{Prob(\vec{a^\star}|C) \cdot Prob(C)}{Prob(\vec{a^\star})} = \frac{1}{Prob(\vec{a^\star})} \cdot Prob(C) \cdot \prod_{k=1}^{n} Prob(a^\star_k|C)$$

and $\vec{a^\star}$ is assigned to the class $C \in \mathcal{C}$ providing the highest value for $Prob(C|\vec{a^\star})$.

The probabilistic approach is the prevailing viewpoint for induction and transduction. Still there exist other options. Let us briefly review them.

[3] Even if deduction can also infer new generic rules from given rules.

Possibility theory [19] also has a Bayesian-like rule, which writes [20],

$$\Pi(C|\vec{a^\star}) \circledast \Pi(\vec{a^\star}) = \Pi(\vec{a^\star}|C) \circledast \Pi(C)$$

where $\circledast = \min$ or $\circledast =$ product. These two options separate the theories of qualitative and quantitative possibilities where conditioning is based on min and product respectively [19]. In case of product-based conditioning on obtains the expression

$$\Pi(C|\vec{a^\star}) = \frac{\Pi(\vec{a^\star}|C)}{\max_{j:C^j \in \mathcal{C}} \Pi(a^\star|C^j)}$$

under the hypothesis of no prior information, i.e., $\Pi(C) = 1$. More generally $\forall j, \Pi(C^j) = 1$, since $\Pi(\vec{a^\star}) = \max_j \Pi(a^\star|C^j) \cdot \Pi(C^j)$. The result is very close to Edwards' notion of likelihood [22], who advocates a non-Bayesian view, i.e., without priors. See, for instance, [7] for experiments with possibilistic classification.

If we use the $\min$-based qualitative conditioning, i.e., $\circledast = \min$, assuming that class C is fully possible (no prior, i.e., $\Pi(C) = 1$), and taking $\Pi(\vec{a^\star}) = 1$, the equality $\Pi(C|\vec{a^\star}) \circledast \Pi(\vec{a^\star}) = \Pi(\vec{a^\star}|C) \circledast \Pi(C)$ reduces to $\Pi(C|\vec{a^\star}) = \Pi(\vec{a^\star}|C)$ and finally with an hypothesis of non-interactivity (logical independence) of attributes we obtain

$$\Pi(C|\vec{a^\star}) = \min_{k=1}^{n} \Pi(a_k^\star|C)$$

It expresses that C is all the more a possible class for $\vec{a^\star}$ as all its attribute values are possible in class C, or better that the less possible is one of the attribute values for the class C, the less possible class C is for $\vec{a^\star}$. We then come closer to a logical analysis of data that we examine now.

Considering a particular class $C \in \mathcal{C}$, a simple logical reading for Boolean data sets can be done from the sets of examples $\mathcal{E}_C$ and counter-examples $\mathcal{E'}_C$. Let $\varphi(C)$ be a logical formula that describes class C, i.e., $\varphi(C)$ is true for the description of any item belonging to $\varphi(C)$. Then we have [21]

$$\bigvee_{i:\vec{a^i} \in \mathcal{E}} (a_1^i \wedge a_2^i \wedge \cdots \wedge a_n^i) \models \varphi(C) \models \bigwedge_{j:\vec{a'^j} \in \mathcal{E}'} (\neg a_1'^j \vee \neg a_2'^j \vee \cdots \vee \neg a_n'^j)$$

This means that the description of any example in terms of the n Boolean attributes makes $\varphi(C)$ true, and that any model of $\varphi(C)$ falsifies at least one attribute value of any counter-example. Provided that the data are not noisy, this provides a consistent bracketing of $\varphi(C)$. This looks like the version space approach [43] where the hypotheses space is bracketed between a lower and an upper bound computed from the examples and counter-examples, except that here no representation bias is introduced.

In such a view, examples and counter-examples are exploited one by one. This is also the case for the possibilistic approach, although differently. In other words, the informational contributions made by each example (or counter-example) are combined, but no comparison of examples belonging to the same class, or to different classes, is made. This contrasts, as already mentioned in the introduction, with the analogical proportion-based inference where items are handled three by three ; see Sections 3 and 4 for details. For their part, k-nearest neighbors methods [24, 55], which are based on ideas similar to the ones at work in case-based reasoning, also handle examples one by one (even if there is some cumulative counting across the k examples considered).

The XIXth English philosopher, logician and economist, John Stuart Mill is known, among many other things, for his "methods of induction". He indeed proposed five methods of induction in his 1843 treatise of logic [40]. Strictly speaking, it is a matter of abduction rather than induction, but it was long before Charles Sanders Peirce distinguished between the two notions! Indeed these "induction" methods look for the simplest and most likely hypothesis that explains some observations. But what contrasts Mill's methods with the previously reviewed approaches to induction is that they heavily rely on the ideas of *agreement* and *difference*, which makes them somewhat similar in that respect to analogical proportion-based inference, where examples are compared within pairs, as we shall see in Section 3.

We cite here only the two main methods. The first one, called the *(Direct) Method of Agreement* is stated like that [40]:

$ABCD$ occur together with $wxyz$
$AEFG$ occur together with $wtuv$

Therefore A is the cause, or the effect, of w

Mill is not at all precise about what the letters in his induction patterns refer to. However the causality flavor suggests to regard letters as properties or attributes that characterize circumstances and their consequences (hence the use of two kinds of letters). The number of attributes involved in the two first statements that differ from one to the other has no particular meaning here. What is important here is that two states of affairs $ABCDwxyz$ on the one hand and $AEFGwtuv$ on the other hand are *compared*.

In other words, what $\{A, B, C, D\}$ and $\{A, E, F, G\}$ have in common, i.e., $\{A\}$), corresponds to what $\{w, x, y, z\}$ and $\{w, t, u, v\}$ have in common, i.e., $\{w\}$.

A similar analysis applies to the *Method of Difference*, which reads [40]:

$ABCD$ occur together with $wxyz$
BCD occur together with xyz

Therefore A is the cause, or the effect, or a part of the cause of w

Observe in the second pattern of inference that the set differences between $\{A, B, C, D\}$ and $\{B, C, D\}$ correspond to the set differences between $\{w, x, y, z\}$ and $\{x, y, z\}$.

Still Mill was making no link between his Methods of Difference and Agreement and his view of analogy, although he considered that analogical reasoning (viewed according to the pattern discussed in the introduction) was also a form of induction. The reader is referred to [4] for some further discussions on links between Mill's methods of induction and the analogical proportions-based inference that we are going to present in the next section.

3 Analogical proportions

This section provides a refresher on Boolean-valued analogical proportions together with a new view in terms of an equivalence relation between pairs.

3.1 Postulates

Generally speaking, analogical proportions are statements of the form "a is to b as c is to d" linking four entities, which are supposed to satisfy the following postulates, according to a long tradition that dates back to Aristotle. These postulates were inspired by a parallel with numerical proportions, namely, arithmetic proportions ($a - b = c - d$) and geometric proportions ($\frac{a}{b} = \frac{c}{d}$) which equalize differences and ratios respectively. Thus, proportions operates a double comparison (inside, and between, pairs) as also suggested by the expression of an analogical proportion "a is to b as c is to d". Such a proportion, considered here as a relation between 4 items, is denoted by $a : b :: c : d$.

These postulates, which are the only ones classically associated to an analogical proportion, are:

P1 *reflexivity*: $a : b :: a : b$;

P2 *symmetry*: $a : b :: c : d \Rightarrow c : d :: a : b$;

P3 stability under *central permutation*: $a : b :: c : d \Rightarrow a : c :: b : d$.

As immediate consequences, an analogical proportion also satisfies

- $a : a :: b : b$ (sameness) ;
- $a : b :: c : d \Rightarrow d : b :: c : a$ (external permutation);
- $a : b :: c : d \Rightarrow b : a :: d : c$ (internal reversal);
- $a : b :: c : d \Rightarrow d : c :: b : a$ (complete reversal).

It is also worth noticing that the 3 postulates *do not allow* for other permutations such that $a : b :: c : d \Rightarrow b : a :: c : d$, or $a : b :: c : d \Rightarrow c : b :: a : d$. In particular, $a : b :: b : a$ is not supposed to hold. Thus one cannot say that "black is to white as white is to black" (even if this might be advocated on the basis that the relation of opposition between 'black' and 'white' is the same as the one between 'white' and 'black'; see [37] for a discussion of a relation-based view of analogical proportions, maybe closer to natural language practice.

The entities involved in an analogical proportion can be of different natures: it may be numbers, words, drawings, images, sentences, ... [52]. This may lead to question the validity of the central permutation postulate, and to replace it by a weaker postulate such as, e.g., the internal reversal property ; see [37, 2] for discussions.

It is assumed in this article that the entities considered can be represented by vectors of Boolean feature values. Moreover, we require that the features used in the representation are applicable to the four entities involved in the analogical proportions we deal with.[4] Note that in this paper we are primarily interested in the inference mechanism associated with the logical modeling of analogical proportions between Boolean-valued vectors. Even if we use some analogical proportion stated in natural language for illustration, we do not intent to discuss analogical proportions between words in ordinary language here ; see [37] on this point.

Thus, any entity is represented here by a vector $\vec{a} = (a_1, ..., a_n)$ where a_i is the value of feature (or attribute) i. We define the analogical proportion relation among n-tuples by applying it componentwise. Namely,

$$\vec{a} : \vec{b} :: \vec{c} : \vec{d} \text{ if and only if } \forall i \in [1, n], a_i : b_i :: c_i : d_i$$

We need now to recall the definition of an analogical proportion for four Boolean variables a_i, b_i, c_i, d_i representing the value of some feature i for four distinct items.

3.2 Boolean proportions

The reflexivity postulate $a : b :: a : b$ forces a Boolean analogical proportion to be true for any values of a and b in $\{0, 1\}$, and therefore an analogical proportion $a : b :: c : d$ is true for the valuations $(0, 0, 0, 0)$, $(0, 1, 0, 1)$, $(1, 0, 1, 0)$, and $(1, 1, 1, 1)$. The unique *minimal* Boolean model that satisfies the three postulates P1, P2, P3 is made up of the 6 valuations shown in Table 1, where the valuations $(0, 0, 1, 1)$, $(1, 1, 0, 0)$ are added due to P3. As can be seen the 6 patterns are symmetrical (i.e., satisfy P2). The $2^4 - 6 = 10$ other valuations

[4] This assumption excludes analogical proportions such as "beer is to the English what wine is to the French" where two different conceptual universes are present (beverages and people, in the example). Note that the central permutation does not hold for such analogical proportions, which require a more sophisticated modeling [3].

a	b	c	d
0	0	0	0
1	1	1	1
0	0	1	1
1	1	0	0
0	1	0	1
1	0	1	0

Table 1: Minimal Boolean model of $a : b :: c : d$

are excluded, namely $a : b :: c : d$ is false for $(0, 1, 1, 0)$, for $(1, 0, 0, 1)$, for the 4 valuations with only one 0, and for the 4 valuations with only one 1 [51]:

There are several remarkable quaternary logical formulas for an analogical proportion $a : b :: c : d$, all of which are logically equivalent. Thus, they are all true only for the 6 valuations of Table 1 (and false for the 10 remaining valuations). The first formula uses *dissimilarity* indicators only, inside pairs (a, b) and (c, d). Indeed we have [39]:

$$a : b :: c : d = ((a \wedge \neg b) \equiv (c \wedge \neg d)) \wedge ((\neg a \wedge b) \equiv (\neg c \wedge d)) \quad (1)$$

It precisely expresses that "a differs from b as c differs from d and b differs from a as d differs from c" (and "when a and b do not differ, c and d do not differ").

A second formula, logically equivalent to expression (1), is [39]:

$$a : b :: c : d = ((a \wedge d) \equiv (b \wedge c)) \wedge ((\neg a \wedge \neg d) \equiv (\neg b \wedge \neg c)) \quad (2)$$

It uses *similarity* indicators only and can be read as "what a and d have in common (positively or negatively) b and c have it also"[5].

Another expression, equivalent to the two above logical formulas [51], which is closer to the description of the contents of Table 1, is given by

$$a : b :: c : d = ((a \equiv b) \wedge (c \equiv d)) \vee ((a \equiv c) \wedge (b \equiv d))$$

Beyond the three postulates P1, P2, P3 and their consequences, the minimal Boolean model of an analogical proportion (described in Table 1) also satisfies two noticeable properties [48]:

- *code independence*: $a : b :: c : d \Rightarrow \neg a : \neg b :: \neg c : \neg d$;

[5] It also says that "when a and d differ (one is true, the other is false) then b and c also differ"! Rewriting it as $a : b :: c : d = ((a \wedge d) \equiv (b \wedge c)) \wedge ((a \vee d) \equiv (b \vee c))$, emphasizes that the conjunctions of the extremes and of the means are equivalent, as well as their disjunctions.

- *transitivity*: $(a : b :: c : d) \wedge (c : d :: e : f) \Rightarrow a : b :: e : f$.

The first property expresses that any feature can be encoded positively or negatively without harming the analogical proportion.

The second property that expresses transitivity does not follow from the postulates either, but is true for Boolean variables.[6]. It has an important consequence (together with reflexivity and symmetry properties): The analogical proportion $a : b :: c : d$ defines *an equivalence relation* between the ordered pairs (a, b) and (c, d) in the Boolean setting.

The description of items may involve *nominal* attributes, i.e., attributes with a finite domain $\mathcal{A}$ whose cardinality is larger than 2. Then $a : b :: c : d$ holds for nominal variables if and only if (as first suggested in [46])

$$(a, b, c, d) \in \{(s, s, s, s), (s, t, s, t), (s, s, t, t) \mid s, t \in \mathcal{A}\} \tag{3}$$

When the cardinality of $\mathcal{A}$ is equal to 2, we retrieve the Boolean model. As can be checked, (3) is the unique nominal model that satisfies P1, P2, and P3. All the properties discussed above still hold for nominal attributes.

Let us take the example mentioned in the Introduction "a calf is to a cow as a foal is to a mare". The animals there can be described in terms of attributes such as `mammal`, `carnivore`, `young`, `adult`, `ruminant`, `single-toed`, or `family` (the last attribute being nominal). The vector describing each animal is given horizontally in Table 2. We can see vertically that a perfect analogical proportion holds component by component.

	mammal	*carnivore*	*young*	*adult*	*ruminant*	*single-toed*	*family*
`calf`	1	0	1	0	1	0	bovidae
`cow`	1	0	0	1	1	0	bovidae
`foal`	1	0	1	0	0	1	equidae
`mare`	1	0	0	1	0	1	equidae

Table 2: A calf is to a cow as a foal is to a mare

[6]Some readers might object that analogical proportions may not be transitive. In a general context, their observation is valid, and this issue becomes more apparent when dealing with multiple attributes. Specifically $\vec{a} : \vec{b} :: \vec{c} : \vec{d}$ may hold with respect to some attributes and $\vec{c} : \vec{d} :: \vec{e} : \vec{f}$ may hold with respect to a different set of attributes leading to a failure of transitivity, as in the following abstract example. Assume $\vec{a}, \vec{b}, \vec{c}, \vec{d}, \vec{e}, \vec{f}$ can be described in terms of 4 Boolean attributes i_1, i_2, i_3, i_4, and $\vec{a} = (1, 1, 0, 0)$, $\vec{b} = (1, 1, 1, 0)$, $\vec{c} = (1, 0, 0, 0)$, $\vec{d} = (1, 0, 1, 1)$, $\vec{e} = (1, 1, 1, 0)$, and $\vec{f} = (1, 1, 1, 1)$. Let us denote by $(\vec{a} : \vec{b} :: \vec{c} : \vec{d})_S$ the fact that the analogical proportion holds componentwise for all attributes $i \in S$. Then it can be easily checked that $(\vec{a} : \vec{b} :: \vec{c} : \vec{d})_{\{i_1, i_2, i_3\}}$ holds as well as $(\vec{c} : \vec{d} :: \vec{e} : \vec{f})_{\{i_1, i_2, i_4\}}$, while $(\vec{a} : \vec{b} :: \vec{e} : \vec{f})_{\{i_1, i_2, i_3, i_4\}}$ does not hold. Still, it can be observed that here transitivity is preserved if we restrict ourselves to the set $S = \{i_1, i_2\}$.

Note that as soon as $\vec{a}, \vec{b}, \vec{c}, \vec{d}$ have at least two components, $\vec{a} : \vec{b} :: \vec{c} : \vec{d}$ can hold with 4 distinct vectors: for instance, in dimension 2, $\vec{a} = (0,0), \vec{b} = (0,1), \vec{c} = (1,0), \vec{d} = (1,1)$ build a proper analogical proportion.

3.3 Analogical proportions in terms of sets

As already said, the entities or items involved in analogical proportions that we consider are represented by vectors of Boolean values, such as $\vec{a} = (a_1, ..., a_n)$, and analogical proportions between vectors are defined componentwise:

$$\vec{a} : \vec{b} :: \vec{c} : \vec{d} \Leftrightarrow \forall i \in \{1, \cdots, n\},\ a_i : b_i :: c_i : d_i$$

To a vector $\vec{a} = (a_1, ..., a_n)$ where $\forall i, a_i \in \{0,1\}$, one can associate the set of features $A = \{i \in \{1, \cdots, n\} \mid a_i = 1\}$ possessed by $\vec{a}$. Clearly, A and $\vec{a}$ are equivalent representations. In the same way, we associate $\vec{b}$, $\vec{c}$ and $\vec{d}$ with B, C and D. The set representation provides a maybe more intuitive view of an analogical proportion.

Indeed, it has been noticed [56] (see also [3]) that $A : B :: C : D$ holds if and only if it exists *non overlapping* subsets U, V, X, Y, and Z, such that

- $A = U \cup X \cup Z$;
- $B = U \cup Y \cup Z$;
- $C = V \cup X \cup Z$;
- $D = V \cup Y \cup Z$.

This makes clear that $A \setminus B = C \setminus D = X$ and $B \setminus A = D \setminus C = Y$, which is in agreement with equation (1). Note also that $A \setminus C = B \setminus D = U$ and $C \setminus A = D \setminus B = V$ in agreement with the stability under central permutation.[7]

Moreover, with this set representation, it can be easily checked that an analogical proportion $A : B :: C : D$ holds as soon as

A: made of what is common to A and C together with what is common to A and B
B: made of what is common to B and D together with what is common to A and B
C: made of what is common to A and C together with what is common to C and D
D: made of what is common to B and D together with what is common to C and D.

[7] This is also in agreement with equation (2) since $A \cap D = B \cap C = Z$ and $A \cup D = B \cup C = U \cup V \cup X \cup Y \cup Z$.

where 'made of' refers to union operation and "what is common" to intersection operation. Interestingly enough, the four descriptions of A, B, C, D above make a perfect analogical proportion, term by term (applying nominal definition (3))!

Remark. The set-based view of analogical proportion allows us to better suggest a relationship between Mill's rules of induction and analogy. If we reconsider Mill's pattern of (Direct) Method of Agreement as the analogical proportion (keeping the notations of Section 2) "$\mathcal{A} = ABCD$ is to $\mathcal{B} = wxyz$ as $\mathcal{C} = AEFG$ is to $\mathcal{D} = wtuv$" (where 'is to' is understood as 'co-occurs with'), we can observe that indeed $\mathcal{A} \setminus \mathcal{C} = \mathcal{C} \setminus \mathcal{A} = A$ co-occurs with $\mathcal{B} \setminus \mathcal{D} = \mathcal{D} \setminus \mathcal{B} = w$. Mill's Method of Difference could be read similarly in an analogical proportion manner in spite of the presence of items of two different natures in those patterns.

3.4 Analogical proportions as equivalence relations between pairs

Analogical proportions are a matter of i) comparing items inside an ordered pair, and then. ii) pairing pairs $(\vec{a}, \vec{b})$ and $(\vec{c}, \vec{d})$. Let us examine these two steps.

Let $\vec{a} = (a_1, ..., a_n)$, $\vec{b} = (b_1, ..., b_n)$, etc. be items described by means of n Boolean attributes or features. Given two vectors $\vec{a}$, $\vec{b}$, their comparison leads to consider the subsets of attributes where they are equal (to 1 or to 0), and the subsets of attributes where they differ (by going from 0 to 1, or from 1 to 0), when we go from $\vec{a}$ to $\vec{b}$. This leads to define

$Equ^0(\vec{a}, \vec{b}) = \{i \mid a_i = b_i = 0\}$,
$Equ^1(\vec{a}, \vec{b}) = \{i \mid a_i = b_i = 1\}$,
$Equ(\vec{a}, \vec{b}) = \{i \mid a_i = b_i\} = Equ^0(\vec{a}, \vec{b}) \cup Equ^1(\vec{a}, \vec{b})$,
and
$Dif^{10}(\vec{a}, \vec{b}) = \{i \mid a_i = 1, b_i = 0\}$,
$Dif^{01}(\vec{a}, \vec{b}) = \{i \mid a_i = 0, b_i = 1\}$,
$Dif(\vec{a}, \vec{b}) = \{i \mid a_i \neq b_i\} = Dif^{01}(\vec{a}, \vec{b}) \cup Dif^{10}(\vec{a}, \vec{b})$.

This allows us to state the following result:

$$\vec{a} : \vec{b} :: \vec{c} : \vec{d} \text{ if and only if } \begin{cases} Equ(\vec{a}, \vec{b}) = Equ(\vec{c}, \vec{d}) \\ Dif^{10}(\vec{a}, \vec{b}) = Dif^{10}(\vec{c}, \vec{d}) \\ Dif^{01}(\vec{a}, \vec{b}) = Dif^{01}(\vec{c}, \vec{d}) \end{cases}$$

We see that what matters in an analogical proportion is the orientation of the differences, whereas it does not matter with which value the equality is realized. Table 3 highlights the structure of an analogical proportion, in three subsets of attribute(s), one where the 4 items are equal, one where they are equal within the pairs, but not in the same way, and finally the

subset of attribute(s) whose value(s) change(s), in the same direction, from $\vec{a}$ to $\vec{b}$ and from $\vec{c}$ to $\vec{d}$.

$items$	$All\ equal$	$Equality\ by\ pairs$	$Change\ (Dif)$
$\vec{a}$	1 0	1 0	1 0
$\vec{b}$	1 0	1 0	0 1
$\vec{c}$	1 0	0 1	1 0
$\vec{d}$	1 0	0 1	0 1

Table 3: The 3 parts of analogical proportion and the associated valuations

As shown in Table 3, the set of attributes with which the four items involved in an analogical proportion $a : b :: c : d$ are supposed to be represented can be partitioned in three subsets corresponding to the way the attribute values are possibly modified from an item to another item. As we can see, the central permutation of $\vec{b}$ and $\vec{c}$ exchanges the content of the columns "Equality by pairs" and "Change" (but does not affect the "All equal" column). Neither of these two subsets must be empty if we want the analogical proportion to be non-trivial, i.e., $\vec{a}$, $\vec{b}$, $\vec{c}$, $\vec{d}$ are distinct vectors (for $n = 2$, $\vec{a} = (1, 1)$, $\vec{b} = (1, 0)$, $\vec{c} = (0, 1)$, $\vec{d} = (0, 0)$ realize an analogical proportion with distinct vectors, as already said). Besides, the subset of attribute(s) "All equal" can be empty. If the subset "Equality by pairs" or the subset "Change" is empty, then $\vec{a} = \vec{c}$ and $\vec{b} = \vec{d}$ or $\vec{a} = \vec{b}$ and $\vec{c} = \vec{d}$ respectively.

Besides, analogical proportions are implicitly present when comparing two items represented with the same set of n attributes ($n \geq 2$): From a formal viewpoint, for any two distinct vectors $\vec{a}$ and $\vec{d}$ *differing on at least two attributes*, there exist two other distinct vectors $\vec{b}$ and $\vec{c}$ such as $\vec{a} : \vec{b} :: \vec{c} : \vec{d}$ [12]. This does not mean that these two vectors $\vec{b}$ and $\vec{c}$ represent items existing in the real world.

As already noticed when dealing with one component, analogical proportions express an equivalence relation between two ordered pairs. Given four distinct vectors $\vec{a}, \vec{b}, \vec{c}, \vec{d}$, we have the following result:

Two ordered pairs $(\vec{a}, \vec{b})$ and $(\vec{c}, \vec{d})$ are in the same equivalence class if and only if[8]

1. $Dif(\vec{a}, \vec{b}) = Dif(\vec{c}, \vec{d})$;
2. $\forall j \in Dif(\vec{a}, \vec{b})\ a_j = c_j$ and $b_j = d_j$.

Condition 1 ensures that the change concerns the same attributes in both pairs, condition 2 that the change applies in the same direction in both pairs. It is clear that any two pairs $(\vec{a}, \vec{b})$ and $(\vec{c}, \vec{d})$ taken in the same equivalence class together form an analogical proportion

[8] A further condition should be added, namely $Dif(\vec{a}, \vec{b}) \neq \emptyset$ and $\exists i\ a_i \neq c_i$ in case the vectors might not be distinct.

$\vec{a} : \vec{b} :: \vec{c} : \vec{d}$. This notion of equivalence class is similar to the idea of "analogical cluster" introduced in [34] in a context of computational linguistics.

4 Analogical proportions-based inference and classification

In the Boolean and nominal cases, analogical inference relies on the solving of analogical equations, i.e. finding $\vec{x}$ such that $\vec{a} : \vec{b} :: \vec{c} : \vec{x}$ holds, working component by component.[9]. This kind of extrapolation is a counterpart of the "rule of three" based on geometric proportions $x = \frac{b \cdot c}{a}$ (its arithmetic counterpart is $x = b + c - a$).

Since a triplet $a, b, c \in \{0, 1\}^3$ may take $2^3 = 8$ values, while $a : b :: c : d$ is true only for six distinct 4-tuples, there are cases where the equation $a : b :: c : x$ in the Boolean case has no solution. Indeed, the equations $1 : 0 :: 0 : x$ and $0 : 1 :: 1 : x$ have no solution. It is easy to prove that the Boolean analogical equation $a : b :: c : x$ is solvable if and only if $(a \equiv b) \vee (a \equiv c)$ holds true. In that case, the *unique* solution is given by $x = a \equiv (b \equiv c)$; thus $x = b$ if $a = c$ and $x = c$ if $a = b$.

The situation in the nominal case is quite similar: $s : t :: t : x$ has no solution (for $s \neq t$). Only the equations $s : t :: s : x$ and $s : s :: t : x$ are solvable, with unique solution $x = t$. In the nominal case, where s, t, u can take more than 2 values, the equation $s : t :: u : x$ is also not solvable as soon as s, t, u are distinct.

Analogical proportion-based inference, as described by the inference rule (4), applies to classification and relies on a simple principle: in the Boolean or nominal cases, if four vectors $\vec{a}$, $\vec{b}$, $\vec{c}$ and $\vec{d}$ make a valid analogical proportion component-wise for each 4-tuple of values pertaining to the same attribute, then it is expected that their class labels also make a valid proportion ([58], see also [9]).

$$\frac{\vec{a} : \vec{b} :: \vec{c} : \vec{d}}{cl(\vec{a}) : cl(\vec{b}) :: cl(\vec{c}) : cl(\vec{d})} \tag{4}$$

Assuming that the class labels for vectors $\vec{a}$, $\vec{b}$ and $\vec{c}$ are known (i.e., they belong to the sample set), the classification of a new Boolean or nominal vector $\vec{d}$ is only possible i) when the equation $cl(\vec{a}) : cl(\vec{b}) :: cl(\vec{c}) : x$ is solvable (since a Boolean or a nominal equation may have no solution if the equation is of the form $s : t :: t : x$ or $s : t :: u : x$), and ii) the analogical proportion $\vec{a} : \vec{b} :: \vec{c} : \vec{d}$ holds true on all components. If these two conditions are met, we take $cl(\vec{d})$ as the unique solution for x.

Clearly, the inference rule (4) offers no guarantee on the truth of the prediction for $cl(\vec{d})$; this prediction may be only regarded as a plausible conclusion, just as in the case of

[9] The three other equations $\vec{x} : \vec{b} :: \vec{c} : \vec{d}$, $\vec{a} : \vec{x} :: \vec{c} : \vec{d}$, $\vec{a} : \vec{b} :: \vec{x} : \vec{d}$ can be equivalently stated in the above form (with x in d position), applying internal and complete reversal properties.

the "analogical jump" inference (recalled in the introduction), of which (4) can be shown to a particular instance [9]. Note that this inference rule is a transduction rule that infers a factual conclusion about the class of a new item from a set of items and their class. It is an extrapolation process [35].

However as already said, there is no guarantee that the conclusion of the inference rule is not erroneous, and it may happen that different predictions coexist for $cl(\vec{d})$ as shown in Table 4, where three pairs $(\vec{a}, \vec{b})$, $(\vec{c}, \vec{d})$, $(\vec{a'}, \vec{b'})$, belonging to the same equivalence class (since each pair has the same Dif-pattern) are put in parallel leading to divide the equality part of the pairs into four subparts $Equ_{sss}, Equ_{sst}, Equ_{sts}, Equ_{stt}$ corresponding to the different possible arrangements of these equality parts while preserving analogical proportions between the first two pairs and between the last two.[10] Indeed, on the one hand, one can check that $\vec{a} : \vec{b} :: \vec{c} : \vec{d}$ holds, and the analogical inference yields $cl(\vec{d}) = cl(\vec{b}) = t$; on the other hand, it can be seen that $\vec{a'} : \vec{b'} :: \vec{c} : \vec{d}$ holds also, and the analogical inference then gives $cl(\vec{d}) = cl(\vec{c}) = s$. This takes place in spite of the fact the pairs $(\vec{a}, \vec{b})$ and $(\vec{a'}, \vec{b'})$ are in the same class of equivalence.

$case$	Equ_{sss}	Equ_{sst}	Equ_{sts}	Equ_{stt}	Dif	$class$
$\vec{a}$	s_1	s_2	s_3	s_4	s_5	s
$\vec{b}$	s_1	s_2	s_3	s_4	t_5	t
$\vec{c}$	s_1	s_2	t_3	t_4	s_5	s
$\vec{d}$	s_1	s_2	t_3	t_4	t_5	t / s
$\vec{a'}$	s_1	t_2	s_3	t_4	s_5	s
$\vec{b'}$	s_1	t_2	s_3	t_4	t_5	s

Table 4: Inconsistent prediction

This situation of inconsistent predictions is very general. The unique exception (where the situation is impossible) is when the classification function is a linear function in case of Boolean attributes [11], or is quasi-linear in case of nominal attributes [13]. Still this does not mean that those special cases are the only cases where the analogical proportion-based inference can be used. Indeed, even if there is some inconsistent predictions, one may just retain the prediction made by the majority of the triplets.

Thus, the brute-force algorithm consists in looking for all triplets $(\vec{a}, \vec{b}, \vec{c})$ for which the corresponding analogical equation on class is solvable, and which makes an analogical proportion with the item $\vec{d}$ for which one wants to predict the class. This has a clear cubic complexity, and is costly. However, the accuracy results are good enough on real data

[10] In Table 4, each of the 5 columns from Equ_{sss} to Dif stands for a subset of attributes and the corresponding s_k is here a sub-vector of Boolean or nominal attribute values.

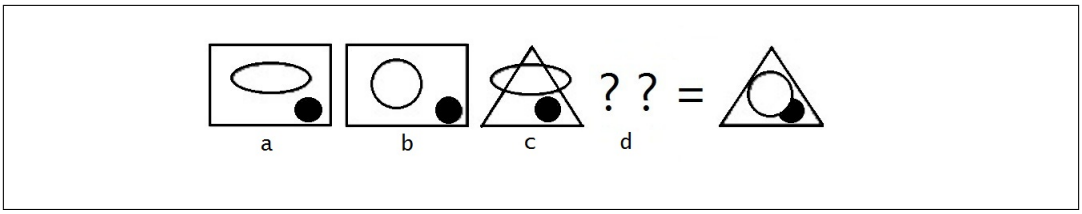

Figure 1: Example of a visual analogical proportion to be completed

benchmarks to be compared with other more classical approaches [38, 9]. It is worth noticing that it is possible to make an offline pre-compilation of pairs, to focus on pairs which are differing on a few attributes only, and to choose $\vec{c}$ as a close neighbor of $\vec{d}$, without harm for the accuracy results [9].

Moreover, a natural idea for restricting the number of triplets is to look only for those triplets involving "competent" pairs. Competent pairs are those in an equivalence class whose pattern for the class is in the majority. For instance, in Table 4, the pairs $(\vec{a}, \vec{b})$ and $(\vec{a'}, \vec{b'})$ are in the the same equivalence class ; their respective pattern for the class are (s, t) and (s, s); the pairs whose pattern for the class is is in the majority in the equivalence class will be considered as "competent" for extrapolating $cl(\vec{d})$ on the basis of such a pair and a $\vec{c}$ forming an analogical proportion $\vec{a} : \vec{b} :: \vec{c} : \vec{d}$ with $\vec{d}$. The use of competent pairs has proved to be experimentally interesting in terms of accuracy and computational cost [35, 21].

5 Towards a logic of creativity

Analogical reasoning has long had a reputation for creativity [47, 17, 29]. First, let us explain in what limited sense we can talk about creativity in the case of analogical proportion-based inference.

For doing that, we use a simple IQ test-like example[11]. The problem is to find, among a given set of candidate solutions, the figure X that gives the best fit to the analogical proportion "A is to B as C is to X". This kind of problem was successfully addressed very early in artificial intelligence by Thomas Evans [23].

In the following, we show that it can be solved *without the knowledge of candidate solutions*. Let us consider the example of Figure 1. The figures in this example of Figure 1 can be encoded with 5 Boolean predicates, namely $hasRectangle(hR)$, $hasBlackDot(hBD)$,

[11] More difficult tests, like Raven IQ tests where a series of instances having the format of a 3×3 matrix whose cells contain diverse geometric figures, except the last cell which is empty and has to be completed by selecting a solution among 8 candidates, can be also solved by analogical proportion-based reasoning without the help of candidate solutions [10].

	hR	hBD	hT	hC	hE
$\boldsymbol{a}$	1	1	0	0	1
$\boldsymbol{b}$	1	1	0	1	0
$\boldsymbol{c}$	0	1	1	0	1
$\boldsymbol{x}$	?	?	?	?	?

Table 5: Solving the example: $\boldsymbol{x} = (0, 1, 1, 1, 0)$

$hasTriangle(hT)$, $hasCircle(hC)$, and $hasEllipse(hE)$. They appear in that order in Table 5, where the example is encoded.

It can be observed that the description of figure x, namely, $hR = 0$, $hBD = 1$, $hT = 1$, $hC = 1$, $hE = 0$ can be obtained by solving the analogical equation in each column, which in this example, has a solution for each feature. Thus, from three items $\vec{a}$, $\vec{b}$, $\vec{c}$, we are able to build (*create*!) a *novel* item $\vec{d}$ different from the three others. Since in the above example we have not encoded the position of the figures in a picture, each figure could be drawn inside, outside or intersecting the other ones. We could also take care of the positions with respect to, e.g., the basis of the rectangles and triangles, by using more attributes. Thus, for instance, the Black Dot will remain at the same place in the different figures. But what is important in the handling of such riddles is to use independent attributes in the representation; but simple relationships, such as "the black dot is outside the ellipse", "the black dot is inside the rectangle" could be also coded directly; see [10] for more discussions.

This form of creativity can be summarized by the following inference pattern (remember $\equiv$ is associative), where $\vec{a}, \vec{b}, \vec{c}$ are vectors of Boolean attribute values defined on the same set of attributes:

$$\frac{\vec{a}, \vec{b}, \vec{c}}{\vec{d} = (\vec{a} \equiv \vec{b} \equiv \vec{c})} \tag{5}$$

This pattern of inference has been considered a fundamental element of human creativity [32]. However, note that we consider this pattern as *valid only if* on each feature i the Boolean analogical equation $a_i : b_i :: c_i : x_i$ is solvable, which requires that

$$(\vec{a} \equiv \vec{b}) \vee (\vec{a} \equiv \vec{c})$$

holds true for each vector component.

Remark 1. *Strictly speaking, we might accept the inference pattern (5) without any restriction, since $\vec{a} \equiv \vec{b} \equiv \vec{c}$ is always defined. Such a view was defended by S. Klein [31] who was a forerunner of the Boolean modeling of analogical proportions used here. But*

this leads to debatable consequences. Indeed Klein's view of analogy was deeply influenced by his anthropological interest in cultural devices such as Navaho sand paintings or mandalas [31]. Such paintings upon the ground have a square structure and can be contemplated from any side (there is no top or bottom); it makes natural a property such as $\vec{a} : \vec{b} :: \vec{c} : \vec{d} \Rightarrow \vec{b} : \vec{c} :: \vec{d} : \vec{a}$*, where* $\vec{a}, \vec{b}, \vec{c}, \vec{d}$ *refer to the descriptions of the four corners of a sand painting. If we iterate the property, namely,* $\vec{b} : \vec{c} :: \vec{d} : \vec{a} \Rightarrow \vec{c} : \vec{d} :: \vec{a} : \vec{b}$*, we see it entails symmetry. But while this property preserves* $0 : 0 :: 0 : 0$ *and* $1 : 1 :: 1 : 1$*, exchanges* $0 : 1 :: 0 : 1$ *and* $1 : 0 :: 1 : 0$*, it changes* $0 : 0 :: 1 : 1$ *into* $0 : 1 :: 1 : 0$ *and* $1 : 1 :: 0 : 0$ *into* $1 : 0 :: 0 : 1$*, thus introducing two patterns excluded by the condition* $(\vec{a} \equiv \vec{b}) \vee (\vec{a} \equiv \vec{c})$*. This latter condition preserves strict analogical proportions (otherwise the undesirable property* $\vec{a} : \vec{b} :: \vec{c} : \vec{d} \Rightarrow \vec{b} : \vec{a} :: \vec{c} : \vec{d}$ *would hold, and break the oriented nature of the comparisons inside the analogical proportion).*

Our aim in the following is to investigate what consequence relation could be defined between ordered pairs. This relation, once symmetrized, must give rise to an equivalence relation between ordered pairs, which must therefore be reflexive, symmetrical and transitive. In a Boolean framework, such a relation corresponds to a logical connector between four variables (two per pair). Analogical proportions are a particular case of logical proportions. This is why we first start by a short journey among logical proportions [48] in the next subsection, looking for those proportions that are reflexive, symmetrical and transitive when considered as operators on pairs. The contents of the rest of this section expands the presentation of a logic outlined in [53].

5.1 Logical proportions

The logical proportions [48] offer a framework, in propositional logic, of quaternary connectors expressing relations between pairs. It is from this Boolean framework, the essence of which we now recall, that we start our investigations.

In the Boolean framework, we have four comparison indicators to relate two variables a to b.

- Two indicators express *similarity*, either *positively* as $a \wedge b$ (which is true if a and b are true), or *negatively* as $\neg a \wedge \neg b$ (which is true if a and b are false).
- The other two are indicators of *dissimilarity* $\neg a \wedge b$ (which is true if a is false and b is true) and $a \wedge \neg b$ (which is true if a is true and b is false).

Definition 1. *A logical proportion* $T(a, b, c, d)$ *is the conjunction of two equivalences between an indicator for* (a, b) *and an indicator for* (c, d)*.*

The expression $((a \wedge \neg b) \equiv (c \wedge \neg d)) \wedge ((a \wedge b) \equiv (c \wedge d))$

provides an example of a logical proportion, where the same dissimilarity operator and the same similarity operator are applied to both pairs. As can be seen, it expresses that "a differs from b as c differs from d" and that "a is similar to b as c is similar to d". It seems to refer to the comparison of the elements within each ordered pair, but we shall see that this is not in the sense of an analogical proportion.

It has been established [48] that there are 120 syntactically and semantically distinct logical proportions. Because of the way they are built, all these proportions share a remarkable property: They are true for exactly 6 patterns of $abcd$ values among $2^4 = 16$ candidate patterns. For instance, the above proportion is true for 0000, 1111, 1010, 0101, 0001, and 0100. The interested reader is invited to consult [48, 49] for in-depth studies of the different types of logical proportions.

Among all 120 logical proportions T, only 6 are reflexive (i.e., $T(a, b, a, b)$ holds true) [48].

Proposition 1. *Only 6 logical proportions are reflexive. They are*
- the analogical proportion

$$A(a, b, c, d) = a : b :: c : d = ((a \wedge \neg b) \equiv (c \wedge \neg d)) \wedge ((\neg a \wedge b) \equiv (\neg c \wedge d))$$

- the so-called paralogy
$P(a, b, c, d) = ((a \wedge b) \equiv (c \wedge d)) \wedge ((\neg a \wedge \neg b) \equiv (\neg c \wedge \neg d))$
- and the 4 following conditional logical proportions
$((a \wedge b) \equiv (c \wedge d)) \wedge ((a \wedge \neg b) \equiv (c \wedge \neg d))$;
$((a \wedge b) \equiv (c \wedge d)) \wedge ((\neg a \wedge b) \equiv (\neg c \wedge d))$;
$((a \wedge \neg b) \equiv (c \wedge \neg d)) \wedge ((\neg a \wedge \neg b) \equiv (\neg c \wedge \neg d))$;
$((\neg a \wedge b) \equiv (\neg c \wedge d)) \wedge ((\neg a \wedge \neg b) \equiv (\neg c \wedge \neg d))$.

The paralogy expresses that "what a and b have in common (positively or negatively), c and d also have, and vice versa". This proportion is (only) true for 0000, 1111, 1010, 0101, 1001, and 0110. It can be checked that $\mathrm{P}(a, b, c, d)) \Leftrightarrow \mathrm{A}(c, b, a, d)$. It has been already mentioned in subsection 3.2 that $\mathrm{A}(a, b, c, d)$ is equivalent to $\mathrm{P}(a, d, b, c)$. The reason of the name "conditional logical proportions" will appear in next subsection.

These 6 logical proportions are also among the 12 logical proportions that are symmetrical ($T(a, b, c, d) \Leftrightarrow T(c, d, a, b)$) and among the 54 logical proportions that are transitive ($T(a, b, c, d), T(c, d, e, f) \Rightarrow T(a, b, e, f)$) [48].

Proposition 2. *A, P and the 4 conditional logical proportions of Proposition 1 are the only logical proportions that define equivalence relations between ordered pairs.*

It should be also clear that while the 6 above logical proportions satisfy the first two postulates P1 and P2 (reflexivity and symmetry), only the analogical proportion is stable under central permutation (i.e., satisfies P3). The paralogy $\mathrm{P}(a, b, c, d))$ is stable under the permutation of the first two items (or the last two, due to symmetry), i.e., we have $\mathrm{P}(a, b, c, d)) \Leftrightarrow \mathrm{P}(b, a, c, d))$. The last 4 logical proportions are not stable for any permutation of two items. This shows that these logical proportions are quite different and serve different purposes.

Indeed it turns out that this result covers two important cases:

- the logic of conditional events which is a basis of non monotonic reasoning, recalled in the next subsection, and which will also be a source of inspiration for the rest of the paper,

- a logic of ordered pairs preserving positive changes, which may contribute to a "controlled" creativity process, outlined in the rest of the section.

We first recall the logic associated with the conditional logical proportions, since the way a logic is associated with these proportions will guide us for building a logic associated with the analogical proportion, and another one associated with the paralogy proportion (which will be only briefly mentioned, since out of the analogical scope of the paper).

5.2 Conditional events as a basis of non monotonic reasoning

Let us consider the 4 conditional proportions which are related to our subject, since they are reflexive, symmetrical and transitive. Let us first explain the word "conditional". It comes from the fact that these proportions express equivalences between conditional statements. Indeed, it was pointed out in [18] that a rule "if a then b" can be considered as a three-valued entity referred as a "conditional object" or a "conditional event", and denoted $b|a$. This entity is *tri-valued* [15] as follows:

- $b|a$ is true if $a \wedge b$ is true. The elements which make true $a \wedge b$ are the *examples* of the rule "if a then b";
- $b|a$ is false if $a \wedge \neg b$ is true. The elements which make true $a \wedge \neg b$ are the *counter-examples* of the rule "if a then b";
- $b|a$ is undefined if a is false. The rule "if a then b" is then not applicable.

Consider the conditional proportion appearing in Proposition 1 and which was our first example of a logical proportion:

$$((a \wedge b) \equiv (c \wedge d)) \wedge ((a \wedge \neg b) \equiv (c \wedge \neg d)) \tag{6}$$

The above logical proportion can then be denoted $b|a :: d|c$ by combining the notation of conditional objects with that of the analogical proportion. Indeed, the proportion $b|a :: d|c$

expresses a semantic equivalence between the two rules "if a then b" and "if c then d" by stating that:

- they have the same examples, i.e., $(a \wedge b) \equiv (c \wedge d)$;
- they have the same counter-examples, i.e., $(a \wedge \neg b) \equiv (c \wedge \neg d)$;
- if $b|a$ is not applicable, i.e., a is false, then necessarily c is false (otherwise (6) would be false), which means that $d|c$ is not applicable.

The logical consequence relation between conditional objects $b|a \models d|c$ is defined as [15]:

$$a \wedge b \vDash c \wedge d \text{ and } c \wedge \neg d \vDash a \wedge \neg b \tag{7}$$

which expresses that the examples of the first conditional object are examples of the second one, and the counter-examples of the second conditional object are counter-examples of the first one. This entailment is naturally associated with the conditional proportion $b|a :: d|c$, since

$$b|a :: d|c \Leftrightarrow b|a \vDash d|c \text{ and } d|c \vDash b|a.$$

The transitivity of the 4 conditional proportions of the Proposition 1 reflects the fact that they express equivalences between conditional objects (and thus between rules), namely respectively $b|a :: d|c$, $a|b :: c|d$, $a|\neg b :: c|\neg d$, and $b|\neg a :: d|\neg c$.

The conditional object $b|a$ must therefore be thought of as a rule "if a then b". A rule may have exceptions. That is, we can have at the same time the rule "if a then b" and a rule "if $(a \wedge c)$ then $\neg b$". The two conditional objects $b|a$ and $\neg b|(a \wedge c)$ do not lead to a contradiction in the presence of the facts a and c (unlike a modeling of rules by material implication), in the setting of a tri-valued logic where the conjunction $\&$ is defined by [18]:

$$b|a \;\&\; d|c \triangleq ((a \rightarrow b) \wedge (c \rightarrow d))|(a \vee c)$$

where $\rightarrow$ is the material implication ($a \rightarrow b \triangleq \neg a \vee b$) and with the following semantics: $val(o_1 \& o_2) = \min(val(o_1), val(o_2))$ where the three truth values are ordered as follows: undefined $>$ true $>$ false.[12]

It can be shown that this quasi-conjunction '&' (that is its name) is associative. It expresses that the set constituted by the two rules "if a then b" and "if c then d" is triggerable if a or c is true, and in this case the triggered rule behaves like the material implication. This logic constitutes the simplest semantics [6] of the system P of non-monotonic inference of Kraus, Lehmann, and Magidor [33]. The reader can consult [18, 6] for more details.

[12]The negation is defined by $\neg(b|a) = (\neg b|a)$; $\neg(b|a)$ is undefined if and only if $b|a$ is.

Although nonmonotonic reasoning only yields plausible conclusions, it is not clear that it might be considered as a special form of induction. Indeed nonmonotonic reasoning is here a two steps-process. First from a set of conditional events representing a set of default rules, we *deduce* a new conditional event whose condition part corresponds exactly to our knowledge of the current situation under consideration, and then - second step - we apply the new default rule thus inferred to the current situation.

We now consider the two other logical proportions that define an equivalence relation between ordered pairs, namely the analogical proportion and the paralogy, and we try to identify what consequence relations can be associated with them.

5.3 A new consequence relation between ordered pairs

In the following subsections, we try to identify some elements of a comparative logic of ordered pairs. The items to be compared are described by vectors of attribute values (here Boolean). When $a_i = 1$ (resp. $a_i = 0$) we understand it as item $\vec{a}$ has (resp. has not) feature / property i.

As usual, logical connectives extend to vectors componentwise:

1. $\neg\vec{a} = (\neg a_1, ..., \neg a_n)$;
2. $\vec{a} \wedge \vec{b} = (a_1 \wedge b_1, ..., a_n \wedge b_n)$;
3. $\vec{a} \vee \vec{b} = (a_1 \vee b_1, ..., a_n \vee b_n)$.

Taking inspiration from the case of conditional logical proportions (namely definition Definition 7), we are led to define the following, new logical consequence relation between pairs (still denoted $\models$) from the definition of an analogical proportion:

$$(\vec{a}, \vec{b}) \models (\vec{c}, \vec{d}) \Leftrightarrow \neg\vec{a} \wedge \vec{b} \models \neg\vec{c} \wedge \vec{d} \text{ and } \vec{c} \wedge \neg\vec{d} \models \vec{a} \wedge \neg\vec{b} \tag{8}$$

When we deal with pairs, the valuation $(a_i, b_i) = (0, 1)$ can be understood as when we go from $\vec{a}$ to $\vec{b}$, we acquire feature i. Thus the meaning of entailment (8) is the following: features that are acquired when going from $\vec{a}$ to $\vec{b}$ remain acquired when going from $\vec{c}$ to $\vec{d}$. Moreover if when going from $\vec{c}$ to $\vec{d}$ a feature is lost, it was already the case when going from $\vec{a}$ to $\vec{b}$.[13]

Proposition 3. *We have the following equivalence:*

$$(\vec{a}, \vec{b}) \models (\vec{c}, \vec{d}) \textit{ and } (\vec{c}, \vec{d}) \models (\vec{a}, \vec{b}) \textit{ iff } A(\vec{a}, \vec{b}, \vec{c}, \vec{d})$$

[13]The choice of definition (8), rather than $(\vec{a}, \vec{b}) \models (\vec{c}, \vec{d}) \Leftrightarrow \vec{a} \wedge \neg\vec{b} \models \vec{c} \wedge \neg\vec{d}$ and $\neg\vec{c} \wedge \vec{d} \models \neg\vec{a} \wedge \vec{b}$, is governed by the need here to privilege the acquisition of features rather than their loss. Indeed the alternative definition given in this footnote says that features that are lost when going from $\vec{a}$ to $\vec{b}$ remain lost when going from $\vec{c}$ to $\vec{d}$, and that if when going from $\vec{c}$ to $\vec{d}$ a feature is acquired, it was already the case when going from $\vec{a}$ to $\vec{b}$.

Proof: Let us see the precise meaning of this definition for pairs. Because we are working componentwise, it is enough to consider the consequence of this definition on one component. Two cases have to be considered:

- Case $a = b$ (representing 8 valuations among the 16 candidates for a, b, c, d). Because $\neg a \wedge b$ and $a \wedge \neg b$ are 0, the only constraint is that $c \wedge \neg d = 0$ which is valid only if $(c, d) \neq (1, 0)$, eliminating (0010) and (1110) as valid valuations, leaving 6 valuations still valid.

- Case $a \neq b$ (representing the 8 remaining valuations): if $(a, b) = (1, 0)$, there is no constraint on (c, d). If $(a, b) = (0, 1)$, only $(c, d) = (0, 1)$ is valid eliminating 3 valuations among the 8: $(0100), (0110), (0111)$

Having both $(\vec{a}, \vec{b}) \vDash (\vec{c}, \vec{d})$ and $(\vec{c}, \vec{d}) \vDash (\vec{a}, \vec{b})$ leads to the truth table of $A(a, b, c, d)$ with exactly 6 valid valuations. □

Because when $(\vec{a}, \vec{b}) \vDash (\vec{c}, \vec{d})$, the 5 valuations (0010), $(1110), (0100)$, (0110), (0111) are *forbidden* for each $(a_i\ b_i\ c_i\ d_i)$, this means that

- $(a_i, b_i) = (0, 1) \Rightarrow (c_i, d_i) = (0, 1)$; (a property acquired from $\vec{a}$ to $\vec{b}$ has to be acquired from $\vec{c}$ to $\vec{d}$)

- $a_i = b_i \Rightarrow (c_i, d_i) \neq (1, 0)$ (when there is no acquisition or loss from $\vec{a}$ to $\vec{b}$, there cannot be a loss from $\vec{c}$ to $\vec{d}$)

Similarly, we have $(\vec{c}, \vec{d}) \vDash (\vec{a}, \vec{b}) \Leftrightarrow \begin{cases} (a_i, b_i) = (1, 0) \Rightarrow (c_i, d_i) = (1, 0) \\ a_i = b_i \Rightarrow (c_i, d_i) \neq (0, 1) \end{cases}$

which forbids the 5 valuations (1000), (1001), (1011), (0001), (1101).

Thus we have, as expected, $(\vec{a}, \vec{b}) \vDash (\vec{c}, \vec{d})$ and $(\vec{c}, \vec{d}) \vDash (\vec{a}, \vec{b}) \Leftrightarrow A(\vec{a}, \vec{b}, \vec{c}, \vec{d})$.

Table 6 exhibits the situations where the entailments (defined by (8) $(a, b) \vDash (c, d)$ and $(c, d) \vDash (a, b)$ are true. The relation $\vDash$ is a clear weakening of the analogical proportion when viewed as a relation between pairs. To support the intuition of the entailment, let us consider the case where a, b, c, d are just Boolean values. As previously explained:

- $(0, 1)$ has 1 consequence $(0, 1)$,

- $(0, 0)$ has 3 consequences $(0, 0), (0, 1), (1, 1)$,

- $(1, 1)$ has 3 consequences $(0, 0), (0, 1), (1, 1)$,

a	b	c	d	$(a,b) \vDash (c,d)$	$(c,d) \vDash (a,b)$	$a : b :: c : d$
0	0	0	0	1	1	1
0	0	0	1	1	0	0
0	0	1	0	0	1	0
0	0	1	1	1	1	1
0	1	0	0	0	1	0
0	1	0	1	1	1	1
0	1	1	0	0	1	0
0	1	1	1	0	1	0
1	0	0	0	1	0	0
1	0	0	1	1	0	0
1	0	1	0	1	1	1
1	0	1	1	1	0	0
1	1	0	0	1	1	1
1	1	0	1	1	0	0
1	1	1	0	0	1	0
1	1	1	1	1	1	1
nb	of	values	'true'	**11**	**11**	**6**

Table 6: Entailments defined by (8) vs. analogical proportion

- $(1,0)$ has 4 consequences $(0,0), (0,1), (1,0), (1,1)$.

A compact way to put it would be to order the set of Boolean pairs such as $(1,0) < \{(0,0),(1,1)\} < (0,1)$ and to consider that any pair entails all pairs at the same level and higher.

Thus, if we consider a pair $(\vec{a}, \vec{b})$ of vectors of dimension 4 where $\vec{a} = (0,0,1,1)$ and $\vec{b} = (0,1,0,1)$, we see that this pair has $3 \times 1 \times 4 \times 3 - 1 = 35$ distinct logical consequences in the sense of $\vDash$ defined by (8).

5.4 Logical combinations of ordered pairs

One may think of defining conjunctive or disjunctive combinations of ordered pairs, but these combinations should agree with the consequence relation (8) and make sense with respect to the interpretation of pairs. Natural componentwise definitions, including negation, seem to be:

$$(\vec{a}, \vec{b}) \wedge (\vec{c}, \vec{d}) = (\vec{a} \wedge \vec{c}, \vec{b} \wedge \vec{d});$$

$$(\vec{a}, \vec{b}) \vee (\vec{c}, \vec{d}) = (\vec{a} \vee \vec{c}, \vec{b} \vee \vec{d}).$$

$$\neg(\vec{a}, \vec{b}) = (\neg\vec{a}, \neg\vec{b})$$

Note that $\neg(\vec{a}, \vec{b}) \neq (\vec{b}, \vec{a})$ in general. However, an involutive operation such as

$$\circlearrowleft(\vec{a}, \vec{b}) = (\vec{b}, \vec{a})$$

looks more interesting, as we shall see, since it reverses the order of comparison in the pair. Besides, as a consequence of the above definitions, we have

$$(\vec{a}, \vec{b}) \wedge (\vec{a}, \vec{b}) = (\vec{a}, \vec{b}) = (\vec{a}, \vec{b}) \vee (\vec{a}, \vec{b})$$

But unfortunately one can check that

$$(\vec{a}, \vec{b}) \wedge (\vec{c}, \vec{d}) \not\models (\vec{a}, \vec{b}) \not\models (\vec{a}, \vec{b}) \vee (\vec{c}, \vec{d}).$$

This failure is simply due to the fact that a feature acquired from $\vec{a} \wedge \vec{c}$ to $\vec{b} \wedge \vec{d}$ may not be a feature acquired from $\vec{a}$ to $\vec{b}$. Indeed starting with $(a_i, b_i, c_i, d_i) = (1, 1, 0, 1)$, we get $(a_i \wedge c_i, b_i \wedge d_i) = (0, 1)$ and $(0, 1) \not\models (1, 1)$.[14]

However, this should not come as a surprise. Indeed, here $\models$ preserves pairs of the form $(0, 1)$, while the conjunction of pairs preserves $(0, 1)$ if it appears in both places of the conjunction, but also when one of the pairs is equal to (1, 1) for some feature. This leads us to introduce a new operation $\wedge\vee$ mixing conjunction and disjunction:

$$(\vec{a}, \vec{b}) \wedge\vee (\vec{c}, \vec{d}) = (\vec{a} \wedge \vec{c}, \vec{b} \vee \vec{d})$$

Obviously, this operator $\wedge\vee$ is commutative and associative by construction. As much as the logical consequence relation between pairs defined by (8) makes sense, the intuition might seem more fragile for the conjunction / disjunction of pairs. However note that $(a_i \wedge c_i, b_i \vee d_i) = (1, 0)$ only if $(a_i, b_i) = (c_i, d_i) = (1, 0)$. By contrast, if (a_i, b_i) or $(c_i, d_i) = (0, 1)$, $(a_i \wedge c_i, b_i \vee d_i) = (0, 1)$.

In a dual manner, one can define

$$(\vec{a}, \vec{b}) \vee\wedge (\vec{c}, \vec{d}) = (\vec{a} \vee \vec{c}, \vec{b} \wedge \vec{d}).$$

Indeed there is a De Morgan duality with respect to the operation $\circlearrowleft$ between $\vee\wedge$ and $\wedge\vee$, namely

$$\circlearrowleft(\circlearrowleft(\vec{a}, \vec{b}) \vee\wedge \circlearrowleft(\vec{c}, \vec{d})) = (\vec{a}, \vec{b}) \wedge\vee(\vec{c}, \vec{d}).$$

[14] There are two other cases of violation when $(a_i, b_i) = (1, 0)$, $(c_i, d_i) = (0, 0)$ or $(c_i, d_i) = (0, 1)$, we get $(a_i \wedge c_i, b_i \wedge d_i) = (0, 0)$, and $(0, 0) \not\models (1, 0)$. Besides, $(\vec{a}, \vec{b}) \not\models (\vec{a}, \vec{b}) \vee (\vec{c}, \vec{d})$ due to three possible situations: i) $(a_i, b_i) = (0, 0)$, $(c_i, d_i) = (1, 0)$ and $(0, 0) \not\models (1, 0)$; ii) & iii) $(a_i, b_i) = (0, 1)$, $(c_i, d_i) = (1, 1)$ or $(c_i, d_i) = (1, 0)$, and $(0, 1) \not\models (1, 1)$.

Note that $(a_i \vee c_i, b_i \wedge d_i) = (0,1)$ only if $(a_i, b_i) = (c_i, d_i) = (0,1)$. But, if (a_i, b_i) or $(c_i, d_i) = (1,0)$, $(a_i \wedge c_i, b_i \vee d_i) = (1,0)$. Then it can be checked that $\vee\wedge$ behaves like a conjunction, and $\wedge\vee$ like a disjunction, in the sense that:

Proposition 4.

$$(\vec{a}, \vec{b}) \vee\wedge (\vec{c}, \vec{d}) \vDash (\vec{a}, \vec{b}) \vDash (\vec{a}, \vec{b}) \wedge\vee (\vec{c}, \vec{d})$$

where $\vDash$ is defined by (8).

Proof. We should first show that $(a \vee c, b \wedge d) \vDash (a, b)$. Indeed it holds since we have 1. $\neg(a \vee c) \wedge b \wedge d \vDash \neg a \wedge b$; 2. $a \wedge \neg b \vDash (a \vee c) \wedge \neg(b \wedge d)$.

It remains to show that $(a, b) \vDash (a \wedge c, b \vee d)$. Indeed it can be checked that we have 1. $\neg a \wedge b \vDash \neg(a \wedge c) \wedge (b \vee d)$; 2. $a \wedge c \wedge \neg(b \vee d) \vDash a \wedge \neg b$. □

Remark. Lines for further research

The conditional events involved in the conditional logical proportions have a tri-valued semantics. From an analogical proportion point of view, a natural way to associate a tri-valuation to an ordered pair of Boolean vectors, is to compute their difference to get a vector belonging to $\{-1, 0, 1\}^n$: $val_A(\vec{a}, \vec{b}) = \vec{a} - \vec{b} = (a_1 - b_1, ..., a_n - b_n) \in \{-1, 0, 1\}^n$.

Then one can check that $A(\vec{a}, \vec{b}, \vec{c}, \vec{d})$ is true if and only if $val_A(\vec{a}, \vec{b}) = val_A(\vec{c}, \vec{d})$. Moreover, if $A(\vec{a}, \vec{b}, \vec{c}, \vec{d})$ is true, we have

$$(\vec{a} \wedge \vec{c}) - (\vec{b} \wedge \vec{d}) = \vec{a} - \vec{b} = \vec{c} - \vec{d} = (\vec{a} \vee \vec{c}) - (\vec{b} \vee \vec{d}).$$

This means that $A(\vec{a}, \vec{b}, \vec{c}, \vec{d})$ entails $A(a \vec{\wedge} \vec{c}, b \vec{\wedge} \vec{d}, \vec{a} \vee \vec{c}, \vec{b} \vee \vec{d})$, but the converse is wrong.[15]

While the analogical proportion insists on the identity of the differences existing in each pair, the paralogy expresses rather a parallel between the pairs at the level of shared properties, positively or negatively. This is reflected in the following result, dual to that for analogy:

$$P(\vec{a}, \vec{b}, \vec{c}, \vec{d}) \text{ iff } \begin{cases} Dif(\vec{a}, \vec{b}) = Dif(\vec{c}, \vec{d}) \\ Equ^1(\vec{a}, \vec{b}) = Equ^1(\vec{c}, \vec{d}) \\ Equ^0(\vec{a}, \vec{b}) = Equ^0(\vec{c}, \vec{d}) \end{cases}$$

We could also define an entailment starting from paralogy, such that
$(\vec{a}, \vec{b}) \vDash_P (\vec{c}, \vec{d}) \Leftrightarrow \vec{a} \wedge \vec{b} \vDash \vec{c} \wedge \vec{d}$ and $\neg\vec{c} \wedge \neg\vec{d} \vDash \neg\vec{a} \wedge \neg\vec{b}$,
or alternatively $(\vec{a}, \vec{b}) \vDash'_P (\vec{c}, \vec{d}) \Leftrightarrow \neg\vec{a} \wedge \neg\vec{b} \vDash \neg\vec{c} \wedge \neg\vec{d}$ and $\vec{c} \wedge \vec{d} \vDash \vec{a} \wedge \vec{b}$,
depending if we privilege the persistence of properties shared positively inside the pairs, or shared negatively, when going from the pair $(\vec{a}, \vec{b})$ to the $(\vec{c}, \vec{d})$.

[15] Indeed $A(a \vec{\wedge} \vec{c}, b \vec{\wedge} \vec{d}, \vec{a} \vee \vec{c}, \vec{b} \vee \vec{d})$ is also true for $(a\ b\ c\ d) = (0\ 1\ 1\ 0)$ or $(1\ 0\ 0\ 1)$.

Moreover, the tri-valuation naturally associated with a pair, from the point of view of paralogy, would be $val_P(\vec{a}, \vec{b}) = (a_1 + b_1, ..., a_n + b_n) \in \{0, 1, 2\}^n$. Indeed it can be checked that $P(\vec{a}, \vec{b}, \vec{c}, \vec{d})$ holds true if and only if $val_P(\vec{a}, \vec{b}) = val_P(\vec{c}, \vec{d})$.

We leave these entailments associated with the paralogy, and the tri-valued logics associated with the analogical proportion and the paralogical proportion for a further study.

5.5 Creative inference

Given a set $\mathcal{S}$ of existing items, each represented by a set of Boolean attribute values, creativity may amount to produce an item not in $\mathcal{S}$, but described by the same set of attributes. Viewed like that, creativity is an easy game: we have just to choose at random the attribute values and to check if the result is not already in $\mathcal{S}$. However with such a process we have no control on the the attribute values that might be desirable. In the following, we present a *creative inference process* that attempts to improve a particular item or entity, taking advantage of a set of ordered pairs of existing items, using an analogical proportion-based mechanism. However, we certainly do not claim that every form of creative analogical inference, taken in the broadest sense, could be captured by the mechanism we propose.

More precisely, let us suppose we have a sample set $\mathcal{S}$ of items from which a set $\mathcal{P}$ of k ordered pairs $(\vec{a}^j, \vec{b}^j)$ with $j \in \{1, \ldots, k\}$ has been extracted where the $\vec{a}^j$'s and $\vec{b}^j$'s are in $\mathcal{S}$. Each vector in $\mathcal{S}$ is a Boolean representation of an object/profile/situation belonging to a real world universe, and then, each pair of vectors $(\vec{a}^j, \vec{b}^j)$, all of the same dimension n, represents legitimate, feasible / allowed changes from $\vec{a}^j$ to $\vec{b}^j$.

Then given a current fixed item represented by a vector $\vec{c} \in \mathcal{S}$ one may wonder what new item(s) $\vec{d}$ could be obtained by applying some change existing in the base of pairs. This item could be the representation of a plausible item in the real world.

A first option is to consider the set of solutions (when the solution exists)

$$\vec{d} \in \{\vec{x} \mid \exists (\vec{a}^j, \vec{b}^j) \in \mathcal{P}, j \in \{1, \ldots, k\} \text{ such that } A(\vec{a}^j, \vec{b}^j, \vec{c}, \vec{x}) \text{ holds}\} \tag{9}$$

This is the approach followed in [1]. When there is no solution or when the values found for $\vec{d}$ are not considered satisfactory enough, we have to consider other options. One idea would be to consider the entailment (8) between pairs associated to the analogical proportion, and then to look for the $\vec{d}$'s such that:

$$\vec{d} \in \{\vec{x} \mid \exists (\vec{a}^j, \vec{b}^j) \in \mathcal{P}, j \in \{1, \ldots, k\} \text{ such that } (\vec{a}^j, \vec{b}^j) \vDash (\vec{c}, \vec{x})\}$$

But this option has two drawbacks. First, the $\vec{d}$'s obtained depend on a unique pair $(\vec{a}^j, \vec{b}^j)$. Second, $\vDash$ is quite permissive and the number of pairs $(\vec{c}, \vec{x})$ obtained is likely to be rather large as seen in subsection 5.3 and there is a risk of losing control.

What seems to be a better idea is to enlarge the initial base of pairs by computing new pairs belonging to the closure of operation $\wedge\vee$ introduced in the previous subsection 5.4. This operation has the merit of "cumulating" the acquisition of features[16]. Extending the initial set $\mathcal{P}$ of pairs gives us more chance to find a plausible $\vec{d}$, perhaps with more desirable features. More precisely we apply (9) where $\mathcal{P}$ is replaced by $\mathcal{P}' = \{(a^k, b^k) |\ (a^k, b^k) = (a^i, b^i) \wedge\vee (a^j, b^j)$ such that $((a^i, b^i), (a^j, b^j)) \in \mathcal{P}^2\}$. We may apply this enlargement of $\mathcal{P}$ recursively to $\mathcal{P} \cup \mathcal{P}'$ and so on. This process ensures that i) the $\vec{d}$ obtained are new, and ii) they are obtained from an existing $\vec{c}$ on the basis of already existing changes, since observed on pairs of existing items. Is $\vec{d}$ thus obtained, valuable ? This a completely different issue.

Note that this way of reasoning parallels non monotonic reasoning with conditional objects, where from a base of default rules "if a^j then b^j" represented by a set of conditional objects $b^j|a^j$, one deduces a new conditional object $d|c$, using entailment (7) and conjunction $\&$, where c corresponds to everything we know in the current context, for which we then conclude d [18].

6 Example and first experiments

Before giving some statistics about the behaviour of our mechanism, we start with a simple example.

6.1 An example freely inspired from a simplified Kaggle dataset

To avoid the creation of a completely artificial dataset, we start by using a Kaggle dataset. Kaggle is a platform renowned for hosting data science competitions, collaborative projects, and educational resources, accessible at https://www.kaggle.com/.

The targeted dataset [16] encompasses the details of 1000 users, characterized by 32 attributes. The final column denotes whether they purchased a bike, forming the basis for a binary classification task. Following the exclusion of rows with missing data, 952 complete rows remain.

In order to facilitate the understanding of our process, we narrow our focus to the first 6 attributes of this dataset, creating a simplified universe where objects are represented as Boolean vectors of dimension 6. In the initial dataset, these 6 first features are Marital Status, Gender, Income, Children, Education, Occupation, but, more generally, each attribute has to be viewed as an individual feature describing an object.

By limiting our analysis to the first 6 attributes, we inevitably encounter duplicates. Depending on the random shuffle of the initial dataset comprising 952 items, we end up

[16] However note that $(0,0) \wedge\vee (1,1) = (1,1) \wedge\vee (0,0) = (0,1)$, which may create some unfeasible change; in such a case the generated pair(s) should not be considered in the further process.

with less than $2^6 = 64$ distinct elements from which we must select pairs.

Consequently, we consider only two pairs for the need of our example, denoted $(\vec{a_1}, \vec{b_1})$, $(\vec{a_2}, \vec{b_2})$ that respect the following constraints:

- The Hamming distance $hamming(\vec{a_i}, \vec{b_i})$ is equal to 2 because a pair should represent a realistic perturbation of $\vec{a}$ into $\vec{b}$.
- We do not loose any option when moving from $\vec{a_i}$ to $\vec{b_i}$, i.e., we forbid to have an attribute j such that $a_j = 1$ and $b_j = 0$. All other combinations are allowed.

At this stage of this experiment, the two pairs are selected in a random manner w.r.t. the previous constraints, from the candidate pairs, since we currently lack specific information about the entire universe. However, in practical scenarios, prior knowledge about the universe could result in more suitable selections. Here is an example of the two pairs constituting $\mathcal{P}$:
- $(\vec{a_1}, \vec{b_1}) = ([0,0,0,0,1,0], [0,1,0,0,1,1])$
- $(\vec{a_2}, \vec{b_2}) = ([0,0,0,0,1,0]), [1,0,0,0,1,1])$
When we extend $\mathcal{P}$ with operator ($\wedge\vee$ but without doing the full closure), we add to P the following pair:
- $(\vec{a_3}, \vec{b_3}) = ([0,0,0,0,1,0], [1,1,0,0,1,1])$
just because $(\vec{a_3}, \vec{b_3}) = (\vec{a_1}, \vec{b_1}) \wedge\vee (\vec{a_2}, \vec{b_2})$
Because, $\vec{a_1} = \vec{a_2}$, obviously $\vec{a_1} = \vec{a_3}$ but this is not necessary. Starting from $\vec{c} = [0,0,0,1,0,0]$, we observe in Table 7 that the 3 corresponding analogical equations are solvable. The solution of the third equation is then a new object, which is distinct from the 5 existing vectors $\vec{a_1}, \vec{b_1}, \vec{b_2}, \vec{x_1}, \vec{x_2}$.

This approach only makes practical sense when considering Boolean representations of relatively large dimensions. That is why we give in the following subsection some figures about what can be expected in higher dimensions.

6.2 Higher dimensions

Indeed, in the context of Boolean vectors with high dimension (let us say larger than 10), the available data $\mathcal{S}$ are generally scarce compared to the whole universe: this is a well-known consequence of the curse of dimensionality. For instance, for vectors of dimension 30, the space of possible profiles is of size $2^{30} \sim 10^9$.

Considering $\mathcal{P}$ as the set of pairs built from two distinct elements from $\mathcal{S}$, we can first inquire about the existence, within $\mathcal{P}$, of pairs $(\vec{c}, \vec{d})$ that are logical consequences of another pair $(\vec{a}, \vec{b})$ (in the sense of (8)) in $\mathcal{P}$.

To answer this question, we conducted experiments in dimension 10 and 30 (with reasonable execution times) while varying the size of the sample $\mathcal{S}$. From a practical perspective, when given a pair $(\vec{a}, \vec{b})$ in $\mathcal{P}$, we determine the quantity of pairs $(\vec{c}, \vec{d})$ in $\mathcal{P}$ that are

	Opt1	Opt2	Opt3	Opt4	Opt5	Opt6
$\vec{a_1}$	0	0	0	0	1	0
$\vec{b_1}$	0	1	0	0	1	1
$\vec{c}$	0	0	0	1	0	0
x_1	0	1	0	1	0	1
	Opt1	Opt2	Opt3	Opt4	Opt5	Opt6
$\vec{a_2}$	0	0	0	0	1	0
$\vec{b_2}$	1	0	0	0	1	1
$\vec{c}$	0	0	0	1	0	0
x_2	1	0	0	1	0	1
	Opt1	Opt2	Opt3	Opt4	Opt5	Opt6
$\vec{a_3}$	0	0	0	0	1	0
$\vec{b_3}$	1	1	0	0	1	1
$\vec{c}$	0	0	0	1	0	0
x_3	1	1	0	1	0	1

Table 7: Example in dimension 6

logical consequences of $(\vec{a}, \vec{b})$. Additionally, we distinguish in this count the pairs that form an analogical proportion $(\vec{a}, \vec{b}) :: (\vec{c} : \vec{d})$. We calculate the average of these two numbers across the total number of pairs $(\vec{a}, \vec{b})$ within $\mathcal{P}$. Finally, we present the average values from these calculations in Table 8 based on 10 tests (or 10 samples $\mathcal{S}$). Obviously, as soon as

Dim	Size $\mathcal{S}$	# pairs	# tests	# analogies	std. dev	# log. cons.	std. dev
10	50	1225	10	0	0	20	4
10	100	4950	10	0	0	113	18
30	100	4950	10	0	0	0	0
30	200	19900	10	0	0	0	0

Table 8: Number of pairs that are logical consequences inside S

the random sample set S has a small size w.r.t. the whole universe size, it is very unlikely to get inside $\mathcal{S}$, four vectors $\vec{a}, \vec{b}, \vec{c}, \vec{d}$ such that $\vec{a} : \vec{b} :: \vec{c} : \vec{d}$. Additionally, we notice that as the dimension increases, we also fail to discover any logical consequence within $\mathcal{P}$ that can be seen as a relaxation of analogical proportion. This lack of logical consequences can be attributed to the fact that a global solution is only acceptable if there is a solution for each individual component. When the number of components increases, the number of constraints also increases, thereby decreasing the probability of obtaining a global solution.

And the size of the sample $\mathcal{S}$, from which $\mathcal{P}$ is derived, cannot compensate these increasing constraints.

If we have a sample $\mathcal{S}$ of size 1000, it is natural to be interested in a "reasonable" extension of the sample. This is where the analogy with analogical extension comes into play, which involves completing the set of initial examples, as seen, for instance, in [11]. But if the analogical extension does not provide enough new elements, we could then implement, initially, the logical consequence of pairs, seen as a way to weaken the analogical constraint as follows:

- Every pair $(\vec{a}, \vec{b})$ from the sample represents a potential variation of the profiles.
- Any pair $(\vec{c}, \vec{d})$ such that $(\vec{a}, \vec{b}) \models (\vec{c}, \vec{d})$ can potentially be regarded as the description of a candidate variation of the profiles.

In the absence of an efficient algorithm, the task of generating logical consequences can prove to be very complex. So, instead of trying to generate via brute force the set of logical consequences, another option is to try to solve the equation $(\vec{a}, \vec{b}) \models (\vec{c}, \vec{d})$ where $\vec{a}, \vec{b}, \vec{c}$ are in $\mathcal{S}$: in that context, instead of looking for a pair $(\vec{c}, \vec{d})$ we just look for at least one element $\vec{d}$, if it exists, that is not in $\mathcal{S}$ and satisfying $(\vec{a}, \vec{b}) \models (\vec{c}, \vec{d})$. Once again:

- Every pair $(\vec{a}, \vec{b})$ from the sample represents a potential variation of the profiles.
- Given another profile $\vec{c}$ from $\mathcal{S}$, a profile $\vec{d} \notin \mathcal{S}$ such that $(\vec{a}, \vec{b}) \models (\vec{c}, \vec{d})$ can be considered plausible and added to the initial sample.

First of all, given a pair $(\vec{a}, \vec{b}) \in \mathcal{P}$, we compute the average number over $\vec{c} \in \mathcal{S}$ of $\vec{d} \notin \mathcal{S}$ satisfying $(\vec{a}, \vec{b}) \models (\vec{c}, \vec{d})$. This number tells us the likelihood of creating a new profile $\vec{d}$ when solving the equation $(\vec{a}, \vec{b}) \models (\vec{c}, \vec{d})$ starting from 3 elements $\vec{a}, \vec{b}, \vec{c} \in S$. Then for a given sample S, we can compute the average number of profile $\vec{d}$ that can be generated from pairs in $\mathcal{P} = \mathcal{S} \times \mathcal{S}$ with the help of a third element $\vec{c} \in \mathcal{S}$. Finally, we average this computation on 10 tests and show the result in Table 9 showing the number of $\vec{d} \notin \mathcal{S}$ with the average standard deviation as last column. We conducted experiments in dimensions 10, 30, and 50 with various sample sizes. Table 9 shows that, in general, the equation $(\vec{a}, \vec{b}) \models (\vec{c}, \vec{d})$ where $\vec{d}$ is the unknown, and $\vec{a}, \vec{b}, \vec{c} \in S$ does not have a solution. As it is the case for Table 8, this might be understood because a global solution $\vec{d}$ is acceptable only if there is a solution componentwise. Augmenting the number of components (from 10 to 30 to 50) increases the number of constraints and reduce the likelihood to having a global solution.

The previous experiments suggest that the use of $\wedge\vee$ operator as a pair creator might be more productive. We will count in Table 10 how many completely new pairs are created when applying the $\wedge\vee$ operator to all pairs derived from the sample $\mathcal{S}$. We display the

Dim.	Size $\mathcal{S}$	# pairs	# tests	# $\vec{d} \notin \mathcal{S}$	std. dev.
10	50	1225	10	0.3	0.2
10	100	4950	10	0.23	0.21
30	100	4950	10	0.03	0.03
30	200	19900	10	0.03	0.04
50	100	4950	10	0.01	0.02
50	200	19900	10	0.007	0.006

Table 9: Number of vectors $\vec{d}$ solutions of the equation

average value on 10 tests with the standard deviation. At this stage, we do not eliminate pairs where at least one component appears as $(0,0) \wedge\vee (1,1)$ or $(1,1) \wedge\vee (0,0)$. See footnote number 8.

Dim.	Size S	# pairs	# tests	# new pairs	std dev.
10	50	1225	10	333	34
10	100	4950	10	552	28
30	100	4950	10	9423	65
30	200	19900	10	Not Avail.	Not Avail.

Table 10: Number of deduced pairs built with vectors not in S

Since the resulting pairs are only retained if both constituting vectors are not in S, we have constructed at least $\#newpairs$ new vectors (a new vector may appear in multiple new pairs).

It is widely admitted that analogical reasoning only leads to plausible consequences. Its application to creativity does not escape this rule. It will certainly be useful in practice to verify, in one way or another, the feasibility of the new pairs obtained.

7 Conclusion

This paper has discussed analogical reasoning based on analogical proportions. We have first singled out this inference as a special form of induction, more precisely of transduction, where comparisons between examples take place. After providing a refresher on analogical proportions defined between entities represented by means of Boolean or nominal features, we have emphasized that, in this case, analogical proportions define *equivalence* relations between ordered pairs of entities.

We have then surveyed how analogical proportions-based inference can be used for classification tasks, before contrasting this use with the generation of a novel entity from three known entities under some conditions. Taking advantage of the belonging of analogical proportions to the setting of logical proportions, we have found out that there exist only two small subsets of logical proportions that define equivalence relations between ordered pairs: the analogical proportion together with a related proportion called paralogy on the one hand, and four conditional logical proportions between conditional events on the other hand.

Taking lessons from the logic of conditional events and its key role in nonmonotonic inference, we have defined an entailment relation between ordered pairs (in agreement with analogical proportions) and appropriate conjunction and disjunction of ordered pairs. Then, we have described a creative inference process using these entailment and operations. First experiments with them have been also reported. However, it should be clear that, in our formal setting, we only capture a particular form of "creative" inference, which is not intended to cover every type of creative analogical inference.

It is clear that the new logic of ordered pairs outlined in the second half of this paper is still in its infancy and many aspects remain to be developed, as well as its possible use in a creative machinery for controlling the derivation of new items from a given entity on the basis of a set of ordered pairs reporting feasible changes between entities. Moreover, we have focused on analogical proportions between entities described by means of Boolean or nominal values, the case of numerical values already investigated in classification, remains to be considered in creativity. Finally, the pairs involved in analogical proportions can be seen as describing changes resulting from actions, which suggests studying relationships with the logic of action.

References

[1] S. Afantenos, H. Prade, and L. Cortez Bernardes. Analogical proportions and creativity: A preliminary study, 2023. arXiv:2310.13500 [cs.CL].

[2] S. D. Afantenos, T. Kunze, S. Lim, H. Prade, and G. Richard. Analogies between sentences: Theoretical aspects - preliminary experiments. In J. Vejnarová and N. Wilson, editors, *Proc. 16th Europ. Conf. on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'21), Prague, Sept. 21-24*, volume 12897 of *LNCS*, pages 3–18. Springer, 2021.

[3] N. Barbot, L. Miclet, and H. Prade. Analogy between concepts. *Artif. Intell.*, 275:487–539, 2019.

[4] N. Barbot, L. Miclet, H. Prade, and G. Richard. Analogical proportions and binary trees. In J.-Y. Béziau, J.-P. Desclés, A. Moktefi, and A. C. Pascu, editors, *Logic in Question. Talks from the Annual Sorbonne Logic Workshop (2011- 2019)*, pages 435–458. Birkhäuser, 2022.

[5] P. F. A. Bartha. *By Parallel Reasoning: The Construction and Evaluation of Analogical Arguments*. Oxford University Press, 2009.

[6] S. Benferhat, D. Dubois, and H. Prade. Nonmonotonic reasoning, conditional objects and possibility theory. *Artificial Intelligence*, 92(1-2):259–276, 1997.

[7] M. Bounhas, K. Mellouli, H. Prade, and M. Serrurier. Possibilistic classifiers for numerical data. *Soft Comput.*, 17(5):733–751, 2013.

[8] M. Bounhas, H. Prade, and G. Richard. Analogy-based classifiers for nominal or numerical data. *Int. J. Approx. Reason.*, 91:36–55, 2017.

[9] M. Bounhas, H. Prade, and G. Richard. Analogy-based classifiers for nominal or numerical data. *Int. J. of Approximate Reasoning*, 91:36 – 55, 2017.

[10] W. Correa Beltran, H. Prade, and G. Richard. Constructive solving of Raven's IQ tests with analogical proportions. *Int. J. Intell. Syst.*, 31(11):1072–1103, 2016.

[11] M. Couceiro, N. Hug, H. Prade, and G. Richard. Analogy-preserving functions: A way to extend Boolean samples. In *Proc. 26th IJCAI Conf*, 1575-1581, 2017.

[12] M. Couceiro, N. Hug, H. Prade, and G. Richard. Behavior of analogical inference w.r.t. boolean functions. In *Proc. 27th IJCAI Stockholm*, pages 2057–2063, 2018.

[13] M. Couceiro, E. Lehtonen, L. Miclet, H. Prade, and G. Richard. When nominal analogical proportions do not fail. In J. Davis and K. Tabia, editors, *Proc. 14th Int. Conf. on Scalable Uncertainty Management (SUM'20), Bozen-Bolzano, Sept. 23-25*, volume 12322 of *LNCS*, pages 68–83. Springer, 2020.

[14] T. Davies and S. Russell. A logical approach to reasoning by analogy. In *Proc. 10th Int. Joint Conf. on Artificial Intelligence (IJCAI-87), Milan, Aug. 23-28*, pages 264–270, 1987.

[15] B. De Finetti. La logique des probabilités. In *Congrès Int. de Philosophie Scientifique, IV. Induction et probabilité*, pages 31–39, Paris,, 1936. Hermann.

[16] H. Dedhia. Bike buyers 1000. `https://www.kaggle.com/datasets/heeraldedhia/bike-buyers`, 2021. Accessed: September 1, 2023.

[17] R. Dreistadt. The use of analogies and incubation in obtaining insights in creative problem solving. *The Journal of Psychology*, 71:159–175, 1969.

[18] D. Dubois and H. Prade. Conditional objects as nonmonotonic consequence relationships. *IEEE Trans. on Syst., Man and Cyber.*, 24:1724–1740, 1994.

[19] D. Dubois and H. Prade. Possibility theory: Qualitative and quantitative aspects. In D. M. Gabbay and Ph. Smets, editors, *Quantified Representation of Uncertainty and Imprecision*, volume 1 of *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, pages 169–226. Kluwer Acad. Publ., 1998.

[20] D. Dubois and H. Prade. An overview of ordinal and numerical approaches to causal diagnostic problem solving. In D. M. Gabbay and R. Kruse, editors, *Handbook of Defeasible Reasoning and Uncertainty Management Systems. Vol. 4. Abductive Reasoning and Learning*, volume 1 of *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, pages 231–280. Kluwer Acad. Publ., 2000.

[21] D. Dubois and H. Prade. Towards a logic-based view of some approaches to classification tasks. In M.-J. Lesot, S. M. Vieira, M. Z. Reformat, J. P. Carvalho, A. Wilbik, B. Bouchon-Meunier,

and R. R. Yager, editors, *Proc. 18th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'20), Lisbon, June 15-19, Part III*, volume 1239 of *CCIS*, pages 697–711. Springer, 2020.

[22] W. F. Edwards. *Likelihood*. Cambridge University Press, 1972.

[23] T. G. Evans. A program for the solution of a class of geometric-analogy intelligence-test questions. In M. L. Minsky, editor, *Semantic Information Processing*, pages 271–353. MIT Press, Cambridge, Ma, 1968.

[24] E. Fix and J. L. Hodges. Discriminatory analysis. nonparametric discrimination: Consistency properties. Report USAF School of Aviation Medicine, Randolph Field, Texas, 1951.

[25] B. Fuchs, J. Lieber, L. Miclet, A. Mille, A. Napoli, H. Prade, and G. Richard. Case-based reasoning, analogy and interpolation. In P. Marquis, O. Papini, and H. Prade, editors, *A Guided Tour of Artificial Intelligence Research. 1 - Knowledge Repreentation, Reasoning and Learning*, pages 307–339. Springer, 2020.

[26] A. Gammerman, V. Vovk, and V. Vapnik. Learning by transduction. In *Proc. 14th Conf. on Uncertainty in AI*, pages 148–155. Morgan Kaufmann, 1998.

[27] D. Gentner. Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7(2):155–170, 1983.

[28] D. Gentner, K. J. Holyoak, and B. N. Kokinov. *The Analogical Mind: Perspectives from Cognitive Science*. Cognitive Science, and Philosophy. MIT Press, Cambridge, MA, 2001.

[29] D. F. Halpern. Analogies as a critical thinking skill. In Dale E. Berger, Kathy Pezdek, and William P. Banks, editors, *Applications of Cognitive Psychology: Problem Solving, Education, and Computing*, pages 75–86. Routledge, 1986.

[30] M. B. Hesse. On defining analogy. *Proceedings of the Aristotelian Society*, 60:79–100, 1959.

[31] S. Klein. Analogy and mysticism and the structure of culture (and Comments & Reply). *Current Anthropology*, 24 (2):151–180, 1983.

[32] S. Klein. The analogical foundations of creativity in language, culture & the arts: the upper paleolithic to 2100 CE. In P. McKevitt, S. O'Nullain, and C. Mulvihill, editors, *Proc. 8th Int. Workshop on the Cognitive Science of Natural Language Processing, Galway, Ireland, Aug. 9-11, 1999*, pages 347–371. Amsterdam: John Benjamin, 2002.

[33] S. Kraus, D. Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44:167–207, 1990.

[34] Y. Lepage and C.-L. Goh. Towards automatic acquisition of linguistic features. In K. Jokinen and E. Bick, editors, *Proc. 17th Nordic Conf. of Computational Linguistics, NODALIDA'09, Odense, Denmark, May 14-16*, pages 118–125. Northern European Association for Language Technology (NEALT), 2009.

[35] J. Lieber, E. Nauer, and H. Prade. Improving analogical extrapolation using case pair competence. In K. Bach and C. Marling, editors, *Proc. 27th Int. Conf. on Case-Based Reasoning (ICCBR '19), Otzenhausen, Sept. 8-12*, volume 11680 of *LNCS*, pages 251–265. Springer, 2019.

[36] J. Lieber, E. Nauer, H. Prade, and G. Richard. Making the best of cases by approximation, interpolation and extrapolation. In M. T. Cox, P. Funk, and S. Begum, editors, *Proc. 26th Int.*

Conf. on Case-Based Reasoning (ICCBR '18), Stockholm, July 9-12, volume 11156 of *LNCS*, pages 580–596. Springer, 2018.

[37] S. Lim, H. Prade, and G. Richard. Classifying and completing word analogies by machine learning. *Int. J. Appr. Reas.*, 132:1–25, 2021.

[38] L. Miclet, S. Bayoudh, and A. Delhay. Analogical dissimilarity: Definition, algorithms and two experiments in machine learning. *J. Artif. Intell. Res.*, 32:793–824, 2008.

[39] L. Miclet and H. Prade. Handling analogical proportions in classical logic and fuzzy logics settings. In *Proc. 10th Eur. Conf. on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'09),Verona*, pages 638–650. Springer, LNCS 5590, 2009.

[40] J. S. Mill. *A System of Logic, Ratiocinative and Inductive, being a connected view of the principles of evidence and the methods of scientific investigation*. Cambridge University Press, 1843. Book III "Of Induction", chap VIII "Of the four method of experimental inquiry", chap. XX: "Of Analogy".

[41] G. Minnameier. Abduction, induction, and analogy. On the compound character of analogical inferences. In L. Magnani, W. Carnielli, and C. Pizzi, editors, *Model-Based Reasoning in Science and Technology: Abduction, Logic, and Computational Discovery*, pages 107–119. Springer Berlin Heidelberg, 2010.

[42] R. Misiewicz. Peirce on analogy. *Trans. of the Charles S. Peirce Society: A Quarterly Journal in American Philosophy*, 56(3):299–325, 2020.

[43] T. M. Mitchell. *Version spaces: An approach to concept learning*. PhD thesis, Stanford University, 1979.

[44] C. S. Peirce. *Philosophical Writings of Peirce*. Dover Publ., New York, selected and edited by J. Buchler, 1955.

[45] C. S. Peirce. *Collected Papers of Charles Sanders Peirce*. Harvard University Press, Cambridge, 1958. Edited by C. Hartshorne & P. Weiss (vol. 1-6), and A. Burks (vol. 7-8).

[46] V. Pirrelli and F. Yvon. Analogy in the lexicon: a probe into analogy-based machine learning of language. In *Proc. 6th Int. Symp. on Human Communication*, 1999. Santiago de Cuba.

[47] G. Polya. *How to Solve It*. Princeton University Press, 2nd ed. 1957, 1945.

[48] H. Prade and G. Richard. From analogical proportion to logical proportions. *Logica Universalis*, 7(4):441–505, 2013.

[49] H. Prade and G. Richard. Homogenous and heterogeneous logical proportions. *IfCoLog J. of Logics and their Applications*, 1(1):1–51, 2014.

[50] H. Prade and G. Richard. Analogical proportions and analogical reasoning - An introduction. In D. W. Aha and J. Lieber, editors, *Proc. 25th Int. Conf. on Case-Based Reasoning (ICCBR'17), Trondheim, June 26-28*, volume 10339 of *LNCS*, pages 16–32. Springer, 2017.

[51] H. Prade and G. Richard. Analogical proportions: From equality to inequality. *Int. J. of Approximate Reasoning*, 101:234 – 254, 2018.

[52] H. Prade and G. Richard. Analogical proportions: Why they are useful in AI. In *Proc. 30th Int. Joint Conf. on Artificial Intelligence (IJCAI-21), (Z.-H. Zhou, ed.) Virtual Event / Montreal, Aug. 19-27*, pages 4568–4576, 2021.

[53] H. Prade and G. Richard. First steps towards a logic of ordered pairs. In Z. Bouraoui and

S. Vesic, editors, *Proc. 17th Europ. Conf. on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'23), Arras, Sept. 19–22*, volume 14294 of *LNCS*, page to appear. Springer, 2023.

[54] S. J. Russell. *The Use of Knowledge in Analogy and Induction*. Pitman, UK, 1989.

[55] G. Shakhnarovich, T. Darrell, and P. Indyk, editors. *Nearest-Neighbor Methods in Learning and Vision*. MIT Press, 2005.

[56] N. Stroppa and F. Yvon. Analogical learning and formal proportions: Definitions and methodological issues. Technical report, ENST, June 2005.

[57] P. H. Winston. Learning and reasoning by analogy. *Commun. ACM*, 23(12):689–703, 1980.

[58] F. Yvon and N. Stroppa. Formal models of analogical proportions. Technical report, Ecole Nationale Supérieure des Télécommunications, no D008, 2006.

Received \jreceived