

# Picking the one that does not fit - A matter of logical proportions

Henri Prade<sup>1</sup> Gilles Richard<sup>1</sup>

<sup>1</sup>IRIT - University of Toulouse - France

## Abstract

Quiz or tests about reasoning capabilities often pertain to the perception of similarity and dissimilarity between situations. Thus, one may be asked to complete a series of entities  $A, B, C$  by an appropriate  $X$ , or to pick the one that does not fit in a list. It has been shown that the first problem can receive a solution by solving analogical proportion equations between the representations of the entities in a logical setting, where we assume that  $X$  should be such that  $A$  is to  $B$  as  $C$  is to  $X$ . In this paper, we focus on the second problem, and we show that it can be properly handled by means of heterogeneous proportions that are the logical dual of the homogeneous proportions involved in the modeling of analogical proportions and related proportions. Thus, the formal setting of logical proportions, to which homogeneous and heterogeneous proportions belong, provides an appropriate framework for handling the two problems in a coherent way. As it already exists for homogeneous proportions, a particular multiple-valued logic extension of heterogeneous proportions is discussed (indeed being an intruder in a group may be a matter of degree).

**Keywords:** Analogy, logical proportions, multiple-valued semantics.

## 1. Introduction

Commonsense reasoning often relies on the perception of similarity as well as dissimilarity between objects or situations. This is why psychological quizzes often test the ability to exploit comparisons in reasoning. There are at least two very different kinds of such tests : i) analogy problems which amount to select some answer  $X$  for completing a series  $A, B, C$  such as  $A$  is to  $B$  as  $C$  is to  $X$  gives the best fit, and ii) “find the odd one out” problems where one has to pick the one that does not fit in a list of 4 or more items.

In the last past years, a *logical* model of analogical proportions, i.e. statements of the form  $A$  is to  $B$  as  $C$  is to  $D$  has been proposed [1, 2, 3], which enables us to compute a representation of  $X$  from the ones of  $A, B$ , and  $C$  (rather than selecting  $X$  from a set of potential candidates). This model takes its roots in a relatively recent work of Lepage [4] who has proposed a set-theoretic view of analogical proportions, and in more ancient works in

formal anthropology by Klein [5] and in psychology by Piaget [6] (see [1, 3] again for the details of the links between these different works). Such a logical view of analogical proportions has been proved to be also successful for solving more difficult IQ tests [7]. Moreover, it has been shown that analogical proportions, which basically express that  $A$  (resp.  $B$ ) differs from  $B$  (resp.  $A$ ) as  $C$  (resp.  $D$ ) differs from  $D$  (resp.  $C$ ), are a special case of so-called *logical proportions* [8]. Roughly speaking, a logical proportion between four terms  $A, B, C, D$  equates similarity or dissimilarity evaluations about the pair  $(A, B)$  with similarity or dissimilarity evaluations about the pair  $(C, D)$ . Among the set of existing logical proportions, 4 *homogeneous* ones (which include the analogical proportion) enjoy outstanding properties and seem particularly attractive for completing missing values in a sequence [3].

In this paper, we introduce and study a set of 4 *heterogeneous* logical proportions, which are dual in some sense of the homogeneous ones. These proportions are then shown to be instrumental for picking out the item that does not fit in a list. The setting of logical proportions appears to be rich enough for coping with two different types of reasoning problems where the ideas of similarity and dissimilarity play a key role. Moreover, a particular graded version of heterogeneous proportions is proposed, which is the counterpart of the one for analogical and related proportions [9]. This gives us some tool for handling situations whose representation involves properties that are a matter of degree.

The paper is organized as follows: we briefly recall the properties of homogeneous proportions and their range of application. In section 3, we provide a graphical representation exhibiting different types of opposition and relationship between analogical proportion, and more generally homogeneous proportions, and the new heterogeneous ones. In section 4, we investigate the core properties of the heterogeneous proportions. Section 5 is devoted to showcase the use of heterogeneous proportions, before concluding in section 6.

## 2. Background on homogeneous proportions

A *logical proportion*  $T(a, b, c, d)$  is a conjunction of 2 equivalences, involving 4 Boolean variables [2], denoted by lower case letters. For instance  $(\bar{a}b \equiv \bar{c}d) \wedge (\bar{a}b \equiv \bar{c}d)$  is the expression of analogical

proportion. Here  $\bar{a}$  is a compact notation for the negation of  $a$ , and  $a\bar{b}$  is short for  $a \wedge \bar{b}$ , and so on.

It has been proved that there exist 120 semantically distinct logical proportions, involving the pair  $(a, b)$  on one side, the pair  $(c, d)$  on the other side of  $\equiv$ , where each variable may be negated or not. Moreover, each logical proportion has exactly 6 lines leading to 1 in its truth table (and the 10 remaining lines lead to 0). Both  $a\bar{b}$  and  $\bar{a}b$  capture the idea of dissimilarities between  $a$  and  $b$ , while  $ab$  and  $\bar{a}\bar{b}$  capture the idea of similarities, positively and negatively. For instance, analogical proportion uses only dissimilarities and could be informally read as *what is satisfied by  $a$  and not by  $b$  is exactly what is satisfied by  $c$  and not by  $d$ , and vice versa*.

In order to identify the logical proportions that may be considered as suitable counterparts of the numerical proportions (which are known as not depending on the representation of the numbers in a particular basis), one property is paramount. This is the so-called *code independency* property:  $T(a, b, c, d) \implies T(\bar{a}, \bar{b}, \bar{c}, \bar{d})$  insuring that the proportion  $T$  holds whether we encode falsity as 0 (resp. truth as 1) or vice versa. Only 8 among the 120 proportions satisfy *code independency* [3] and, in that perspective, these 8 proportions stand out from the crowd: they are shown in Table 2. They can be divided in two groups: the 4 so-called “homogeneous” proportions denoted  $A, R, P, I$  which have been deeply investigated and which are briefly recalled below, and the 4 so-called “heterogeneous” proportions denoted  $H_a, H_b, H_c, H_d$  (for a reason that will be made clear in the following), which are considered and studied in section 4 for the first time.

Homogeneous proportions			
A		R	
$a\bar{b} \equiv c\bar{d} \wedge \bar{a}b \equiv \bar{c}d$		$a\bar{b} \equiv \bar{c}d \wedge \bar{a}b \equiv cd$	
P		I	
$ab \equiv cd \wedge \bar{a}\bar{b} \equiv \bar{c}\bar{d}$		$ab \equiv \bar{c}\bar{d} \wedge \bar{a}\bar{b} \equiv cd$	
Heterogeneous proportions			
$H_a$		$H_b$	
$\bar{a}b \equiv \bar{c}d \wedge \bar{a}b \equiv cd$		$a\bar{b} \equiv \bar{c}d \wedge \bar{a}b \equiv cd$	
$H_c$		$H_d$	
$\bar{a}\bar{b} \equiv \bar{c}d \wedge ab \equiv \bar{c}d$		$\bar{a}\bar{b} \equiv cd \wedge ab \equiv \bar{c}d$	

Table 1: 8 outstanding proportions

## 2.1. Properties of homogeneous proportions

Their truth tables (restricted to the 4-tuples leading to the truth value 1), are derived from their Boolean expressions, and shown in Table 2.1. We observe that there are only 8 distinct 4-tuples appearing in the tables, among 16 candidates, each one having an even number of 0 and 1, and leaving out the 8 remaining tuples having an odd number of

0 and 1. This emphasizes their collective coherence as the whole class of homogeneous proportions. We observe on these truth tables that any combination of 2 homogeneous proportions is satisfiable by 4 4-tuples, any combination of 3 homogeneous proportions is satisfiable by only 2 4-tuples, and that the conjunction  $A(a, b, c, d) \wedge R(a, b, c, d) \wedge P(a, b, c, d) \wedge I(a, b, c, d)$  is not satisfiable.

$A(a, b, c, d)$  is the *analogical* proportion which expresses that  $a$  (resp.  $b$ ) differs from  $b$  (resp.  $a$ ) as  $c$  (resp.  $d$ ) differs from  $d$  (resp.  $c$ ).  $R(a, b, c, d)$  is the *reverse analogical* proportion, where  $R(a, b, c, d) = A(a, b, d, c)$  ( $a$  is to  $b$  as  $d$  is to  $c$ ).  $P(a, b, c, d)$  is named *paralogy* and expresses that what  $a$  and  $b$  have in common,  $c$  and  $d$  have it also. Finally,  $I(a, b, c, d)$  denotes the *inverse paralogy* and expresses that what  $a$  and  $b$  have in common,  $c$  and  $d$  do not have it, and vice-versa. Most of the semantical properties of homogeneous proportions can be easily checked from their truth tables, and may be viewed as counterparts of properties of the numerical (geometrical) proportion  $\frac{a}{b} = \frac{c}{d}$ . For instance, the property  $\frac{a}{b} = \frac{\frac{1}{a}}{\frac{1}{b}}$  parallels the property  $T(a, b, \bar{b}, \bar{a})$  (called *exchange mirroring*) for a logical proportion  $T$  where the negation operator plays the role of the inverse. It has then been shown that, among the homogeneous proportions, only  $A$  and  $I$  satisfy *exchange mirroring*. Table 2.1 summarizes the results: the third column enumerates the homogeneous proportions satisfying the property respectively named and described in the 1st and 2nd columns. Note that the 4 homogeneous proportions satisfy the symmetry property  $T(a, b, c, d) = T(c, d, a, b)$ : the pairs  $(a, b)$  and  $(c, d)$  play symmetrical roles.

A	R	P	I
0 0 0 0	0 0 0 0	0 0 0 0	1 1 0 0
1 1 1 1	1 1 1 1	1 1 1 1	0 0 1 1
0 0 1 1	0 0 1 1	1 0 0 1	1 0 0 1
1 1 0 0	1 1 0 0	0 1 1 0	0 1 1 0
0 1 0 1	0 1 1 0	0 1 0 1	0 1 0 1
1 0 1 0	1 0 0 1	1 0 1 0	1 0 1 0

Table 2: A, R, P, I Boolean truth tables

Property name	Formal definition	Proportion
full identity	$T(a, a, a, a)$	A, R, P
1-full identity	$T(1, 1, 1, 1) \wedge \neg T(0, 0, 0, 0)$	none
0-full identity	$T(0, 0, 0, 0) \wedge \neg T(1, 1, 1, 1)$	none
reflexivity	$T(a, b, a, b)$	A, P
reverse reflexivity	$T(a, b, b, a)$	R, P
sameness	$T(a, a, b, b)$	A, R
symmetry	$T(a, b, c, d) \rightarrow T(c, d, a, b)$	A, R, P, I
permut. of means	$T(a, b, c, d) \rightarrow T(a, c, b, d)$	A, I
permut. of extremes	$T(a, b, c, d) \rightarrow T(d, b, c, a)$	A, I
all permutations	$\forall i, j, T(a, b, c, d) \rightarrow T(p_{i,j}(a, b, c, d))$	I
transitivity	$T(a, b, c, d) \wedge T(c, d, e, f) \rightarrow T(a, b, e, f)$	A, P
semi-mirroring	$T(a, b, \bar{a}, \bar{b})$	R, I
exchange mirroring	$T(a, b, \bar{b}, \bar{a})$	A, I
negation compatib.	$T(a, \bar{a}, b, \bar{b})$	P, I

Table 3: Properties of homogeneous proportions

As can be seen from Table 2.1, homogeneous proportions, and in particular the analogical proportion  $A$ , enjoy properties that parallel properties of numerical proportions. It makes homogeneous pro-

portions a satisfactory Boolean counterpart of numerical proportions.

## 2.2. Equation solving

The idea of proportion is closely related to the idea of extrapolation, i.e. to guess/compute a new value on the ground of existing values. In other words, if for some reason, it is believed or known that a proportion holds between 4 binary items, 3 of them being known, then one may try to infer the value of the 4th one, at least in the case this extrapolation leads to a unique value. In our context, the problem can be stated as follows. Given a logical proportion  $T$  and a 3-tuple  $(a, b, c)$ , does it exist a Boolean value  $x$  such that  $T(a, b, c, x) = 1$ , and in that case, is this value unique? It is easy to see that there are always cases where the equation has no solution since the triple  $a, b, c$  may take  $2^3 = 8$  values, while any proportion  $T$  is true only for 6 distinct 4-tuples. For instance, when we deal with analogy  $A$ , the equations  $A(1, 0, 0, x)$  and  $A(0, 1, 1, x)$  have no solution. And it has been proved, for instance that the analogical equation  $A(a, b, c, x)$  is solvable iff  $(a \equiv b) \vee (a \equiv c)$  holds. In that case, the solution is unique, and is given by  $x = a \equiv (b \equiv c)$ . Similar results hold for the 3 remaining homogeneous proportions.

## 2.3. Applications

Homogeneous proportions lead to surprisingly successful applications in reasoning and in classification tasks. In real world, we generally have to deal with objects which cannot be simply coded with a unique Boolean value and it is necessary to extend what has been done to Boolean vectors. For a given proportion  $T$ , the extension is done componentwise:

$$T(a, b, c, d) \text{ iff } \forall i \in [1, n], T(a_i, b_i, c_i, d_i)$$

All the previous properties still hold for homogeneous proportions and the equation solving process, when successful, provides a complete Boolean vector instead of a unique Boolean value. For instance, let us consider the simple sequence of Figure 2.3. A basic way to encode the problem involves the 5 Boolean predicates  $hasRectangle(hR)$ ,  $hasBlackDot(hBD)$ ,  $hasTriangle(hT)$ ,  $hasCircle(hC)$ ,  $hasEllipse(hE)$  in that order. This leads to the code shown in Figure 2.3. Then the equation solving process provides the expected solution whose code is  $d = (0, 1, 1, 1, 0)$ . Sometimes, one can avoid an encoding involving high level notions such as circle, triangle, etc., by considering that one has a sequence of 3 bitmap pictures, each picture being represented as a huge Boolean vector with much more than 5 components (if the bitmap are  $256 \times 256$ , each vector belongs to  $\mathbb{B}^{2^{16}}$ ). In that case, the method may apply and provide exactly the same solution [7]. Moreover, in [7], this method has been extended to much more

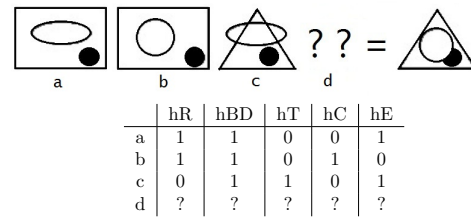


Figure 1: An analogical sequence of pictures and its Boolean coding

sophisticated IQ tests known as the Raven Progressive Matrices (RPM) [10], providing exact and complete solutions for 32 out of 36 tests of the Advanced RPM batch (each time the solution is built rather than chosen in a set of potential candidates).

## 3. Structures of opposition among proportions

The construction of the logical proportions rely on the interplay of similarity and dissimilarity. When this interplay is “homogeneous” with respect to the pairs of binary variables considered, we obtain the 4 homogeneous proportions. The union of the set of 4-tuples that makes true at least one proportion (in fact 3 of them) among the 4 homogeneous ones is made of the 8 4-tuples that have an even number of 0 (and then an even number of 1). Thus, there are 8 remaining 4-tuples that are never true for any homogeneous proportion. Those 8 4-tuples have an odd number of 0 (and then an odd number of 1). It turns out that this set of 8 4-tuples (with an odd number of 0) is also the union of the 4-tuples making true at least one (in fact 3 of them) of a set of 4 logical proportions. These 4 logical proportions are the *heterogeneous* proportions already mentioned at the beginning of Section 2. We call them ‘heterogeneous’ because of the oddity of their truth tables.

Before further investigating the heterogeneous proportions, we use an hexagonal device for laying bare and visualizing the structures of opposition linking analogical proportion to the heterogeneous proportions.

In order to do that, let us provide a brief reminder about the hexagonal extension of the Aristotelian square of opposition proposed and advocated by Blanché [11, 12] (see also [13]). Indeed, it can be easily checked that three *mutually exclusive* situations  $J, K, L$  (represented by subsets of a referential) always give birth to a hexagon as in Figure 2. This hexagon is outstanding for many reasons. It also exhibits the pairwise union of these situations and thus gives birth to 3 Aristotelian squares of opposition whose vertices are respectively  $(J, K, K \cup L, J \cup L)$ ,  $(K, L, L \cup J, K \cup J)$ , and  $(L, J, J \cup K, L \cup K)$ . In each of the 3 squares in Figure 2:

- the diagonals relate disjunct subsets.

- the arrows, e.g. from  $K$  to  $J \cup K$ , express inclusion.
- the thick black lines join vertices which correspond to disjoint sets, whose union does not cover the referential.
- the double lines join overlapping subsets, whose union is the referential.

Note also that each of the vertices  $J, K, L$  is the intersection of its 2 neighboring vertices in the hexagon.

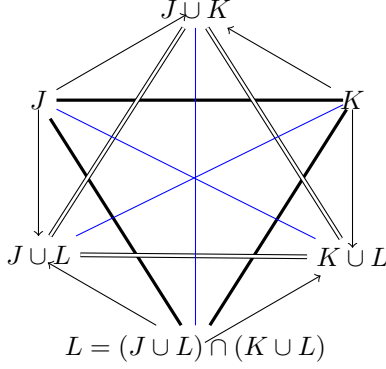


Figure 2: Hexagon induced by a tri-partition  $(J, K, L)$

We are going now to apply the fact that a tri-partition leads to a hexagon of opposition, by considering tri-partitions of sets of 4-tuples of binary valuations. It is clear that there is a unique quaternary connective that makes true a particular subset of 4-tuples of binary valuations, and which is false on all the other subsets of 4-tuples of valuations. We start by considering the whole set of the 16 possible 4-tuples. The partition that we are going to consider is the following one  $J = \{0110, 1001\}$ ,  $K = \{0001, 0010, 0100, 1000, 1110, 1101, 1011, 0111\}$ ,  $L = \{0000, 1111, 1010, 0101, 1100, 0011\}$ . As can be seen, in this partition of the 16 possible 4-tuples into 3 mutually exclusive subsets,  $L$  corresponds to the truth table of the analogy  $A$ ,  $K$  gathers the 8 patterns including an odd number of '1' (and thus of '0'), while  $J$  gathers the two characteristic patterns 0110 and 1001 of the reverse analogy. The corresponding hexagon is pictured in Figure 3. This hexagon has a nice interpretation in terms of analogical dissimilarity in the sense of [14]. Indeed the analogical dissimilarity of the 6 valuation patterns in  $L$  is 0, since they correspond to the 6 cases where  $A$  holds true; the analogical dissimilarity of the 8 valuation patterns in  $K$  is 1, since in each case it is enough to switch one bit for getting a pattern for which proportion  $A$  is true; the analogical dissimilarity of the 2 patterns in  $J$  is 2 since one needs to change 2 bits to get a pattern where proportion  $A$  is true. It is also noticeable that the set of patterns in  $J \cup L$  corresponds to the truth table of a quaternary connective that exactly corresponds to S. Klein [15, 5]'s view of analogy, defined as:

$$Klein(a, b, c, d) \triangleq (a \equiv b) \equiv (c \equiv d),$$

which is closely related to the equation solving problem for proportions  $A, R, P$ . The vertex corresponding to  $K \cup L$ , named "approximate similarity" in Figure 3, is associated with an operator which is true in the cases where the analogical dissimilarity is not maximal. The hexagon of Figure 3 is clearly associated with the analogical proportion  $A$ . There are 3 other similar hexagons associated with each of the 3 other homogeneous proportions, changing  $L$  and  $J$  in the appropriate way. The vertex opposite to Klein's operator gathers the 8 heterogeneous 4-tuples from which the truth tables of the 4 heterogeneous proportions are made.

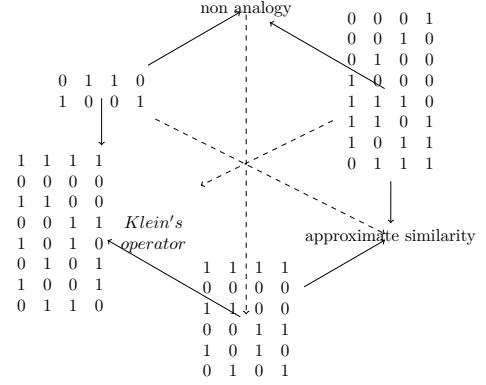


Figure 3: Hexagon induced by analogical dissimilarity

#### 4. Heterogeneous proportions

As highlighted in the introduction, there are 4 other outstanding proportions, namely the heterogeneous proportions, whose truth tables are given in Table 4.

$H_a$	$H_b$	$H_c$	$H_d$
1 1 1 0	1 1 1 0	1 1 1 0	1 1 0 1
0 0 0 1	0 0 0 1	0 0 0 1	0 0 1 0
1 1 0 1	1 1 0 1	1 0 1 1	1 0 1 1
0 0 1 0	0 0 1 0	0 1 0 0	0 1 0 0
1 0 1 1	0 1 1 1	0 1 1 1	0 1 1 1
0 1 0 0	1 0 0 0	1 0 0 0	1 0 0 0

Table 4:  $H_a, H_b, H_c, H_d$  Boolean truth tables

It is stunning to note that these truth tables exactly involve the 8 missing tuples of the homogeneous tables, i.e. those ones having an odd number of 0 and 1. It is remarkable that they satisfy the same association properties as the homogeneous ones: Indeed we observe on their truth tables that any combination of 2 heterogeneous proportions is satisfiable by 4 4-tuples, any combination of 3 heterogeneous proportions is satisfiable by only 2 4-tuples, and that the conjunction  $H_a(a, b, c, d) \wedge H_b(a, b, c, d) \wedge H_c(a, b, c, d) \wedge H_d(a, b, c, d)$  is not satisfiable. This fact contributes to make the heterogeneous proportions the perfect dual of the homogeneous ones.

##### 4.1. Properties

The formal definitions given in Table 2 lead to immediate equivalences between heterogeneous and

homogeneous proportions that we summarize in Table 4.1. Obviously, the heterogeneous proportions

$\mathbf{H_a}$	$\mathbf{H_b}$
$H_a(a, b, c, d) \equiv I(\bar{a}, b, c, d)$	$H_b(a, b, c, d) \equiv I(a, \bar{b}, c, d)$
$H_a(a, b, c, d) \equiv P(\bar{a}, b, \bar{c}, \bar{d})$	$H_b(a, b, c, d) \equiv P(a, \bar{b}, \bar{c}, \bar{d})$
$H_a(a, b, c, d) \equiv P(a, b, c, d)$	$H_b(a, b, c, d) \equiv P(\bar{a}, b, c, d)$
$\mathbf{H_c}$	$\mathbf{H_d}$
$H_c(a, b, c, d) \equiv I(a, b, \bar{c}, d)$	$H_d(a, b, c, d) \equiv I(a, b, c, \bar{d})$
$H_c(a, b, c, d) \equiv P(\bar{a}, \bar{b}, \bar{c}, d)$	$H_d(a, b, c, d) \equiv P(\bar{a}, \bar{b}, \bar{c}, \bar{d})$
$H_c(a, b, c, d) \equiv P(a, b, c, d)$	$H_d(a, b, c, d) \equiv P(a, b, c, d)$

Table 5: Equivalences between heterogeneous and homogeneous proportions

are strongly linked together: for instance, using the symmetry of  $I$ ,

$$H_a(a, b, c, d) \equiv I(\bar{a}, b, c, d) \equiv I(c, d, \bar{a}, b) \equiv H_c(c, d, a, b)$$

In the following we discuss two different ways of viewing these proportions:

- *a semantic viewpoint*: the *full identity* postulate  $T(a, a, a, a)$  asserts that proportion  $T$  holds between identical values. Negating only one variable position generates an *intruder*, as in  $T(\bar{a}, a, a, a)$ ,  $T(a, \bar{a}, a, a)$ ,  $T(a, a, \bar{a}, a)$  and  $T(a, a, a, \bar{a})$ , and leads to new postulates respectively denoted  $\text{Int}_a, \text{Int}_b, \text{Int}_c$  and  $\text{Int}_d$ . For a proportion, satisfying the property  $\text{Int}_a$  means that *the first variable may be an intruder*. Since each postulate  $\text{Int}_a, \text{Int}_b, \text{Int}_c$  and  $\text{Int}_d$  is validated by 2 distinct 4-tuples, it is clear that 3 of them are enough to define a logical proportion having exactly 6 valid 4-tuples. There is no proportion satisfying all these postulates since it would lead to 8 valid 4-tuples, which excludes any logical proportion. It can be easily checked that  $H_a$  satisfies  $\text{Int}_b, \text{Int}_c, \text{Int}_d$  and does not satisfy  $\text{Int}_a$ : We can interpret  $H_a(a, b, c, d)$  as the following assertion *the first position is not an intruder and there is an intruder among the remaining positions*. In other words, having  $H_a$  true means “there is an intruder which is not  $a$ ”. As a consequence,  $H_a(a, b, c, d)$  does not hold when there is no intruder (i.e. an even number of 0) or when  $a$  is the intruder. The same reasoning applies to  $H_b, H_c, H_d$ .

- *a syntactic viewpoint*: here we start from the inverse paralogy  $I$  definition:  $(ab \equiv \bar{c}\bar{d}) \wedge (\bar{a}\bar{b} \equiv cd)$ . To get the definition of an heterogeneous proportion satisfying postulates where the intruder is in position 1, 2 or 4 for instance, we add a negation on the 3rd variable in both equivalences defining  $I$ . Here we get  $H_c$  as:

$$(ab \equiv \bar{c}\bar{d}) \wedge (\bar{a}\bar{b} \equiv \bar{c}d)$$

This process, allowing us to generate the 4 heterogeneous proportions, shows that, in some sense, they are “atomic perturbations” of  $I$ : for this reason and since they are heterogeneous proportions, they have

been respectively denoted  $H_a, H_b, H_c$  and  $H_d$  where the subscript corresponds to:

- the postulate *which is not satisfied* by the corresponding proportion or, equivalently,
- the negated variable in the equivalence with  $I$ .

For instance,  $H_a(a, b, c, d) \equiv I(\bar{a}, b, c, d)$ ,  $H_a$  satisfies  $\text{Int}_b, \text{Int}_c, \text{Int}_d$  and does not satisfy  $\text{Int}_a$ . So, when  $H_a(a, b, c, d) = 1$ , there is an intruder, and  $\bar{a}$  is the value of the intruder, since  $a$  is not the intruder. The different possible cases where  $H_a(a, b, c, d) = 1$  are as follows:

- $(\bar{a}, b, c, d) = (1, 1, 0, 0)$  or  $(0, 0, 1, 1)$  and the intruder is  $b$ .
- or  $(\bar{a}, b, c, d) = (0, 1, 0, 1), (1, 0, 1, 0), (0, 1, 1, 0)$ , or  $(1, 0, 0, 1)$  and the intruder is  $c$  in the first two cases, and  $d$  in the two others.

In other words, there is an intruder in  $(a, b, c, d)$ , which is not  $a$ , iff the properties common to  $\bar{a}$  and  $b$  (positively or negatively) are not those common to  $c$  and  $d$ , and conversely.

As in the case of homogeneous proportions, the semantic properties of heterogeneous proportions are easily derived from their truth tables, which we summarize in Table 4.1. It is clear on their truth

Property name	Formal definition	Proportion
full identity	$T(a, a, a, a)$	none
1-full identity	$T(1, 1, 1, 1) \wedge \neg T(0, 0, 0, 0)$	none
0-full identity	$T(0, 0, 0, 0) \wedge \neg T(1, 1, 1, 1)$	none
reflexivity	$T(a, b, a, b)$	none
reverse reflexivity	$T(a, b, b, a)$	none
sameness	$T(a, a, b, b)$	none
symmetry	$T(a, b, c, d) \rightarrow T(c, d, a, b)$	none
means permut.	$T(a, b, c, d) \rightarrow T(a, c, b, d)$	$H_a, H_d$
extremes permut.	$T(a, b, c, d) \rightarrow T(d, b, c, a)$	$H_c, H_b$
all permutations	$\forall i, j, T(a, b, c, d) \rightarrow T(p_{i,j}(a, b, c, d))$	none
transitivity	$T(a, b, c, d) \wedge T(c, d, e, f) \rightarrow T(a, b, e, f)$	none
$\text{Int}_a$	$T(\bar{a}, a, a, a)$	$H_b, H_c, H_d$
$\text{Int}_b$	$T(a, \bar{a}, a, a)$	$H_a, H_c, H_d$
$\text{Int}_c$	$T(a, a, \bar{a}, a)$	$H_a, H_b, H_d$
$\text{Int}_d$	$T(a, a, a, \bar{a})$	$H_a, H_b, H_c$

Table 6: Properties of heterogeneous proportions

tables, that none of the heterogeneous proportions satisfy neither symmetry nor transitivity. From a practical viewpoint, these proportions are closely related with the idea of spotting the odd one out (the intruder), or if we prefer of picking the one that doesn’t fit among 4 items. This will be further discussed in Section 5, but we first consider the extension of heterogeneous proportions to the case of graded properties with intermediate truth values.

## 4.2. Multi-valued semantics

When we have to handle properties whose satisfaction is a matter of levels, an extension of the Boolean interpretation to multiple-valued models, where the truth values belong to  $[0, 1]$ , is necessary. This has been done for homogeneous proportions in [9, 16] and we study it here for heterogeneous proportions. Roughly speaking, in the case

of  $H_a$ , we may expect that the graded truth value of  $H_a(a, b, c, d)$  estimates something as the extent to which there is an intruder other than  $a$ .

A simple way to proceed for defining a multiple-valued extension for the heterogeneous proportions is to start from the definition given in [9] for multiple-valued paralogy  $P$ , leading to 15 4-tuples fully true, and 18 fully false. The 48 remaining patterns get intermediate truth values given by the general formula (1):

$$P(a, b, c, d) = \min(1 - |\max(a, b) - \max(c, d)|, 1 - |\min(a, b) - \min(c, d)|) \quad (1)$$

which, thanks to the symmetry of  $P$  and stability w.r.t. the permutation of its two first variables, has the following behavior

general case	case $u = v$
$P(1, 1, u, v) = \min(u, v)$	$P(1, 1, u, u) = u$
$P(1, 0, u, v) = \min(\max(u, v), 1 - \min(u, v))$	$P(1, 0, u, u) = \min(u, 1 - u)$
$P(0, 0, u, v) = 1 - \max(u, v)$	$P(0, 0, u, u) = 1 - u$

As a consequence of the equivalences given in Table 4.1, we get the multiple-valued definition for  $H_a$  (we get similar definitions for  $H_b, H_c, H_d$ ), still leading to 15 true patterns, 18 false patterns and 48 with intermediate values:

$$H_a(a, b, c, d) = \min(1 - |\max(a, 1 - b) - \max(c, d)|, 1 - |\min(a, 1 - b) - \min(c, d)|)$$

Let us note that  $H_a(0, 0, u, v) = H_a(1, 1, u, v)$  due to the equivalences  $H_a(0, 0, u, v) = P(0, 1, u, v) = P(1, 0, u, v)$ .

general case	case $u = v$
$H_a(1, 1, u, v) = \min(\max(u, v), 1 - \min(u, v))$	$H_a(1, 1, u, u) = \min(u, 1 - u)$
$H_a(1, 0, u, v) = \min(u, v)$	$H_a(1, 0, u, u) = u$
$H_a(0, 1, u, v) = 1 - \max(u, v)$	$H_a(0, 1, u, u) = 1 - u$

Let us analyze 2 examples to suggest how our definition fits with the intuition.

- Considering the general pattern  $(1, 0, 0, u)$ , its truth value is:
  - $u$  for  $P$ : if  $u$  is close to 1, we are close to the fully true paralogical pattern  $(1, 0, 0, 1)$  and the truth value is high. In the opposite case,  $u$  is close to 0 and we are close to a fully false paralogical pattern  $(1, 0, 0, 0)$ .
  - $1 - u$  for  $H_b, H_c, H_d$ : if  $u$  is close to 1, we are close to the pattern  $(1, 0, 0, 1)$  which is definitely not a valid pattern for  $H_b, H_c, H_d$ : so  $1 - u$  is a low truth value. But if  $u$  is close to 0, we are close to the pattern  $(1, 0, 0, 0)$  which is valid for  $H_b, H_c, H_d$  and  $1 - u$  is a high truth value.
  - finally 0 for  $H_a$  whatever the value of  $u$ : indeed in the 4-tuple  $(1, 0, 0, u)$  there is "an intruder in first position, or no intruder at all", while the semantics of  $H_a$  is just the opposite.
- Back to the graded pattern  $(0.7, 1, 1, 0.9)$  introduced above:

- regarding  $P$ , the truth value as given by the formula is 0.8, i.e. the pattern is close to be a true paralogy.

- regarding the heterogeneous proportions, we understand that we have 2 candidate intruders namely  $a = 0.7$  and  $d = 0.9$ . But they are not equivalent in terms of intrusion and it is more likely to be  $a$  than  $d$ . This is consistent with the fact that the truth value of  $H_a(0.7, 1, 1, 0.9)$  is 0.1 (very low), but the truth value of  $H_d(0.7, 1, 1, 0.9)$  is 0.3 (a bit higher).

- in fact,  $(0.7, 1, 1, 0.9)$  does not give a strong belief that there is an intruder, and it is not a surprise that  $H_b(0.7, 1, 1, 0.9) = H_c(0.7, 1, 1, 0.9) = 0.4$ .

However there are patterns, where the multiple-valued extensions of  $H_x(a, b, c, d)$  do not behave exactly as expected:

- It can be checked that  $H_a(0.5, 0.5, 0.5, 0.5) = H_b(0.5, 0.5, 0.5, 0.5) = H_c(0.5, 0.5, 0.5, 0.5) = H_d(0.5, 0.5, 0.5, 0.5) = 1$ , since  $\forall x \in \{a, b, c, d\}$ ,  $H_x(u, u, u, u) = 2 \min(u, 1 - u)$ . Thus, it is only in the Boolean case that we get the expected value 0. This behavior appears to be related to the fact that  $x \wedge \neg x$  is no longer a contradiction in general. In fact, it can be checked that we also have  $H_a(0, u, 0, u) = H_a(0, u, u, 0) = H_a(1, u, 1, u) = H_a(1, u, u, 1) = 2 \min(u, 1 - u)$ , while  $H_a(u, 0, u, 0) = H_a(u, 0, 0, u) = H_a(u, 1, u, 1) = H_a(u, 1, 1, u) = \min(u, 1 - u)$ . This indicates that when two values are close to 0.5, there are situations where the values of  $H_x$  tend to indicate that there is an intruder, while it is not the case.

- Besides, we have  $H_a(0.9, 0.9, 0.9, 1) = 0.2$ ,  $H_b(0.9, 0.9, 0.9, 1) = 0.2$ ,  $H_c(0.9, 0.9, 0.9, 1) = 0.1$ ,  $H_d(0.9, 0.9, 0.9, 1) = 0.2$ . Note that  $c$  does not behave as  $a$  and  $b$ , as we might expect. More generally  $H_c(u, u, u, 1) = 1 - u$ , while  $H_a(u, u, u, 1) = 1 - u = H_b(u, u, u, 1)$  if  $u \leq 0.5$  and  $H_a(u, u, u, 1) = \min(u, 2(1 - u)) = H_b(u, u, u, 1)$  if  $u \geq 0.5$ ,  $H_d(u, u, u, 1) = \min(u, 1 - |2u - 1|)$ . As can be seen, for the low values of  $u$ ,  $H_a = H_b = H_c$  are high, while  $H_d$  is low, which nicely extends the Boolean case. For high values of  $u$  (less than 1),  $H_a, H_b, H_c$ , and  $H_d$  are low (as expected), but their relative values may be found misleading with regard to a graded view of intruder.

## 5. Finding the odd one out

With homogeneous proportions, as recalled in section 2, we can solve sophisticated IQ tests where a missing item has to be found. Namely, a sequence of items being given and assumed to be incomplete, the problem is to find the item that completes the sequence in a "natural" way among a given set of candidate solutions, or better, to build it. As it is the case for homogeneous proportions, heterogeneous proportions can be also related to the solving of another type of quiz problem. Namely, the

“Finding the odd one out” problem where a complete sequence of 4 or more items being given, we have to find the item that does not fit with the other ones and which is, in some sense, an intruder or an anomaly. On this basis, a complete battery of IQ tests has been recently developed in [17]. The problem of spotting anomalies can be found in other areas as well, e.g., [18]. Solving Odd One Out tests (which are visual tests) has been recently tackled in [19] by using analogical pairing between fractal representation of the pictures. It is worth noticing that the approach of these authors takes its root in the idea of analogical proportion. However, this method relies on the use of similarity/dissimilarity measures rather than referring to a formal logical view of analogical proportion. In the following, we show that heterogeneous proportions provide a convenient way to code and tackle this problem.

Let us first consider the case of 4 items. The idea of intruder in a set of Boolean values amounts to have a unique value distinct from all the others (which are thus identical). Note that when a homogeneous proportion holds between 4 Boolean values, there is no intruder among them. More generally when dealing with Boolean vectors (describing items), we may extend this view by requiring that for at least one component, there is an intruder value in the previous sense. In that case, the corresponding vector (and the associated item) is an *intruder w.r.t. that component*. Then the candidate intruder, if it exists, is the item whose vector is an intruder w.r.t. a maximum number of components. As a consequence, there is no intruder when there is no candidates, or when there are more than one. Besides, note that although in practical quizzes it is often the case that the items are identical on many components, this is not always true in situations where the notion of intruder makes sense.

Let us consider the simple example (*lorry, bus, bicycle, car*) (where the obvious intruder is *bicycle*) shown in Figure 4, with an encoding in terms of 5 binary attributes having a straightforward meaning (‘specLicense’ is short for ‘requires a specific driving license’). When

Figure 4: A simple quiz and its Boolean coding

	canMove	hasEngine	specLicense	has4Wheel	canFly
lorry	1	1	1	1	0
bus	1	1	1	1	0
bicycle	1	0	0	0	0
car	1	1	0	1	0

considering the item componentwise, we see that:

- for  $i = 1, 3, 5$ ,  $H_a(a_i, b_i, c_i, d_i) = H_b(a_i, b_i, c_i, d_i) = H_c(a_i, b_i, c_i, d_i) = H_d(a_i, b_i, c_i, d_i) = 0$ .
- for  $i = 2, 4$ ,  $H_a(a_i, b_i, c_i, d_i) = H_b(a_i, b_i, c_i, d_i) = H_d(a_i, b_i, c_i, d_i) = 1$ .
- for  $i = 2, 4$ ,  $H_c(a_i, b_i, c_i, d_i) = 0$

The indices 1, 3 and 5 are not useful to pick up the intruder because all the heterogeneous proportions have the same truth value. This is not the

case for the indices 2 and 4:  $H_a$  for instance, being equal to 1, insures that there is an intruder (which is not the first element). The intruder is then given by the proportion having the value 0: for instance,  $H_a(a_j, b_j, c_j, d_j) = 1$  together with  $H_c(a_j, b_j, c_j, d_j) = 0$  mean that the assertion  $c$  is *not an intruder* is false, which implies that  $c$  is the intruder w.r.t. component  $j$ . In our example,  $c = \text{bicycle}$  is the unique intruder w.r.t. components 2 and 4, and is thus the global intruder. Note that it could be the case that the values of the proportions refer to *distinct* items as potential intruders. Then, a special procedure is needed to identify a global intruder, when possible, as explained now.

Let us assume that each item  $a$  is represented as a Boolean vector  $(a_1, \dots, a_n)$  where  $n$  is the number of attributes and  $a_i \in \{0, 1\}$ . For each  $x \in \{a, b, c, d\}$ , we compute the number  $\mathcal{N}_x$  of components where  $H_x$  is equal to 0 (rather than 1), namely

$$\mathcal{N}_x = |\{i \in \{1, \dots, n\} \text{ s.t. } H_x(a_i, b_i, c_i, d_i) = 0\}|$$

In case where  $\mathcal{N}_a = \mathcal{N}_b = \mathcal{N}_c = \mathcal{N}_d = n$ , it means that homogeneous proportions hold for each component, and obviously, there is no intruder. More generally, in case where  $\mathcal{N}_a = \mathcal{N}_b = \mathcal{N}_c = \mathcal{N}_d = p < n$ , it means that on  $n - p$  components, heterogeneous proportions hold for each of the items  $a, b, c, d$ . Thus,  $a, b, c, d$  are equally referred to as potential intruders (they are intruders w.r.t. the same number of components), and there is no intruder. Lastly,  $x$  is an intruder only if  $\mathcal{N}_x > \mathcal{N}_y$  for each  $y \neq x$ . Otherwise,  $\exists u, v \in \{a, b, c, d\}$  s.t.  $u \neq v$  and  $\mathcal{N}_u = \mathcal{N}_v$ , and there is no clear intruder.

In the case where we have to find the odd one out among more than 4 items, an option is to consider all the subsets of 4 items. For each such subset, we apply the previous method to exhibit an intruder (if any). Then the global intruder will be the one which is intruder for the maximum number of subsets.

Thanks to the multiple-valued extension, we may think of generalizing this method to the non Boolean case where each item  $a$  is represented as a real vector  $(a_1, \dots, a_n)$  and  $a_i \in [0, 1]$ . According to the end of section 4.2, we are in position to detect an intruder w.r.t. a component, only if 3 heterogeneous proportions are high, and the remaining one is low. When such a condition is satisfied to some extent on one or several components, and if all these components identify the same intruder  $x$ , this  $x$  is the global intruder as in the Boolean case. When the considered components refer to different intruders, it would be necessary to take into account the degree to which each of these components points out a candidate as an intruder. This degree may be estimated as an increasing function of the difference between the two smallest truth-values among  $H_a, H_b, H_c, H_d$  (the smallest referring to the intruder). Vectors of intruder degrees associated with each candidate are then compared on a leximax basis.

## 6. Concluding remarks

The Boolean modeling of analogical proportions leads to a set of 120 derived proportions: the logical proportions. Among this set, 4 proportions, the homogeneous ones, have already been deeply investigated, not only in terms of formal properties but also in terms of practical applications. In this paper, we have shown that 4 other proportions stand out from the set of logical proportions, namely the heterogeneous proportions, because of their syntactic definitions and their semantics. From a syntactic viewpoint, they differ from the homogeneous ones with the addition of a negation in their definition, which leads to heterogeneous truth tables where an odd number of 1 is necessary for such a proportion to hold. From a semantic viewpoint, they satisfy properties which are not satisfied by any of the homogeneous proportions. More than that, while the use of homogeneous proportions leads to mechanisms allowing to deal with IQ tests of the type “Find the Missing One”, heterogeneous proportions underlies a mechanism allowing us to deal with quiz of the type “Find the Odd One Out”. Thus, both from a formal viewpoint and from an applicative viewpoint, heterogeneous proportions appear as a perfect dual of the homogeneous ones. Ultimately, logical proportions provide an elegant framework to deal with IQ tests, from Raven Progressive Matrices to Find the Odd One Out quizzes, in a uniform way. Still the investigation of the most appropriate multiple-valued extension(s) for heterogeneous proportions remains an open question.

## References

- [1] 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.
- [2] H. Prade and G. Richard. Reasoning with logical proportions. In *Proc. 12th Int. Conf. on Principles of Knowledge Representation and Reasoning, KR 2010, Toronto, May 9-13, 2010 (F. Z. Lin, U. Sattler, M. Truszczynski, eds.)*, pages 545–555. AAAI Press, 2010.
- [3] H. Prade and G. Richard. Homogeneous logical proportions: Their uniqueness and their role in similarity-based prediction. In G. Brewka, T. Eiter, and S. A. McIlraith, editors, *Proc. 13th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'12)*, Roma, June 10-14, pages 402–412. AAAI Press, 2012.
- [4] Y. Lepage. De l’analogie rendant compte de la commutation en linguistique. *Habilit. à Diriger des Recher.*, Univ. J. Fourier, Grenoble, 2003.
- [5] S. Klein. Analogy and mysticism and the structure of culture (and Comments & Reply). *Current Anthropology*, 24 (2):151–180, 1983.
- [6] J. Piaget. *Logic and Psychology*. Manchester Univ. Press, 1953.
- [7] W. Correa, H. Prade, and G. Richard. When intelligence is just a matter of copying. In *Proc. 20th Eur. Conf. on Artificial Intelligence, Montpellier, Aug. 27-31*, pages 276–281. IOS Press, 2012.
- [8] H. Prade and G. Richard. Logical proportions - Typology and roadmap. In E. Hüllermeier, R. Kruse, and F. Hoffmann, editors, *Computational Intelligence for Knowledge-Based Systems Design: Proc. 13th Int. Conf. on Inform. Proc. and Manag. of Uncertainty (IPMU'10)*, Dortmund, June 28 - July 2, volume 6178 of LNCS, pages 757–767. Springer, 2010.
- [9] H. Prade and G. Richard. Multiple-valued logic interpretations of analogical, reverse analogical, and paralogical proportions. In *Proc. 40th IEEE Int. Symp. on Multiple-Valued Logic (ISMVL'10)*, pages 258–263, Barcelona, 2010.
- [10] J. Raven. The Raven’s progressive matrices: Change and stability over culture and time. *Cognitive Psychology*, 41(1):1 – 48, 2000.
- [11] R. Blanché. Sur l’opposition des concepts. *Theoria*, 19:89–130, 1953.
- [12] R. Blanché. *Structures Intellectuelles. Essai sur l’Organisation Systématique des Concepts*. Vrin, Paris, 1966.
- [13] J.-Y. Béziau. The power of the hexagon. *Logica Universalis*, 6 (1-2):1–43, 2012.
- [14] L. Miclet, S. Bayoudh, and A. Delhay. Analogical dissimilarity: definition, algorithms and two experiments in machine learning. *JAIR*, 32, pages 793–824, 2008.
- [15] S. Klein. Culture, mysticism & social structure and the calculation of behavior. In *Proc. 5th Europ. Conf. in Artificial Intelligence (ECAI'82)*, Paris, pages 141–146, 1982.
- [16] H. Prade and G. Richard. Analogical proportions and multiple-valued logics. In L. C. van der Gaag, editor, *Proc. 12th Eur. Conf. on Symb. and Quantit. Appr. to Reasoning with Uncertainty (ECSQARU'13)*, Utrecht, Jul. 7-10, LNAI 7958, pages 497–509. Springer, 2012.
- [17] A. Hampshire. The odd one out tests of intelligence. 2010. <http://www.cambridgebrainsciences.com/browse/reasoning/test/oddoneout>.
- [18] L. Akoglu, M. McGlohon, and C. Faloutsos. Oddball: Spotting anomalies in weighted graphs. In M. Javeed Zaki, J. Xu Yu, B. Ravindran, and V. Pudi, editors, *Proc. 14th Pacific-Asia Conf. Advances in Knowledge Discovery and Data Mining (PAKDD'10)*, Hyderabad, India, June 21-24, Part II, volume 6119 of LNCS, pages 410–421. Springer, 2010.
- [19] K. McGregor and A. K. Goel. Finding the odd one out: a fractal analogical approach. In *8th ACM conference on Creativity and cognition*, pages 289–298, 2011.