# Analogical Proportions: Why They Are Useful in AI

**Henri Prade and Gilles Richard**
IRIT, CNRS and Université Paul Sabatier, Toulouse, France
{henri.prade, gilles.richard }@irit.fr

## Abstract

This paper presents a survey of researches in analogical reasoning whose building block are analogical proportions which are statements of the form "$a$ is to $b$ as $c$ is to $d$". They have been developed in the last twenty years within an Artificial Intelligence perspective. After discussing their formal modeling with the associated inference mechanism, the paper reports the main results obtained in various AI domains ranging from computational linguistics to classification, including image processing, I.Q. tests, case based reasoning, preference learning, and formal concepts analysis. The last section discusses some new theoretical concerns, and the potential of analogical proportions in other areas such as argumentation, transfer learning, and XAI.

## 1 Introduction

Exploiting analogies has been recognized for a long time as a useful approach to problem solving. For instance, Polya's famous book *How to solve it?* [Polya, 1945] is a good advocacy in favor of an extensive search for analogies in order to discover solutions of mathematical problems. Even if analogical inference is brittle since it offers no guarantee on the validity of its derived conclusions, it has the merit of laying bare candidate answers or solutions that further investigations may confirm as being ultimately valid. Analogies, as a useful parallel between two situations enabling us to foresee that what is true in the first situation may be true as well in the second situation, has been extensively studied not only by philosophers and psychologists [Hesse, 1966; Gentner *et al.*, 2001; Hofstadter and Mitchell, 1995; Hofstadter and Sander, 2013], but also by AI scientists [Boy de la Tour and Caferra, 1987; Veloso and Carbonell, 1993; Melis and Veloso, 1998]. This survey focuses on *analogical proportions* (APs for short in this survey), i.e., statements of the form "$a$ is to $b$ as $c$ is to $d$" such as "cow is to calf as mare is to foal", usually denoted $a : b :: c : d$. APs can be related to analogies between two situations if the pairs of items $(a, b)$ and $(c, d)$ refer respectively to these two situations and the items are put in correspondence [Hesse, 1959]. Starting in the late 1990's there has been a new trend of research in analogical studies dealing with APs: this is what is reviewed here. Although the idea of analogy seems to primarily focus on similarity, APs deal as much with dissimilarity. APs take place as soon as we are comparing items described by multiple attributes. An AP states that the comparisons between $a$ and $b$, and between $c$ and $d$, in terms of similarity and dissimilarity, yield results that are very much alike. APs are quite pervasive; in particular, if $a$ and $d$ differ on at least two attributes, one can find (several) pairs $(x, y)$ such that $a : x :: y : d$ holds and the four items are distinct [Couceiro *et al.*, 2018]. Analogical reasoning is often associated with the idea of creative thinking. APs are creative as well, since the equation $a : b :: c : x$ between items described by multiple attributes, when it is solvable, has a solution $d$ different from $a$, $b$, $c$ in general. However, feature by feature, $(c, x)$ will be a copy of $(a, b)$, as we shall see. These properties and others presented in the next section, make APs quite useful in many tasks. The paper is organized into three main sections. Section 2 presents how APs can be modeled and used for inference task. Other proportions, strongly related to APs, are also described. Section 3 surveys domains where APs have been used with success. Section 4 highlights tracks worth to be investigated.

## 2 Modeling and Related Issues

The idea of AP between word items dates back to Aristotle [Aristotle, 2011] (at least), and could have been inspired by a parallel with (geometric) numerical proportions, namely $\frac{a}{b} = \frac{c}{d}$; see [Prade and Richard, 2013] for details. This parallel explains the postulates for APs recalled below. Note that it fits as well with *arithmetic* proportions that state the equality of two differences (rather than two ratios) between numbers: $a - b = c - d$ (can be got as a logarithmic transform of $\frac{a}{b} = \frac{c}{d}$).

### 2.1 Postulates

Given a set of items $X$, APs is a quaternary relation supposed to obey the 3 following postulates (e.g. [Lepage, 2001]): $\forall a, b, c, d \in X$,

1. $a : b :: a : b$ (*reflexivity*);
2. $a : b :: c : d \to c : d :: a : b$ (*symmetry*);
3. $a : b :: c : d \to a : c :: b : d$ (*central permutation*).

These postulates have straightforward consequences like:

- $a : a :: b : b$ (*identity*);
- $a : b :: c : d \to b : a :: d : c$ (*internal reversal*);

- $a : b :: c : d \rightarrow d : b :: c : a$ (*extreme permutation*);
- $a : b :: c : d \rightarrow d : c :: b : a$ (*complete reversal*).

Among the 24 permutations of $a, b, c, d$, the previous postulates induce 3 distinct classes each containing 8 distinct proportions regarded as equivalent due to postulates: $a : b :: c : d$ has in its class $c : d :: a : b$, $c : a :: d : b$, $d : b :: c : a$, $d : c :: b : a$, $b : a :: d : c$, $b : d :: a : c$, and $a : c :: b : d$. But $b : a :: c : d$ and $a : d :: c : b$ do not belong to the class of $a : b :: c : d$ and are elements of the two other classes.

## 2.2 Models

There are three main approaches to the modeling of APs that we briefly present: logical, algebraic, and complexity-based.

**Logical View**

There exist two modelings of analogical reasoning respectively based on first order logic [Russell, 1989], and second order logic [Gust *et al.*, 2006]. We do not discuss them further since they are not based on APs. An AP can be viewed as a quaternary propositional logic connective when $X$ is the Boolean set $\mathbb{B} = \{0, 1\}$ [Miclet and Henri Prade, 2009]:

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

It makes explicit that "$a$ differs from $b$ as $c$ differs from $d$ (and vice-versa)". It is easy to check that this formula is only valid for the 6 valuations in Table 1. This set of 6 valuations is

| | |
|---|---|
| $0 : 0 :: 0 : 0$ | $1 : 1 :: 1 : 1$ |
| $0 : 1 :: 0 : 1$ | $1 : 0 :: 1 : 0$ |
| $0 : 0 :: 1 : 1$ | $1 : 1 :: 0 : 0$ |

Table 1: 6 Valid valuations for strong Boolean AP

the *minimal* (w.r.t. set inclusion) Boolean model [Prade and Richard, 2018] obeying the 3 postulates of analogy. $a : b :: c : d$ can be equivalently written in a way that reminds a well-known property of numerical proportions ($a \cdot d = b \cdot c$ and $a + d = b + c$):

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

Note that this writing emphasizes similarity, and not dissimilarity as the first one. It can be seen on Table 1 that 1 and 0 play a symmetrical role, which makes the definition *code-independent*. The Boolean modeling of APs can be related to a set-based view [Lepage, 2003; Miclet and Henri Prade, 2009]. Besides, it has its roots in an early proposal by S. Klein [Klein, 1983] who suggested to also include $(0, 1, 1, 0)$ and $(1, 0, 0, 1)$, which would validate the highly debatable consequence $a : b :: c : d \rightarrow b : a :: c : d$. When $a, b, c, d$ take their values in a finite set $X$ such that $|X| > 2$ (nominal case where the values are discrete or categorical), APs can be taken as true for 3 patterns, namely $(s, s, s, s), (s, t, s, t), (s, s, t, t)$ for $s, t \in X$, which clearly include as a particular case the 6 patterns of Table 1. This generalization still agrees with the postulates.

When $X = [0, 1] \subset \mathbb{R}$, there are basically two options [Dubois *et al.*, 2016] for extending the logical definition of $a : b :: c : d$ into multiple-valued logic expressions, where $a : b :: c : d$ becomes a matter of degree. They both generalize the Boolean and the nominal cases, and are code-independent (with respect to the negation $\nu(a) = 1 - a$)). One is such that $a : b :: c : d = 1$ only for $x : y :: x : y$ and $x : x :: y : y$ (with possibly $x = y$) where $x, y \in [0, 1]$. The other is such that $a : b :: c : d = 1$ if and only if $a - b = c - d$. See [Dubois *et al.*, 2016] for expressions and details. Thus one can define APs for Boolean and nominal features, as well as numerical features (once renormalized within $[0, 1]$). Let us mention the special case of "continuous AP" of the form $a : b :: b : c$ [Prade and Richard, 2018; Essid *et al.*, 2020], which are non trivial only in a numerical setting.

To deal with items represented by *vectors* of feature values, AP definitions are extended componentwise from $X$ to $X^n$:

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

Using arithmetic proportion $a - b = c - d$ as a model for $\mathbb{R}$ [Rumelhart and Abrahamson, 1973] leads in $\mathbb{R}^n$ to the well-known parallelogram view considering $a, b, c, d$ as vectors in $\mathbb{R}^n$. It agrees with Table 1, but then $a - b \in \{-1, 0, 1\}$.

**Algebraic View**

Stroppa and Yvon [Yvon and Stroppa, 2006] have given another more general definition of AP, based on the notion of *factorization*, when the set of items is a commutative semi-group $(X, \oplus)$.

$(x, y, z, t) \in X^4$ is an AP ($x : y :: z : t$) if:

1) either $(y, z) \in \{(x, t), (t, x)\}$,
2) or $\exists (x_1, x_2, t_1, t_2) \in X^4$ such that $x = x_1 \oplus x_2$, $y = x_1 \oplus t_2$, $z = t_1 \oplus x_2$ and $t = t_1 \oplus t_2$.

This definition satisfies the postulates of APs. The second part of the definition formally expresses that $x$ and $y$ (resp. $z$ and $t$) have in common $x_1$ (resp. $t_1$), while $x$ and $y$, as $z$ and $t$, differ from $x_2$ to $t_2$, such as in `mpr:mps::qr:qs`. This definition, which coincides with the logical view on a Boolean lattice, allows us to define APs when $X$ is equipped with various structures (e.g., [Miclet *et al.*, 2014], or non-distributive lattices in relation with formal concept analysis [Barbot *et al.*, 2019a]).

**Complexity-Based View**

A machine learning-oriented view where APs are interpreted in terms of Kolmogorov complexity has been presented in [Cornuéjols, 1996]; (see also [Murena *et al.*, 2020]). The main idea is to measure the size of the minimal program $pr(a, b)$ able to transform $a$ into $b$ ($a$ and $b$ supposed to be formal expressions), in other words, to consider Kolmogorov complexity as the standard for complexity definition (instead of Shannon entropy for instance). If the size of $pr(a, b)$ and the size of $pr(c, d)$ are approximately equal, then $a : b :: c : d$ holds, still expressing that $a$ and $b$ differ as $c$ and $d$ differ [Bayoudh *et al.*, 2012]. Roughly speaking, it amounts to consider the problem of solving an AP equation $a : b :: c : x$ from an inductive learning perspective: knowing $a, b, c$, find a $d$ minimizing the complexity of $a : b :: c : d$ considered as a formula. The truth table of Boolean APs has been also analyzed in terms of Kolmogorov complexity, showing that the complexity of the patterns that make an AP true is smaller than the complexity of any other pattern [Prade and Richard, 2018].

**Other Views**

From a functional viewpoint, $x : f(x) :: y : f(y)$ looks like a prototypical AP, meaning that "$x$ is to $f(x)$ as $y$ is to $f(y)$", just applying the same function $f$ for obtaining $f(x)$ and $f(y)$ from $x$ and $y$ respectively. This underlies the approach developed in COPYCAT [Hofstadter and Mitchell, 1995] for completing $a, b, c$ with a plausible $d$. Note that this view does not fit with the standard arithmetic proportion (e.g., $a : a^2 :: b : b^2$, but $a - a^2 \neq b - b^2$). Due to central permutation, one is led to postulate that we have $x : y :: f(x) : f(y)$ and then it also exists $g$ such that $x : g(x) :: f(x) : f(g(x))$ where $f$ and $g$ commute; see [Barbot *et al.*, 2019b] for a detailed discussion.

The relation-based view, generalizing the functional view, is as simple as: $a : b :: c : d$ holds if there exists a relation $R \subseteq X \times X$ such that $R(a, b) = R(c, d)$. It raises a number of questions, including a risk of triviality if $R$ is too general, and the problem of accommodating (or not) central permutation. See [Lim *et al.*, 2021] for a discussion.

### 2.3 Inference

This inference principle, first proposed in [Pirrelli and Yvon, 1999] for nominal values, can be stated as follows:

$$\frac{\forall i \in \{1, ..., n\},\ \ a_i : b_i :: c_i : d_i \text{ holds}}{\forall j \in \{n+1, ..., m\},\ \ a_j : b_j :: c_j : d_j \text{ holds}}$$

As can be seen, we transfer knowledge from some components of source vectors to their remaining components, implicitly assuming that the values of the $n$ first components determine the values of the others. When $n + 1 = m$, this inference pattern can be used to predict, e.g., the class of $d$, say $cl(d)$, from $cl(a), cl(b), cl(c)$, after checking that $a : b :: c : d$ holds componentwise. This amounts to assume that the equation $cl(a) : cl(b) :: cl(c) : x$ has a solution. But it is not always the case (e.g., if $0 : 1 :: 1 : x$). It is easy to see that the AP equation $a : b :: c : x$ is solvable if and only if $a = b$ or $a = c$ in the nominal case. Then, the *unique* solution is given by $x = c$ if $a = b$ and $x = b$ if $a = c$. The inference pattern can be generalized when $a, b, c, d$ involve numerical features [Bounhas *et al.*, 2017a]. It is the basis of the AP solving process: finding $x$ such that $a : b :: c : x$. In [Bounhas *et al.*, 2017a], it is shown that the above inference pattern can be derived from a basic analogical pattern that derives $Q(y)$ from $P(x)$, $P(y)$ and $Q(x)$ (e.g., [Russell, 1989]).

### 2.4 Related Proportions

A logical proportion[1] $T(a, b, c, d)$ [Prade and Richard, 2013] is the conjunction of two logical equivalences, each one between indicators for $(a, b)$ and indicators for $(c, d)$. By indicators, we mean operators such that $a \wedge \neg b$, $\neg a \wedge b$ expressing dissimilarity, or such that $a \wedge b$, $\neg a \wedge \neg b$ expressing positive and negative similarity respectively, and the like with $c$ and $d$. It has been established that there are 120 syntactically and semantically distinct logical proportions. Each of them are only true for 6 patterns (and false for the $10 = 2^4 - 6$ others). Among them, 8 are particularly noticeable as the only ones to be code-independent (i.e., $T(a, b, c, d) = T(\neg a, \neg b, \neg c, \neg d)$).

These 8 proportions split into 4 *homogeneous* proportions that are symmetrical (one can exchange $(a, b)$ with $(c, d)$) and 4 *heterogeneous* ones that are not symmetrical. Homogeneity here refers to the fact that in the expression of the proportions, both equivalences link indicators of the same kind (similarity or dissimilarity), while in the case of heterogeneous proportions they link indicators of opposite kinds. Moreover, the homogeneous proportions are true only for 6 patterns involving even numbers of 0 and 1, while the 4 heterogeneous proportions are true for 6 patterns of the form $(s, t, t, t)$, $(t, s, t, t)$, $(t, t, s, t)$, $(t, t, t, s)$ with $s,\ t \in \{0, 1\}$. The four heterogeneous proportions express that there is an intruder among $\{a, b, c, d\}$, which is not $a$, which is not $b$, which is not $c$, and which is not $d$, respectively [Prade and Richard, 2014].

Among the 4 code independent homogeneous proportions, 3 of them are directly related to the AP $a : b :: c : d$. Apart AP itself, they are named "reverse analogy" (Rev) and "paralogy" (Par) and are directly related to AP through a permutation, namely $Rev(a, b, c, d)) = b : a :: c : d$ and $Par(a, b, c, d)) = c : b :: a : d$. Interestingly enough, paralogy can be written $\text{Par}(a, b, c, d) = ((a \wedge b) \equiv (c \wedge d)) \wedge ((\neg a \wedge \neg b) \equiv (\neg c \wedge \neg d))$, which expresses that what $a$ and $b$ have in common (positively or negatively), $c$ and $d$ have it also, and conversely. The existence of these three related proportions mirrors the fact that given 3 non aligned points in $\mathbb{R}^2$, there are 3 ways to build a parallelogram from them [Prade and Richard, 2013]. The fourth homogeneous proportion is called "inverse paralogy" (Inv) and defined by $Inv(a, b, c, d) = ((a \wedge b) \equiv (\neg c \wedge \neg d)) \wedge ((\neg a \wedge \neg b) \equiv (c \wedge d))$, which expresses a kind of "orthogonality" between the pairs $(a, b)$ and $(c, d)$: "what $a$ and $b$ have in common, $c$ and $d$ do not have it and conversely". Remarkably enough $Inv$ is the unique proportion among the 120 that is stable under any permutation of the 4 terms [Prade and Richard, 2013].

## 3 Some Achievements

APs (as well as some other related proportions) have been used with success in diverse fields of AI that we survey now.

### 3.1 Computational Linguistics

Natural language processing is a field of choice to investigate the use of APs such as the famous "man:king::woman:queen" example. The pioneering work of [Rumelhart and Abrahamson, 1973] leads the way on this track: it is all about solving APs on concepts represented as vectors. Solving analogies on words, i.e., given 3 words $a, b, c$, looking for a word $d$ such that the proportion $a : b :: c : d$ holds can be tackled from diverse angles. For instance, if $a$ and $b$ belong to a language $L_1$, $c$ belonging to another language $L_2$, it allows to automatically translate the word $b$ into its counterpart in $L_2$. This is done for instance in [Lepage, 1998; Lepage and Denoual, 2005]. More recently, this has also been developed in [Bayoudh *et al.*, 2012; Murena *et al.*, 2020] with a constraint of minimal complexity (in the sense of Kolmogorov complexity).

[1] Piaget [Piaget, 1953] uses this name for logical expressions of the same form as the second Boolean formula we gave for $a : b :: c : d$.

From a different perspective, in the past 10 years, the assumption that words that appear in "similar" contexts tend to have "similar" meanings has been the foundation of all the successful embedding techniques such as word2vec [Mikolov *et al.*, 2013], GloVe ([Pennington *et al.*, 2014], etc. Turning words $a, b, c, d$ into vectors $a', b', c', d'$ and considering the parallelogram perspective telling that, for the word analogy $a : b :: c : d$ to hold, $(a', b', c', d')$ should constitute a parallelogram in the embedding space[2], has led to powerful mechanisms to achieve diverse tasks such as:

- Solving the analogical equation $a : b :: c : x$ where $x$ is an unknown word. It is amazing to note that solving word analogies has become one of the most popular benchmarks for quality of word embeddings [Linzen, 2016; Allen and Hospedales, 2019]. The arithmetic view ($a - b = c - d$) applied to words represented as vectors, is often not sufficient to find an $x$ referring to (or close to) an existing word. More sophisticated formulas have been proposed to better match an existing word ([Levy and Goldberg, 2014; Drozd *et al.*, 2016]). The problem can also be considered as a regression task in a vector space as in [Lim *et al.*, 2021].
- Focusing on what is known as relation induction [Bouraoui *et al.*, 2018; Lu *et al.*, 2019], i.e., seeking to classify pairs of words $(a, b)$. With the implicit assumption that if 2 pairs of words $(a, b)$ and $(c, d)$ are in the same class, then $a : b :: c : d$ should constitute an AP.
- Recently, in [Diallo *et al.*, 2019; Zhu and de Melo, 2020], AP between words have been extended to sentences. Combining classical word embeddings to get "sentence embeddings", the authors are able to solve APs with practical application to chatbots for instance. The underlying model still is the parallelogram.

### 3.2 Image Processing

Analogies between images has long been applicable to a wide variety of image related problems, including image filtering, texture synthesis, texture transfer, artistic filters, etc. [Hertzmann *et al.*, 2001]. Recently, image processing has largely benefited from the emergence of deep neural networks, involving subtle operators like convolution, pooling, dropout, etc. One of the most important tools is the concept of distance, leading to a lot of works in what is known as distance metric learning (see [Law *et al.*, 2017] for a more extensive description). When considering images, the concept of AP still makes sense as can be seen in the work of [Hwang *et al.*, 2013; Law *et al.*, 2017]. For instance, when talking about pictures, $pic(collie) : pic(dalmatian) :: pic(lion) : pic(leopard)$ (with informal notations) is visually acceptable. [Hwang *et al.*, 2013] use the concept of Analogy-preserving Semantic Embedding (ASE) to classify pictures. The analogical view is still the parallelogram: the authors learn an embedding $u$ of pictures $a, b, c, d$ into a vector space where analogy is preserved thanks to a specific regularizer $R$ enforcing the proportion. $R$ (to be minimized) is

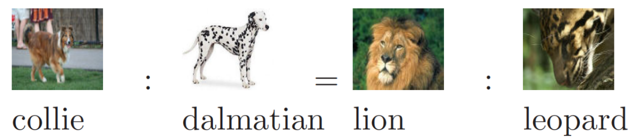


Figure 1: Analogical proportion between pictures

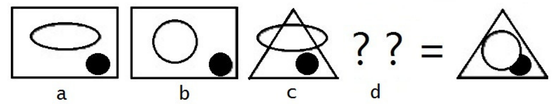


Figure 2: A simple analogical sequence of pictures

just a weighted sum of $||u(a) - u(b)|| + ||u(c) - u(d)||)$ and $||u(a) - u(c)|| + ||u(b) - u(d)||)$ (involving $L^2$ norms), taking into account the central permutation postulate. With a distance learning objective, [Law *et al.*, 2017] exploit analogical constraints over quadruplets of images such as "image $a$ is closer to image $b$ than image $c$ is to image $d$", formalized as $dist(a, b) \leq dist(c, d)$ where $dist$ has to be learned, which is the basis of their *Qwise* method. This has suggested the development in [Prade and Richard, 2018] of a logical framework for analogical *inequalities* for having an *oriented* modeling of such constraints (not captured by a distance since $dist(a, b) = dist(b, a)$).

### 3.3 I.Q. Tests

Consider a completion test such as the example of Figure 2 where pictures $a$, $b$, $c$ are given, and $d$ has to be found (in the usual tests, among a set of potential solutions). The problem can be encoded using the 5 Boolean predicates $hasSquare(hS)$, $hasBlack\ Dot(hBD)$, $hasTriangle(hT)$, $hasCircle(hC)$, $hasEllipse(hE)$ (taken in that order).
This leads to the code of Figure 3. Applying componentwise the AP solving process (described in subsection 2.3), we get $\vec{x} = (0, 1, 1, 1, 0)$, i.e., the code of the expected solution. The approach is *constructive* since the missing picture $\vec{x}$ is obtained by computation from $\vec{a}, \vec{b}, \vec{c}$. This contrasts with classical AI approaches to this problem, pioneered by Thomas Evans [Evans, 1968], where $d$ has to be chosen among a set of candidate pictures containing the right answer: the change between $\vec{a}$ and $\vec{b}$ is compared with the change between $\vec{c}$ and $\vec{x}$ for each potential $\vec{x}$, leading to choose $\vec{x}$ as the one maximizing similarity between the changes. In the above example, the solution could also be obtained by applying the AP solving

| | $hS$ | $hBD$ | $hT$ | $hC$ | $hE$ |
|---|---|---|---|---|---|
| **a** | 1 | 1 | 0 | 0 | 1 |
| **b** | 1 | 1 | 0 | 1 | 0 |
| **c** | 0 | 1 | 1 | 0 | 1 |
| **x** | ? | ? | ? | ? | ? |

Figure 3: A Boolean coding for Figure 2

[2]It is worth to remember that this was exactly the idea of [Rumelhart and Abrahamson, 1973] but with another representation space.

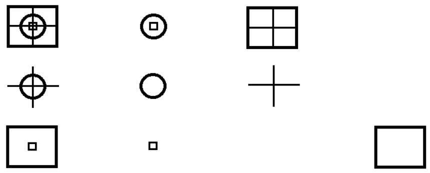

Figure 4: A Raven-like test and its solution

process to a pixel level representation of the pictures [Beltran *et al.*, 2016]. [Ichien *et al.*, 2021] compares similarity-based human performance with deep learning on $A : B :: C : ?$ problems made with images of cars and of parts thereof.

Raven progressive matrices are difficult IQ tests: faced with a $3 \times 3$ matrix with 8 cells containing pictures, one has to guess the right contents of the empty 9th cell, among 8 proposed solutions. An example (where an original Raven test has been replaced by an isomorphic example (in terms of logical encoding)) is given with its solution (a simple big square) in Figure 4. The idea of the AP-based approach is to postulate that in a line (and maybe in a column), the picture of the 3rd cell is to the pictures of the first two pictures as the picture of the 3rd cell is to the pictures of the first two pictures in the next line (or column). It amounts to dealing with proportions of the form $(\vec{cell}_1, \vec{cell}_2) : \vec{cell}_3 :: (\vec{cell}'_1, \vec{cell}'_2) : \vec{cell}'_3$ where the $\vec{cell}_i$'s refer to vectors describing the features. Then the application of the analogical inference amounts to copying patterns observed in other lines or columns, feature by feature. Again the solution is built, and not chosen among candidates via an optimization process as in [Lovett *et al.*, 2010]. See [Beltran *et al.*, 2016] for details and discussions. See also [Ragni and Neubert, 2012] on this topic.

Other logical proportions are involved in the solving of other puzzles. Thus, inverse paralogy is instrumental in the feature-based representation and solving of Bongard problems (see [Prade and Richard, 2016]), which are visual puzzles where we have two sets $A$ and $B$ of relatively simple pictures. All the pictures in set $A$ have a common feature, which is lacking in all the ones in set $B$. The problem is to find this feature. Besides, heterogeneous proportions have been shown to be appropriate for solving puzzles of the type "Finding the odd one out" [Prade and Richard, 2014].

### 3.4 Classification

It has been observed [Barbot *et al.*, 2019b] that taxonomies based on cascade of dichotomous properties induce many APs, which suggests that APs have something to do with classification. Classification is probably one of the most investigated topics of modern AI. In that perspective, [Bayoudh *et al.*, 2007] is clearly a pioneering work, leveraging APs (in the case of items described by nominal features) for the purpose of transferring information to the prediction target. This is the first work to show the effectiveness of APs-based classification. It is based on an *analogical dissimilarity* measure closely related to APs [Miclet and Henri Prade, 2009]. More recently, other works have followed such as [Bounhas *et al.*, 2017a; Essid *et al.*, 2020] covering the case of numerical features.

Still related to machine learning in the large, [Miclet *et al.*, 2008; Couceiro *et al.*, 2017] have shown that APs could also be used for data extension, which is a well known process for task such as image classification or segmentation. The power and theoretical limitations of analogy-based methods have been precisely described in [Couceiro *et al.*, 2017; Couceiro *et al.*, 2018], showing that APs-based inference never fails in case the Boolean function underlying the classification is linear; it has been extended to the case of nominal attributes [Couceiro *et al.*, 2020].

As shown in [Bounhas *et al.*, 2017b; Bounhas *et al.*, 2018], *heterogeneous* logical proportions can also be used for evaluating how much an item is at odd with a class, and are thus a tool of interest in classification.

A main issue for analogy-based classification is that it relies on a cubic algorithm (due to the search of appropriate triplets $(a, b, c)$ for classifying a $d$), which is not satisfactory in a world where a standard training set counts more than 100000 items. Nevertheless, such a cubic procedure can be much improved by i) restricting $c$ to be a neighbor of $d$, and ii) looking for a limited number of pairs $(a, b)$ that are particularly "competent" as shown in [Lieber *et al.*, 2019] (the search for these pairs can be done in a pre-processing step; thus there is no need to walk through the whole domain to get accurate classification). Let us also mention the use of APs for completing missing values [Correa Beltran *et al.*, 2014].

### 3.5 Case-Based Reasoning

Case based reasoning is similarity-based. Given a repertory of experienced cases describing problems $p_i$'s with solutions $s_i$'s, and facing a new problem $p^*$, it amounts to looking for (a) similar problem(s) in the repertory and to adapting its/their solution(s) to this new problem. It already involves APs since one can say that the new solution is to the new problem $p^*$ as the solution $s_i$ is to problem $p_i$ (when $p^*$ is similar to $p_i$). But one cannot apply the AP-based inference pattern here because problems and solutions are not described with the same features. But, one can say that "$(p_a, s_a)$ is to $(p_b, s_b)$ as $(p_c, s_c)$ is to' $(p_d, s_d)$". Thus from 3 cases, one can extrapolate an unknown $s_d$, using the AP-based inference, provided that $p_a : p_b :: p_c : p_d$ holds and that $s_a : s_b :: s_c : x$ is solvable. Note that this extrapolation incorporates the adaptation step of case-based reasoning for producing $s_d$ [Lieber *et al.*, 2018].

### 3.6 Preference Learning

The first implemented approach to preference learning based on AP inference to preference learning is due to Fürnkranz and Hüllermeier [Fahandar and Hüllermeier, 2018], and focuses on learning to rank user preferences based on the evaluation of a loss function. It is based on an inference pattern stating that since $a$ differs from $b$ as $c$ differs from $d$ (and vice-versa), and $b$ is preferred to $a$, $d$ should be preferred to $c$ as well. In [Bounhas *et al.*, 2019], the focus is on predicting preferences (rather than getting a ranking) using the previous inference pattern, or a more sophisticated one that states that if $a : b :: c : d$ and $a' : b' :: c' : d'$ hold and $a$ is preferred to $a'$, $b$ is preferred to $b'$, $c$ is preferred to $c'$, then $d$

should be preferred to $d'$. A comparison of the these different approaches can be found in [Bounhas *et al.*, 2019].

### 3.7 Formal Concept Analysis

With analogical reasoning, categorization is a noticeable activity of the mind. This suggests a possible bridge between APs and formal concept analysis (FCA). FCA relies on a so-called formal context, which is a relation $R$ linking items with (Boolean) properties, from which one can define a formal concept as a pair made of a subset of items $E$ and a subset of properties $I$ that characterize them (where $E \times I$ is maximal in $R$). Formal concepts can be organized into a lattice.

As shown in [Barbot *et al.*, 2019a], APs can be defined between formal concepts (in the sense of formal concept analysis) and algorithms to compute them are proposed, exploiting special subcontexts, called analogical complexes. Moreover, APs between concepts can be related to another form of analogy, called "relational proportion", which involves two universes of discourse, such that "Carlsen is to chess as Mozart is to music", or more compactly "Carlsen is the Mozart of chess", which is not anymore a relation between four items of the same kind, but can be extracted from a formal context. See [Barbot *et al.*, 2019b] for a link between dichotomous trees, FCA and analogical complexes.

## 4 Toward Future Developments

The usages of APs as a first class concept or as a side tool for AI tasks have to be better understood in order to improve their effectiveness. In the next subsections, we provide a list of tracks that are worth exploring.

### 4.1 New Theoretical Concerns

Still many theoretical issues are open. Here are two examples. The central permutation postulate is sometimes debatable (e.g., [Barbot *et al.*, 2019a]). It is legitimate to abandon it and to replace it by the weaker internal reversal property ($a : b :: c : d \rightarrow b : a :: d : c$), which still entails *complete reversal* assuming symmetry. The *minimal* Boolean model of an AP $a : b ::_w c : d$ obeying only reflexivity, symmetry and internal reversal postulates is given by $a : b ::_w c : d = (a \equiv c) \wedge (b \equiv d)$ (which gets rid of the two patterns $(0, 0, 1, 1)$ and $(1, 1, 0, 0)$ in Table 1). Another theoretical issue has its root in an old treatise by French mathematician Gaspard Monge [Monge, 1793]. In this book, relations between strengths and lengths are described in terms of APs and generalized APs such as $a : b : c :: d : e : f$! Such double APs deserves attention in further research (see [Prade and Richard, 2021]).

### 4.2 Argumentation

According to [Bartha, 2009], an analogical argument is an explicit representation of analogical reasoning that cites accepted similarities between two systems in support of the conclusion that some further similarity exists. There are still few works on analogical argumentation [Amgoud, 2020]. Besides, a sequence of analogical arguments may lead to consider APs. Indeed in a debate, a discussant $d$ may state that situation $S_2$ is like situation $S_1$ and that what took place in $S_1$ will happen in $S_2$ as well. The opponent discussant $d'$ may argue that there is an (important) feature where they differ, and that what took place in $S_1$ may not happen in $S_2$. Then $d$ may produce another pair of situations $(S_3, S_4)$, where the same difference can be observed without affecting the conclusion advocated by $d$ for $S_2$. Then $d'$ may counter-argue if he knows another pair of situations $(S'_3, S'_4)$ where the same difference does lead to a different conclusion. Thus this kind of exchange can be analyzed in terms of APs.

### 4.3 Case-Based Decision

Case-based decision [Gilboa and Schmeidler, 1995] is based on similarity only. AP-based inference may be of interest for taking advantage of both similarity and dissimilarity between situations. Let us just illustrate the basic idea using a simple scenario; see [Billingsley *et al.*, 2017] for details and various scenarii. Suppose that a decision $\delta$ was experienced in two different situations $sit_1$ and $sit_2$ in the presence or not of some "special circumstances", leading to "good" or "bad" results respectively depending on the absence or on the presence of these special circumstances. Suppose we have in our repository two cases pertaining to $sit_1$ with and without the presence of special circumstances, and a case pertaining to $sit_2$ with special circumstances, then using AP-based inference, one can conclude if decision $\delta$ in $sit_2$ without the special circumstances may be good or not.

### 4.4 XAI

With the emergence of sophisticated AI/ML techniques, the need for human understandable explanations of the output has become a major research theme known as eXplainable AI (XAI). In [Hüllermeier, 2020], the idea of analogy-based explanations is developed. Because most machine learning methods do obey the regularity assumption underlying similarity-based inference, similarity-based explanations are very natural. Not only may a meaningful measure of similarity require complex individual aggregations, but it can also happen that, for a given item, we cannot find "similar" items among the available data. This is where an analogy-based approach becomes relevant because the notion of similarity at item level is not involved. With an AP $a : b :: c : d$, we do not have to compare $a$ to $b$ or $c$ and $d$: we just have to estimate the relation between $a$ and $b$ which has to be the same as the relation between $c$ and $d$. As such, APs can be used in a model-agnostic way, without taking into account the very details of the underlying ML model. This is work in progress. In [Hug *et al.*, 2019], APs are still used to suggest explanations in recommender systems (see also [Sakaguchi *et al.*, 2011]). Performances of recommender systems, largely based on matrix factorization, are generally made via diverse metrics such as precision, recall, coverage, surprise, etc. Whatever their performances w.r.t. these metrics, they are still considered as "black-boxes", suggesting a movie without explaining why (Netflix for instance). Mining APs in incomplete database-known dataset for recommender systems) allows to find analogies like "Star Wars (1977) is to Raiders of the Lost Ark (1981) as Return of the Jedi (1983) is to Indiana Jones and the Last Crusade (1989)", then to bring an explanation to "Last Crusade (1989)" recommendation.

### 4.5 Transfer Learning

The idea of transfer learning, which may be viewed as a kind of analogical reasoning performed at the meta level, is to take advantage of what has been learnt on a source domain in order to improve the learning process in a target domain. Re-using knowledge across different learning tasks (involving different domains), also known as multi-tasks learning, has long been addressed in the machine learning literature [Caruana, 1998; Daumé III and Marcu, 2006; Dai *et al.*, 2009], generally assuming that the tasks are related at a low level (sharing the same feature space, or the same parametric family of models for instance). From a practical viewpoint, in the context of deep learning, transfer learning is usually implemented by retraining an existing neural network on new data to only update the weights corresponding to the last layers. In [Wang and Yang, 2011] the problem of transfer learning is addressed using structural analogy [Gentner, 1989] between two domains with completely different low-level representations. By making use of sophisticated statistical tools, the authors are able to estimate high level dependencies across domains, then to successfully validate the approach on a large number of transfer learning scenarios.

## 5 Conclusion

Drawing on past experiences to effectively solve current problems is a fundamental insight of human intelligence. Using APs as a formal model linking past pairs (problem, solution) to new observations is a natural way to describe this learning paradigm. As we have seen, APs can be used in a large variety of contexts, from computational linguistics to image processing, via multi-tasks learning. APs are not only a matter of similarity, but also involved dissimilarity. In a prediction perspective, APs bear at least a superficial link with conformal predictors where non-conformity is treated as a first class citizen for ML purposes [Bounhas *et al.*, 2017b].

Ultimately, APs could help to implement a model of human high-level perception leading to the flexible build-up of representations appropriate to a given context, paving the way for creativity in AI.

## References

[Allen and Hospedales, 2019] C. Allen and T. M. Hospedales. Analogies explained: Towards understanding word embeddings. In *Proc. 36th Int. Conf. on Machine Learning (ICML'19)*, pages 223–231, 2019.

[Amgoud, 2020] Leila Amgoud. Evaluation of analogical arguments by Choquet integral. In *Proc. 24th ECAI*, pages 593–600, 2020.

[Aristotle, 2011] Aristotle. *Nicomachean Ethics*. Univ. of Chicago Press, 2011. Trans. by R. Bartlett and S. Collins.

[Barbot *et al.*, 2019a] Nelly Barbot, Laurent Miclet, and Henri Prade. Analogy between concepts. *Artif. Intell.*, 275:487–539, 2019.

[Barbot *et al.*, 2019b] Nelly Barbot, Laurent Miclet, Henri Prade, and Gilles Richard. A new perspective on analogical proportions. In *Proc. 15th ECSQARU*, pages 163–174. LNCS 11726, Springer, 2019.

[Bartha, 2009] Paul F. Bartha. *By Parallel Reasoning: The Construction and Evaluation of Analogical Arguments*. Oxford Univ. Press, 2009.

[Bayoudh *et al.*, 2007] Sabri Bayoudh, Laurent Miclet, and Arnaud Delhay. Learning by analogy: A classification rule for binary and nominal data. In *Proc. 20th IJCAI*, pages 678–683, 2007.

[Bayoudh *et al.*, 2012] Meriam Bayoudh, Henri Prade, and Gilles Richard. Evaluation of analogical proportions through Kolmogorov complexity. *Knowledge-Based Systems*, 29:20–30, 2012.

[Beltran *et al.*, 2016] William Correa Beltran, Henri Prade, and Gilles Richard. Constructive solving of Raven's IQ tests with analogical proportions. *Int. J. Intell. Syst.*, 31(11):1072–1103, 2016.

[Billingsley *et al.*, 2017] Richard Billingsley, Henri Prade, Gilles Richard, and Mary-Ann Williams. Towards analogy-based decision - A proposal. In *Proc. 12th Conf. FQAS*, LNAI 10333, pages 28–35. Springer, 2017.

[Bounhas *et al.*, 2017a] Myriam Bounhas, Henri Prade, and Gilles Richard. Analogy-based classifiers for nominal or numerical data. *Int. J. Approx. Reason.*, 91:36–55, 2017.

[Bounhas *et al.*, 2017b] Myriam Bounhas, Henri Prade, and Gilles Richard. Oddness / evenness-based classifiers for Boolean or numerical data. *Int. J. Approx. Reason.*, 82:81–100, 2017.

[Bounhas *et al.*, 2018] Myriam Bounhas, Henri Prade, and Gilles Richard. Oddness-based classification: A new way of exploiting neighbors. *Int. J. Intell. Syst.*, 33(12):2379–2401, 2018.

[Bounhas *et al.*, 2019] Myriam Bounhas, Marc Pirlot, Henri Prade, and Olivier Sobrie. Comparison of analogy-based methods for predicting preferences. In *LNCS 11940*, pages 339–354. Springer, 2019.

[Bouraoui *et al.*, 2018] Zied Bouraoui, Shoaib Jameel, and Steven Schockaert. Relation induction in word embeddings revisited. In *Proc. 27th COLING*. 1627-1637, Assoc. Computat. Ling., 2018.

[Boy de la Tour and Caferra, 1987] Thierry Boy de la Tour and Ricardo Caferra. Proof analogy in interactive theorem proving: A method to express and use it via second order matching. In *Proc. 6th AAAI*, pages 95–99, 1987.

[Caruana, 1998] Rich Caruana. Multitask learning. In *Learning to Learn*, pages 95–133. (S. Thrun, L. Y. Pratt, eds.), Springer, 1998.

[Cornuéjols, 1996] Antoine Cornuéjols. Analogy as minimization of description length. *Machine Learning and Statistics: The interface*, 1996. (G. Nakhaeizadeh, C. Taylor, eds.), 321-336, Wiley.

[Correa Beltran *et al.*, 2014] William Correa Beltran, Hélène Jaudoin, and Olivier Pivert. Analogical prediction of null values: The numerical attribute case. In *Proc. 18th East Europ. Conf. Advances in Databases and Inf. Syst. (ADBIS'14)*, volume 8716 of *LNCS*, pages 323–336. Springer, 2014.

[Couceiro *et al.*, 2017] Miguel Couceiro, Nicolas Hug, Henri Prade, and Gilles Richard. Analogy-preserving functions: A way to extend Boolean samples. In *Proc. 26th IJCAI, Melbourne*, pages 1575–1581, 2017.

[Couceiro *et al.*, 2018] Miguel Couceiro, Nicolas Hug, Henri Prade, and Gilles Richard. Behavior of analogical inference w.r.t. Boolean functions. In *Proc. 27th IJCAI, Stockholm*, pages 2057–2063, 2018.

[Couceiro *et al.*, 2020] Miguel Couceiro, Laurent Miclet Erkko Lehtonen, Henri Prade, and Gilles Richard. When nominal analogical proportions do not fail. *Proc. 14th Int. Conf. SUM*. LNCS, 12322, pages 68–83. Springer, 2020.

[Dai *et al.*, 2009] Wenyuan Dai, Yuqiang Chen, Gui-rong Xue, Qiang Yang, and Yong Yu. Translated learning:transfer learning across different feature spaces. In *Adv.in NIPS*. Curran Ass. Inc., 2009.

[Daumé III and Marcu, 2006] Hal Daumé III and Daniel Marcu. Domain adaptation for statistical classifiers. *JAIR*, 26(1):101–126, 2006.

[Diallo *et al.*, 2019] Aïssatou Diallo, Markus Zopf, and Johannes Fürnkranz. Learning analogy-preserving sentence embeddings for answer selection. In *Proc. 23rd CoNLL*, pages 910–919. ACL., 2019.

[Drozd *et al.*, 2016] Aleksandr Drozd, Anna Gladkova, and Satoshi Matsuoka. Word embeddings, analogies, and machine learning: Beyond king - man + woman = queen. In *Proc. COLING*, pages 3519–3530, 2016.

[Dubois *et al.*, 2016] Didier Dubois, Henri Prade, and Gilles Richard. Multiple-valued extensions of analogical proportions. *Fuzzy Sets and Systems*, 292:193–202, 2016.

[Essid *et al.*, 2020] Marouane Essid, M yriam Bounhas, and Henri Prade. Continuous analogical proportions-based classifier. In M.-J. Lesot et al., editor, *Proc. 18th Int. Conf.*, volume CCIS 1237, pages 541–555. Springer, 2020.

[Evans, 1968] Thomas 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.

[Fahandar and Hüllermeier, 2018] Mohsen A. Fahandar and Eyke Hüllermeier. Learning to rank based on analogical reasoning. In *Proc. 32nd AAAI Conf. on AI*, pages 2951–2958. AAAI Press, 2018.

[Gentner *et al.*, 2001] Dedre Gentner, Keith J. Holyoak, and Boicho N. Kokinov, editors. *The Analogical Mind: Perspectives from Cognitive Science*. MIT Press, 2001.

[Gentner, 1989] Dedre Gentner. The mechanisms of analogical learning. In *Similarity and Analogical Reasoning*, pages 197–241. Cambridge Univ. Press, New York, 1989.

[Gilboa and Schmeidler, 1995] Itzhak Gilboa and David Schmeidler. Case-based decision theory. *Quart. J. of Econ.*, 110:605–639, 1995.

[Gust *et al.*, 2006] Helmar Gust, Kai-Uwe Kühnberger, and Ute Schmid. Metaphors and heuristic-driven theory projection (HDTP). *Theoretical Computer Science*, 354(1):98 – 117, 2006.

[Hertzmann *et al.*, 2001] Aaron Hertzmann, Charles E. Jacobs, Nuria Oliver, Brian Curless, and David H. Salesin. Image analogies. *Proc. ACM SIGGRAPH*, 2001.

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

[Hesse, 1966] Mary B. Hesse. *Models and Analogies in Science*. 1st ed. Sheed & Ward, London, 1963; 2nd augmented ed. University of Notre Dame Press, 1966.

[Hofstadter and Mitchell, 1995] Douglas Hofstadter and Melanie Mitchell. The Copycat project: A model of mental fluidity and analogy-making. In *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought*, pages 205–267. Basic Books, Inc., 1995.

[Hofstadter and Sander, 2013] Douglas Hofstadter and Emmanuel Sander. *Surfaces and Essences: Analogy as the Fuel and Fire of Thinking*. Basic Books, 2013.

[Hug *et al.*, 2019] Nicolas Hug, Henri Prade, Gilles Richard, and Mathieu Serrurier. Analogical proportion-based methods for recommendation – First investigations. *Fuzzy Sets and Syst.*, 366:110–132, 2019.

[Hüllermeier, 2020] Eyke Hüllermeier. Towards analogy-based explanations in machine learning. In *Proc. 17th Int. Conf. on Modeling Decisions for AI (MDAI)*, pages 205–217. LNCS 12256, Springer, 2020.

[Hwang *et al.*, 2013] Sung Ju Hwang, Kristen Grauman, and Fei Sha.Analogy-preserving semantic embedding for visual object categorization*30thICML*, pages1676–1684,2013.

[Ichien *et al.*, 2021] Nicholas Ichien, Qing Liu, Shuhao Fu, Keith J. Holyoak, Alan Yuille, and Hongjing Lu. Visual analogy: Deep learning versus compositional models. *Proc. 43rd Annual Meeting Cognit. Sci. Soc.*, 2021.

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

[Law *et al.*, 2017] Marc T. Law, Nicolas Thome, and Matthieu Cord. Learning a distance metric from relative comparisons between quadruplets of images. *Int. J. Comput. Vis.*, 121(1):65–94, 2017.

[Lepage and Denoual, 2005] Yves Lepage and Etienne Denoual. Purest ever example-based machine translation: Detailed presentation and assessment. *Machine Translation*, 19(3-4):251–282, 2005.

[Lepage, 1998] Yves Lepage. Solving analogies on words: An algorithm. In *Proc. 36th COLING,* Montréal, pages 728–735, 1998.

[Lepage, 2001] Yves Lepage. Analogy and formal languages. *Electr. Notes Theor. Comput. Sci.*, 53, 2001.

[Lepage, 2003] Yves Lepage. De l'analogie rendant compte de la commuta tion en linguistique. HDR, U. J. Fourier, Grenoble, 2003.

[Levy and Goldberg, 2014] Omer Levy and Yoav Goldberg. Dependency-based word embeddings. In *Proc. EMNLP*, pages 302–308, 2014.

[Lieber *et al.*, 2018] Jean Lieber, Emmanuel Nauer, Henri Prade, and Gilles Richard. Making the best of cases by approximation, interpolation and extrapolation. In *Proc. 26th ICCBR*, LNCS 11156, pages 580–596. Springer, 2018.

[Lieber *et al.*, 2019] Jean Lieber, Emmanuel Nauer, and Henri Prade. Improving analogical extrapolation using case pair competence. *Proc 27th ICCBR*. Springer, LNCS 11680, pages 251–265, 2019.

[Lim *et al.*, 2021] Suryani Lim, Henri Prade, and Gilles Richard. Classifying and completing word analogies by machine learning. *Int. J. Appr. Reas.*, 132:1–25, 2021.

[Linzen, 2016] Tal Linzen. Issues in evaluating semantic spaces using word analogies. *CoRR*, abs/1606.07736, 2016.

[Lovett *et al.*, 2010] Andrew Lovett, Kenneth Forbus, and Jeffrey Usher. A structure-mapping model of Raven's progressive matrices. In *Proc. 32nd Conf. of the Cognitive Science Soc., Portland, OR*, 2010.

[Lu *et al.*, 2019] Hongjing Lu, Ying Nian Wu, and Keith J. Holyoak. Emergence of analogy from relation learning. *Proc. National Acad. of Sciences*, 116:4176–4181, 2019.

[Melis and Veloso, 1998] Erica Melis and Manuela Veloso. Analogy in problem solving. In *Handbook of Practical Reasoning: Computational and Theoretical Aspects*. Oxford Univ. Press, 1998.

[Miclet and Henri Prade, 2009] Laurent Miclet and Henri Prade. Handling analogical proportions in classical logic and fuzzy logic settings. In *Proc. 10th ECSQARU*, LNCS 5590, pages 638–650. Springer, 2009.

[Miclet *et al.*, 2008] Laurent Miclet, Sabri Bayoudh, and Arnaud Delhay. Analogical dissimilarity: Definition, algorithms and two experiments in machine learning. *JAIR*, 32:793–824, 2008.

[Miclet *et al.*, 2014] Laurent Miclet, Nelly Barbot, and Baptiste Jeudy. Analogical proportions in a lattice of sets of alignments built on the common subwords in a finite language. In *Computational Approaches to Analogical Reasoning: Current Trends*, pages 245–260. Springer, 2014.

[Mikolov *et al.*, 2013] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. In *Advances in NIPS*, pages 3111–3119. Curran Associates Inc., 2013.

[Monge, 1793] Gaspard Monge. *Traité Elémentaire de Statique, A l'Usage des Ecoles de la Marine, 3rd, ed.* Baudouin, Paris, 1793. 8 eds. in French to 1846. English transl. An Elementary Treatise on Statics, Baker, Woods, 1851.

[Murena *et al.*, 2020] Pierre-Alexandre Murena, Marie Al-Ghossein, Jean-Louis Dessalles, and Antoine Cornuéjols. Solving analogies on words based on minimal complexity transformation. In *Proc. 29th IJCAI*, pages 1848–1854, 2020.

[Pennington *et al.*, 2014] Jeffrey Pennington, Richard Socher, and Christopher Manning. GloVe:Global vectors for word representation In *EMNLP*, pages 1532–1543, 2014.

[Piaget, 1953] Jean Piaget. *Logic and Psychology*. Manchester Univ. Press, 1953.

[Pirrelli and Yvon, 1999] Vito Pirrelli and François Yvon. Analogy in the lexicon: a probe into analogy-based machine learning of language. In *Proc. 6th Int. Symp. on Human Communic., Cuba*, 1999.

[Polya, 1945] George Polya. *How to Solve It*. Princeton Univ. Press, 1945.

[Prade and Richard, 2013] Henri Prade and Gilles Richard. From analogical proportion to logical proportions. *Logica Univers.*, 7:441–505, 2013.

[Prade and Richard, 2014] Henri Prade and Gilles Richard. Homogenous and heterogeneous logical proportions. *If-CoLog J. of Logics and their Applications*, 1(1):1–51, 2014.

[Prade and Richard, 2016] Henri Prade and Gilles Richard. On different ways to be (dis)similar to elements in a set. Boolean analysis and graded extension. In *Proc. 16th Int. Conf., IPMU*, volume 611 of *CCIS*, pages 605–618. Springer, 2016.

[Prade and Richard, 2018] Henri Prade and Gilles Richard. Analogical proportions: from equality to inequality.*Int. J. Appr. Reas.*,101:234–254, 2018.

[Prade and Richard, 2021] Henri Prade and Gilles Richard. Multiple analogical proportions. *AI Com*, to appear, 2021.

[Ragni and Neubert, 2012] Marco Ragni and Stefanie Neubert. Solving Raven's IQ-tests: An AI and cognitive modeling approach. In *Proc. 20th ECAI, Montpellier, Aug. 27-31*, pages 666–671, 2012.

[Rumelhart and Abrahamson, 1973] David E. Rumelhart and Adele A. Abrahamson. A model for analogical reasoning. *Cognitive Psychol.*, 5:1–28, 1973.

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

[Sakaguchi *et al.*, 2011] Takatoshi Sakaguchi, Yuya Akaho, Kazuhiro Okada, Torahiko Date, Tomohiro Takagi, Naoki Kamimaeda, Masanori Miyahara, and Tomohiro Tsunoda. Recommendation system with multi-dimensional and parallel-case four-term analogy. *IEEE Int. Conf. on Systems, Man, and Cybern.*, pages 3137–3143, 2011.

[Veloso and Carbonell, 1993] Manuela M. Veloso and Jaime G. Carbonell. Derivational analogy in PRODIGY: Automating case acquisition, storage, and utilization. *Mach. Learn.*, 10:249–278, 1993.

[Wang and Yang, 2011] Hua-Yan Wang and Qiang Yang. Transfer learning by structural analogy. In *Proc. 25 AAAI Conf. on AI*, pages 513–518. AAAI Press, 2011.

[Yvon and Stroppa, 2006] François Yvon and Nicolas Stroppa. Formal models of analogical proportions. Technical Report D008, ENST, Paris, 2006.

[Zhu and de Melo, 2020] Xunjie Zhu and Gerard de Melo. Sentence analogies: Linguistic regularities in sentence embeddings. In *Proc. COLING*, pages 3389–3400, 2020.