Analogy
as
Minimization of Description Length

Antoine CORNUEJOLS
Inference and Learning Group
Laboratoire de Recherche en Informatique (L.R.I.)
Building 490 University of Paris-Sud
91405 ORSAY Cédex
email: antoine@lri.fr phone: (33 1) 69 41 63 00

February 22, 2021

Abstract

If analogy is indeed a form of induction, it is a very special one in that (i) it works from the specific to the specific apparently without relying on vast numbers of observations, d (ii) it does not necessarily involve general rules that would apply both to the 'source' of the analogy and to its 'target'. It thus appears quite remote from the realms of statistics. A survey of the current computational approaches to analogy shows that they all embrace the "analogy as a matching proces" perspective differing only on constraining factors and search mechanisms. These matching-based theories gives ways to some dissatisfaction however. This paper proposes a new point of view on analogy that offers possible answers to these queries. Analogy is seen as a realization of an economy principle that minimizes the complexity of the transformation from the source to the target as measured by description length. These description lengths in turn are dependent upon statistical properties of the concepts and abstractions used to account for the analogy. This is where learning can take place so as to facilitate further analogies in the same domain. Tests on a classical domain task confirm that the application of this principle correctly predicts the best analogies.

Content Areas: Analogy, Inductive Learning, Minimum Description Length Principle.

1 Introduction: analogy, induction and computational models

1.1 Analogy and induction

Peculiarities of analogy that make it different from other, traditionally considered as more typical forms of induction, are that (i) it works from the specific to the specific apparently without relying on vast numbers of observations, and (ii) it does not necessarily involve general rules that would apply both to the 'source' of the analogy and to its 'target'. Models of analogy must threrefore propose mechanisms allowing for specific to specific inferencing and this possibly between different domains. Analogy thus appears quite remote from the realms of statistics.

1.2 Current models of analogical thought

A survey of existing computational studies shows that for solving the specific to specific inference problem, the hypothesis of analogy as a mapping process between the representations of the analogues has been unanimously adopted following [**?**]. Recognizing then that a combinatorial number of possible mappings between graph-representations exists, the problem have become one of taming this complexity to allow focussing on relevant and most promising ones. To this effect, a number of intuitively appealing hypotheses have been made, that we have no place to describe here (see [**?**, **?**, **?**, **?**, **?**, **?**, **?**, **?**, **?**]). However these models still leave place for dissatisfaction.

First, there is no formal ground for chosing the analogy as mapping paradigm. Second, all models mentioned so far are highly dependent on the a priori designed representation scheme (see [**?**, **?**] for critics and illustrations). The latter is a serious drawback, even more so that at the same time it seems quite undisputable that analogy is not only mapping between representations but also and foremost the making of these representations appropriate for a given context through perception. Finally, it has not been proved that existing models can account for observed properties of analogy. It seems that a well-founded theory of analogy must explain these properties.

1.3 Properties of analogy

We list here six properties of analogy that we feel are telltale symptoms and accordingly important test beds for any would-be theory of analogy.

1. Analogy is *pervasive* in cognition.

2. *Abstract* analogies are generally preferred over more litteral ones.

3. There is generally a good *agreement in the ratings of analogies* among different people.

4. Analogy is *non involutive* ($f \circ f \neq Id$). In other words analogy is non symmetric. One can readily find examples where transferring properties from the source to a target and then back from the initial target to the initial source, brings discrepencies between the initial properties of the source and the transferred back ones.

5. Analogy is *non idempotent* ($f \circ f \neq f$). If A is used to infer analogically properties about B, and then B is used to augment the knowledge about C, the result may differ from the one obtained by directly using A analogically on C. It must be noticed here that properties 4 and 5 call into question the feasibility of Case Based Reasoning where the case library is supposed to grow incrementaly with the system's history. Indeed it follows from 4 and 5 that inconsistent case bases might be created.

6. Analogy making *can be improved with practice*. At least some domain-dependent forms of analogical reasoning can be learned.

This is in part while trying to account for these properties that the point of view on analogy reported in this paper has gradually emerged. The overall spirit of this view is presented in the next section. Then, in section **??**, its formal foundations are reviewed and explained. Section **??** illustrates the application of this approach to examples drawn from a classical analogy domain. Finally, the conclusion section sums up the current findings and underlines the deep relationships between inferencing and statistics.

2 Analogy as an economy principle

2.1 The classical schema of analogy

We limit ourselves here to one component of analogical reasoning: the transfer of some properties of the source to the target, ignoring the problem of conjuring up in mind a good analogue as well as the difficulties related to the validation and tuning of the result in the target domain. We will assume for the time being that one can legitimately isolate such components.

Analogy making is, in part, *perceiving* some aspects of the structures of two situations _the essences of those situations, in some sense_ as identical. By perception, we mean here the process by which some conceptual primitives and structures are summoned up and used in order to make sense of the raw data and to represent what is judged relevant in them. When perception is tuned to analogy making, the conceptual structures that come up amount to the filtering of the saillant part of the data that make them analogous.

In a way, one might consider the essences of the analogues as the necessary information that, when provided, allows one to easily transform part of one situation into part of the other, these parts being the relevant ones in the context of analogy. But how to bring these essential aspects out ?

Suppose we try to transform, by way of symbolic manipulations, one analogue, the source, belonging to universe U_1, into the other, the target, belonging to universe U_2. We know that if we find an economic transformation means, there are reasons to believe that we hit on an appropriate information supply, that is some essential commonalties between the analogues.

In the classical schema (see figure ref fig-ana-classic) where ($S_1 \Longrightarrow R_1$) is the *source* and S_2 is the *target* situation or problem with R_2 being the unknown, analogy making is trying to find ways of representing S_1, R_1, and S_2 so that (i) an economic transformation α using these representations exist between S_1 and S_2, and (ii) β_1 the relevant dependency $S_1 \Longrightarrow R_1$ be represented.

It is then a simple matter to translate β_1 into β_2 by way of the mapping $\beta_2 = \alpha(\beta_1)$

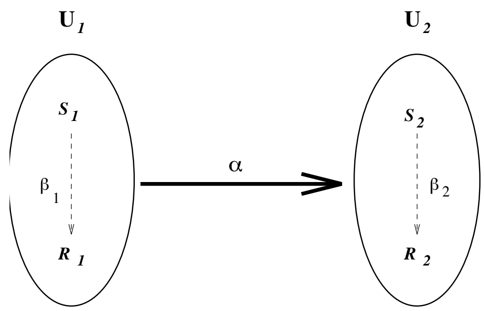

Figure 1: The classical schema of analogy. The *source* in universe U_1 is translated into the target in universe U_2 by way of the transformation α. The simplest the transformation α, the closest look the source and the target.

2.2 Analogy and Kolmogorov complexity

In its naive version, induction consists in drawing general rules from specific observations. These general rules or hypotheses are subject to three constraints: (i) to be more general than the original observations, otherwise they have little value; (ii) to be testable so that later observations may falsify them; and (iii) to inspire confidence, that is to be most probable given the observations.

A general principle that dates back at least to William of Ockham (1290?-1349?) seems to satisfy these three criteria to the best. This is the Principle of Simplicity which asserts that the "simplest" hypothesis or explanation is the most reliable . During the last thirty years, "simplicity" has been equated with "shortest effective description", so that if there are alternative explanations for a given body of observations, one should select the one with the shortest description. In the first formalization ([**?**, **?**, **?**]) called algorithmic or Kolmogorov complexity, hypotheses were considered as the effective computable procedures able to generate as output the observations. The best hypothesis was thus the shortest procedure as measured with the Turing machine encoding scheme, that is the one that allowed maximum compressibility of the data.

We are going to use the same principle here. The conceptual primitives and structures used in the perception and representation of the analogues willstand for the hypotheses, and the raw data will stand for the observations. This way, we will be able to compare the quality of various analogies between some source and target by measuring the corresponding Kolmogorov complexities. The analogies with the lowest algorithmic complexities will be deemed the best.

As has been noted in the introduction, analogy is a special form of induction in which the data consist in a single compound object: the source and its target. It is therefore a great advantage for our purpose, and one that distinguishes it from statistical approaches, that algorithmic complexity applies equally well to sets of observations and to single objects. However, compressibility through Turing machines is not a method that is to be used lightheartedly in Artificial Intelligence. Indeed, one usually looks for explanations expressed within some model classes, and not explanations expressed as abstruse bit strings. This is where the Minimum Description Length Principle (M.D.L.P.) ([**?**, **?**]) intervenes.

2.3 The Minimum Description Length Principle (MDLP)

Rissanen starts from the observation that scientific theories often involve two steps. First, the formulation of a set of possible alternative hypotheses (for which he does not offer any mechanism), and, second, the selection of one hypothesis as the most likely one. Rissanen proposes that this selection mechanism obeys the Minimum Description Length Principle which states that:

> The best theory to explain a set of data is the one which minimizes the sum of
>
> - the length, in bits, of the description of the theory; and,
> - the length, in bits, of data when encoded with the help of the theory.

There is a deep relationship between the MDLP and the Bayesian approach. Indeed, one can derive the former from the latter by observing that from Bayes' Rule:

$$P(H \mid D) = \frac{P(D \mid H) \times P(H)}{P(D)}$$

where H is an hypothesis, and D is the set of observed data, and where we are looking for the hypothesis H that maximizes $P(H \mid D)$ it follows :

$$-\log P(H \mid D) = -\log P(D \mid H) - \log P(H) + \log P(D)$$

Minimizing this expression is then equivalent to minimizing:

$$-\log P(D|H) - \log P(H)$$

since the length of D is fixed for any H. This yields the MDLP.

2.4 Practical problems: encoding scheme and intractability

The problems when trying to apply the MDLP are twofold. First, the length of an explanation will depend on the languages or codes used for describing both the theory and the data. Second, Kolmogorov has proved that searching for the shortest description of an object is NP-complete.

The first problem is known as the *encoding problem.* An answer to it is that, following the equivalence just mentioned between the MDLP and BayesŔule, the code used should reflect, when possible, our prior expectations about the environment: that is descriptions of common or important concepts should be shorter than descriptions of unusual or unimportant ones. We will therefore require that the coding schemes be efficient, i.e. that they provide optimal encodings of the theories with respect to their a priori probability of occurrence, and of the data with respect to each theory. This means that what constitutes a good theory will always be dependent on our expectations about the world. If this seems disappointing, it is exactly these expectations that make the induction problem tractable.

To the second problem, the absence of any effective way of calculating the best theory, there is no other answer than be content with searching for satisfying theories only.

3 Analogy and M.D.L.P.

3.1 Analogy as an economical perception of the source and target

Looking at an analogy problem, we can consider the data as the descriptions of, and the theories as the conceptual constructions that allows to represent S_1, S_2, R_1 and R_2 as well as the transformations and β_1. Now what makes α_S and β_1 interdependent is that they are built using as many common parts or abstractions as possible. This allows the economical description of the analogy. Because there does not always exist a common generalization of the analogues (they might belong to different domains), it might be necessary to introduce different sets of abstractions or models, M_1 and M_2. These models correspond to the ontologies or primitive descriptions that allow the "perception" of the relevant features of the initial data S_1, S_2 and R_1 (see figure **??**).

We propose the following formalization of the MDLP when applied to analogy. M_1 is considered as the theory, and we look for efficient ways of encoding in this theory M_2, α and β_1. The overall complexity of the analogy is then measured using the formula:

$$L(Analogy) = L(M_1) + L(\beta_1 \mid M_1) + L(\alpha_M) + L(\alpha_S \mid \alpha_M) \tag{1}$$

It must be noted that $\alpha_M = I(M_1 \mid M_2)$ the information required to generate M_2, when knowing M_1 , and that α and β_1 are similarly defined.

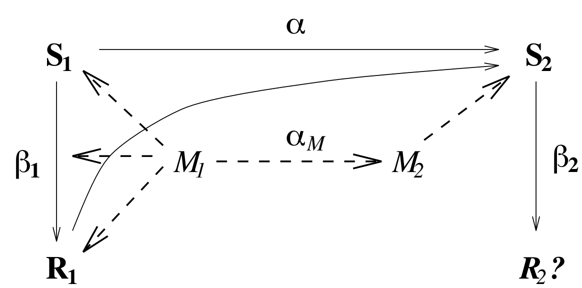

Figure 2: The interplay between conceptual structures M_1 and M_2 and the representation of S_1, S_2 and R_1 as well as α_S and α_M.

3.2 Overview of the task domain and the experiments

As a task domain, we have chosen the microworld developed by Hofstadter et al. for the COPYCAT project [**?**]. In this domain, the analogy problems consists in finding how a letter string is transformed given, as an example, another string and its transform. For instance, given that **abc** $\Longrightarrow$ **abd**, what becomes of **iijjll** $\Longrightarrow$ **?**.

This microworld offers several useful features. It is simple in its definition and it is reasonably straightforward to find adequate knowledge representation primitives for it. It nonetheless provides much of the richness of the analogy problem. And, finally, it is easy to rate the relative merits of the possible solutions for the test riddles.

As a first step towards a complete account of analogy as a Minimum Description Length problem, a preliminary check for the feasibility of the project is to build an encoding scheme for the abstractions, rules and concepts than can enter analogy making, and then see if the ensuing description lengths of possible solutions for given test problems reflect the expectations about their quality. For instance, it is expected that, to the afore-mentioned problem, **iijjll** is a better solution than **iijjld**. Is the description length associated with the first solution accordingly shorter than the description length for the second one ?

3.3 Proposal for an encoding scheme

In an analogy problem, we can consider that the data are the descriptions of the source and the target, and the theories are the conceptual constructions that allows to express the transformation of the source into the target and the relevant properties of the source that need to be transferred (see figure **??** in section **??**). We now have to describe the primitives that allow to represent both the data and the theories. At the same time we have to assign them a coding length that obeys as much as possible the efficiency criteria underlined in section xxx.

The knowledge representation primitives are mostly the ones used in [**?**]. They include: the 26 *letters* of the alphabet, *numbers*, concepts of *relative positions* such as *leftmost*, *rightmost* and *middle*, types of objects called *unit-base* such as *letters* and *groups* of letters, *directions* for reading the strings: *left* and *right*, primitives for *successor laws* like *succ(i,x)* meaning taking the ith successor of unit-base x, and *pred(i,x)* for the corresponding ith predecessor of unit-base x.

Using these primitives, we define templates for the descriptions of the letter strings and for the descriptions of the transformations between letter strings. These templates constitute the theories that are to be compared with respect to the simplicity of the description they will allow. The templates or conceptual primitives that yield the shortest description of the analogy will be selected as th best ones. By using them, we single out the most relevant aspects of the analogy.

3.3.1 Description for the letter strings

The letter strings (e.g. **abc** or **iijjkk**) can be described using a template formulated as a grammar. For each string we may specify the following characteristics: a *read-direction*, the *unit-base*, the *successor-law*, the *length*, and the *starting unit-base*. Each of these "attributes" can possibly be itself described recursively using other attributes. For instance,

iijjkk	=	a *read-direction* (right)
		unit-base (*group* of same *unit-base* (letters) of *length* (2))
		successor-law (letter $\mapsto$ succ(1,letter)) of *length* (3)
		starting-with (*unit-base* of (letters = 'I'))

This description corresponds to the perception of the string **iijjkk** as the grouping of three successive groups of 2 identical letters. Of course, other perceptions are possible, as for instance a very litteral and myopic one:

iijjkk = concat(i,i,j,j,k,k) which does not convey any structural information.

3.3.2 Description for the transformations

A transformation between letter strings or concepts or abstractions must specify what has to be changed in the description so as to obtain the transformed object from the original one. This corresponds to the conditional information needed, in addition to the information contained in the original object, to get the new one. The best induction is obtained when this conditional information is maximally compressed. Here the application of the MDLP results in the choice of certain abstractions and constructions thereof.

For instance, following the works of Hofstadter et al., we will restrain the description of the dependence relation β to the mold *"Replace _ by _"* where the _ can stand for diverse concepts of various levels of abstraction e.g. *letter*, *unit-base* (...), "A", succ(i,x).

The resemblance relation α on the other hand will state the correspondence for each modified attribute of a description. For instance, between the descriptions of *abc* and *ijk*, the α relation may be stated as:

$$\alpha \quad = \quad \textit{starting-with}\ (\textit{unit-base} = \text{“A”}) \\ \mapsto \textit{starting-with}\ (\textit{unit-base} = \text{“I”})$$

It could as well be stated as:

α = “A” $\mapsto$ “I” ; “B” $\mapsto$ “J” ; “C” $\mapsto$ “K”

Figure **??** gives an example of an analogy using these description schemes.

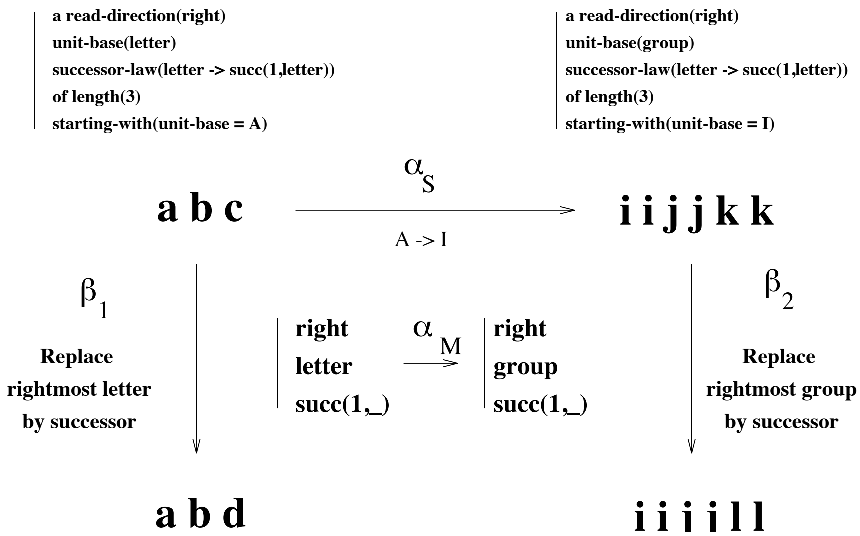

Figure 3: A simple analogy described using a certain set of primitives.

3.4 The encoding scheme: yardsticks for description lengths

In the absence of prior probabilities on the various concepts and compound descriptions, it is natural to resort to measures of the relative specificity of the concepts: the more general ones being simpler to specify than less general ones need correspondingly shorter descriptions. It is customary to organize concepts in hierarchies. In the letter microworld for instance, “group of letters” is more general than “group of 1 letter”, which is itself more general than, say, the letter ‘A’. We get therefore the hierarchies of figure **??**.

Each node of these hierarchies is assigned a probability within this hierarchy according to its specificity. Now, following well-known rules for efficient encoding, the optimal coding length for a concept of probability P should be $-\log(P)$. This is what will be taken here. For instance the description length of “succ(3,_)” is then: $-\log(1/16) = 4$ bits.

The description length of compound abstractions will be taken as the sum of the description lengths of their components. Hence:

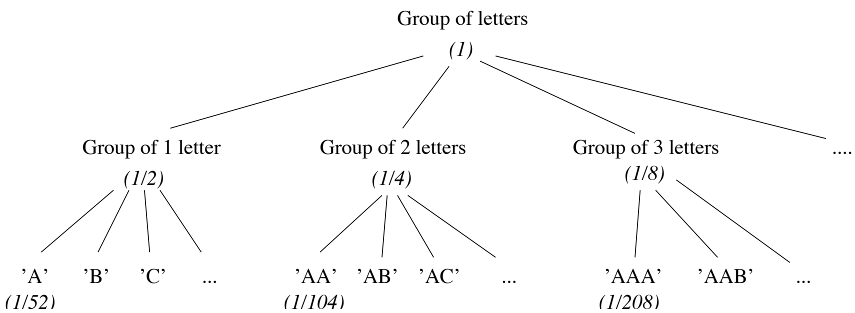

Figure 4: Exemple of hierarchies of concepts with their attached prior probabilites

$$\begin{array}{rcl} \text{L(letter} \mapsto \text{succ(2,letter))} & = & -2\log P(\text{group of 1 letter}) \text{ -log } P(\text{succ(2,\_)}) \\ & = & 1 + 3 = 4 \text{ bits} \end{array}$$

3.5 Illustration

We illustrate this with the following example: **abc** $\Longrightarrow$ **abd** ; **iijjkk** $\Longrightarrow$ **?**

$$\begin{array}{rcl} S_1 & = & \text{a } \textit{read-direction}(\text{right}) \\ & & \textit{unit-base}(\text{letter}) \\ & & \textit{successor-law}(\text{letter} \mapsto \text{succ(1,letter)}) \\ & & \text{of } \textit{length}(3) \\ & & \textit{starting-with}(\text{letter = “A”}) \end{array}$$

$$\begin{array}{rcl} S_2 & = & \text{a } \textit{read-direction}(\text{right}) \\ & & \textit{unit-base}(\textit{group} \text{ of same } \textit{letters} \text{ of } \textit{length}(2)) \\ & & \textit{successor-law}(\text{letter} \mapsto \text{succ(1,letter)}) \\ & & \text{of } \textit{length(3)} \\ & & \textit{starting-with } (\text{letter = “I”}) \end{array}$$

$$\begin{array}{rcl} M_1 & = & \text{a } \textit{read-direction}(\text{right}) \\ & & \textit{unit-base}(\text{letter}) \\ & & \textit{successor-law}(\text{letter} \mapsto \text{succ(1,letter)}) \\ & & \text{of } \textit{length(3)} \end{array}$$

$$\begin{array}{rcl} M_2 & = & \text{a } \textit{read-direction}(\text{right}) \\ & & \textit{unit-base}(\textit{group} \text{ of same } \textit{letters} \text{ of } \textit{length}(2)) \\ & & \textit{successor-law}(\text{letter} \mapsto \text{succ(1,letter)}) \\ & & \text{of } \textit{length(3)} \end{array}$$

Therefore:

$$\begin{array}{lcl} \alpha_M & \equiv & \textit{unit-base}(\text{letter}) \mapsto \textit{unit-base}(\textit{group} \text{ of same } \textit{letters} \text{ of } \textit{length}(2)) \\ (\alpha_S \mid \alpha_M) & \equiv & \textit{starting-with}(\text{letter} = \text{``A''}) \mapsto \textit{starting-with}\ (\text{letter} = \text{``I''}) \\ \beta_1 & \equiv & \text{``Replace the } \textit{right}\text{most } \textit{letter} \text{ by its } \textit{successor}\text{''} \end{array}$$

We then get using an extended version of the encoding scheme described above:

$$\begin{array}{lclcl} L(M_1) & = & -\log(1/2) - \log(1/2) + 4 + 2 & = & 8.00 \text{ bits} \\ L(\beta_1 \mid M_1) & = & -\log(1/3) & = & 1.58 \\ L(\alpha_M) & = & -\log(26/262) & = & 4.70 \\ L(\alpha_S \mid \alpha_M) & = & -\log(1/52) & = & 5.70 \end{array}$$

So that the overall complexity of this analogy is: 20 bits.

4 Experimental results

In the work partially reported here, we examined three questions. First, does the *analogy = economical perception* perspective seem to have some validity? Second, what exactly is transferred in an analogy? And last, what is the optimal level of abstraction, in the context of an analogy, for the intermediary concepts M_1 and M_2?

To this aim, we have set to manually translate miscellaneous analogy problems in the letter microworld with various competing solutions, and then to compute their corresponding complexity. We compared then the results obtained both with intuitive ratings of the diverse analogies and with attached frequency results measured with a set of subject and reported in [**?**] (multiple answers were possible). We made also a comparison with Copycatś behavior so as to test if the fact that we share with this system a lot of the same representation primitives induces a bias towards similar results. The expectation, if, our model is correct is that less complex analogies should correspond to the preferred ones. At the same time, we inspected the abstractions used in the best analogies with regard to the second and third questions above. For lack of space, only sketchy details of some experiments are given thereunder, more can be found in [**?**].

Problem 1: **abc $\Longrightarrow$ abd ; iijjkk $\Longrightarrow$?**

Solution 1: "Replace rightmost group of letters by its successor" iijjkk $\Longrightarrow$ iijjll
Solution 2: "Replace rightmost letter by its successor" iijjkk $\Longrightarrow$ iijjkl
Solution 3: "Replace rightmost letter by D" iijjkk $\Longrightarrow$ iijjkd
Solution 4: "Replace third letter by its successor" iijjkk $\Longrightarrow$ ijkkk
Solution 5: "Replace Cs by Ds" iijjkk $\Longrightarrow$ iijjkk

Problem 2: **abc $\Longrightarrow$ abd ; srqp $\Longrightarrow$?**

Solution 1: "Replace rightmost letter by its predecessor" srqp $\Longrightarrow$ srqo
Solution 2: "Replace leftmost letter by its successor" srqp $\Longrightarrow$ trqp

Problem 3: **abc $\Longrightarrow$ abd ; xcg $\Longrightarrow$?**

Solution 1: "Replace rightmost letter by its successor" xcg $\Longrightarrow$ xch
Solution 2: "Replace Cs by Ds" xcg *Longrightarrow* xdg
Solution 3: "Replace rightmost letter by D" xcg *Longrightarrow* xcd

	P1;S1	P1;S2	P1;S3	P1;S4	P1;S5
Complexity	20 bits	23.3	23.4	22.7	29.8
Hum. subj.	1/26	26/26	2/26	N/A	N/A
COPYCAT	81%	16.5%	0.3%	0%	0%
My rating	1	2	3	4	5

	P2;S1	P2;S2	P3;S1	P3;S2	P3;S3
Complexity	16.7	16.7	22.1	25.8	24.8
Hum. subj.	10/34	7/34	43/49	6/49	4/49
COPYCAT	56%	18.6%	97.4%	1.4%	1.2%
My rating	1	1	1	2	3

The figures for the complexity measure were obtained using formula **??** of section **??** and a fully developed version of the encoding scheme briefly presented in sections **??** and **??**. Because this encoding scheme leaves place for arbitrariness in some places, the absolute values of the complexity numbers should be taken with a grain of salt. Their relative values however are more interesting. They can indeed be interpreted as relative probabilities of occurrence. Thus, $Prob(P1;S2)/Prob(P1;S1) = 2^{23.3-20} = 2^{3.2} = 10.24$. The observation of the overall table shows that if general trends of the complexity measure are in accord with experimental evidences with human subjects and COPYCAT, the comparison at this stage is yet not conclusive.

5 Conclusion

This paper has presented a new perspective where analogy is seen as the result of an economical perception of the analogues. A formal account for this was given implying a form of the Minimum Description Length Principle. Experimental tests on toys problems reveal: (i) that the encoding problem, particularly assigning description lengths to conceptual primitives, is difficult, and (ii) that nonetheless first results confirm that less complex analogies are also the preferred ones. In addition, the proposed model offers some bases for explaining the properties of analogy listed in section 1.3. Analogical thinking is pervasive because it is intrinsically an economical mode of thinking. Abstract analogies are preferred because they correspond to the most economical way of perceiving two things as similar. Good agreement in the ratings of analogies by different people would result from the sharing of the same optimal way of perceiving things. Analogy is non involutive and non idempotent

because of the asymmetry between the source and the target. And finally, analogy making can be improved in our model by the learning of prior probabilities of the perceptual primitives, therefore altering their description length, and thence the overall relative complexities of the various possible analogies. On the other hand, this last property is also responsible for the sensitivity to the coding prescription that renders it a delicate task.
This research reveals the deep interdependencies that may exist between statistics and forms of inference like analogy that seem to imply only symbolic manipulations of the descriptions of single situations. If we consider indeed, as we did here, any inferencing as the search for theories that provide economic encodings for the evidences, and if we recognize that an efficient coding scheme must reflect our prior expectations about the world, then it follows that inferencing will at the end rely on statistics. This is why symbolic reasoning and statistical analysis are two sides of the same coin that cannot be disjoined in Artificial Intelligence.

References

[Burstein 86] Burstein M.H. (1986): *"Concept Formation by Incremental Analogical Reasoning and Debugging"*. In Machine Learning: An Artificial Intelligence Approach (vol.2), Michalski, Carbonell, Mitchell (eds). Morgan Kaufmann, 1986.

[Chaitin 87] Chaitin G. (1987): *Information, Randomness and Incompletness - Papers on Algorithmic Information Theory.* World Scientific Press, 1987.

[Clem-Gent 91] Clement C.A. & Gentner D. (1991): *"Systematicity as a Selectional Constraint in Analogical mapping"*. Cognitive Science 15, pp.89-132.

[Cornuéjols 95] Cornuéjols A. (1995): *"Analogy as a minimization principle in operation: the details"*. L.R.I. technical report in preparation.

[Falkenheimer 87] Falkenheimer B. (1987): *"An examination of the third stage in the analogy process: Verification-based analogical learning"*. Proc. of the 10th International Joint Conference on Artificial Intelligence (IJCAI-87), Milan, Italy, 1987, pp.260-263.

[Gentner 83] Gentner D. (1983): *"Structure Mapping: A Theoretical Framework for Analogy"*. Cognitive Science, vol.7, 1983.

[Gick-Holyoack 80] Gick M.L. & Holyoak K.J. (1980): *"Analogical Problem Solving"*. Cognitive Psychology, vol.15, 1980.

[Gick-Holyoack 83] Gick M.L. & Holyoak K.J. (1983) *"Schema Induction and Analogical transfer"*. Cognitve Psychology 15, pp.1-38.

[Hofstädter 85] Hofstädter D. (1985): *Methamagical Themas: Questing for the Essence of Mind and Pattern.* Basic Books, 1985. See chapter 24: *"Analogies and Roles In Human and Machine Thinking"*, pp.547-603.

[Holyoack & Thagart 89] Holyoak K. & Thagard P. (1989): *"Analogical Mapping by Constraint Satisfaction"*. Cognitive Science 13:295-355, 1989.

[Keane 88] Keane Mark (1988): *"Analogical mechanisms"*. Artificial Intelligence Review (1988) 2:229-250.

[Kedar-Cabelli 85] Kedar-Cabelli S.T. (1985): *"Purpose-directed Analogy"*. In Proc. of the Cognitive Science Society Conference, 1985.

[Kolmogorov 65] Kolmogorov A. N. (1965): *"Three approaches to the quantitative definition of infomation"*. Problems Inform. Transmission, 1(1): 1-7, 1965.

[Li & Vitani 93] Li M. & Vitanyi P. (1993): *An Introduction to Kolmogorov Complexity and its Applications*. Springer Verlag, Texts and Monographs in Computer Science, 1993.

[Mitchell 93] Mitchell M. (1993): *Analogy-Making as Perception*. MIT Press, 1993.

[Pearl 78] Pearl J. (1978): *"On the connection between the complexity and credibility of inferred models"*. Int. J. Gen. Syst., 4:255-264, 1978.

[Rissanen 89] Rissanen J. (1989): *Stochastical Complexity and Statistical Inquiry*. World Scientific Publishing Company, 1989.

[Solomonoff 64] Solomonoff R. (1964): *"A formal theory of inductive inference"*. Information and Control, 7, 1-22 & 224-254, 1964.

[Wallace & Boulton 68] Wallace C. & Boulton D. (1968): *"An Information measure for classification"*. Computing Journal, 11:185-195, 1968.

[Winston 80] Winston P.H. (1980): *"Learning and Reasoning by Analogy"*. Communications of the ACM, vol.23, 1980.