South American Journal of Logic
Advance Access
ISSN: 2446-6719

Polya-Veloso-type Solutions for Metapuzzles

Frank Thomas Sautter

Abstract

We introduce a general method for solving metapuzzles inspired by Veloso's General Theory of Problems, a formal development of Polya's Problem Solving Technique. This method is not the most efficient one, but its virtues include faithful modeling of information flow and logical standardization through the same first-order predicates in all its applications.

Keywords: General Theory of Problems, Logical Puzzle, Logical Education, Semantic Information

Introduction

Raymond Smullyan introduces two types of puzzles in his books on recreational logic: a more straightforward one, related to knights, who always tell the truth, and knaves, who always lies, and a more sophisticated one called *metapuzzle*. He characterizes metapuzzles in the following manner:

> We are given a puzzle without sufficient data to solve it, and then we are given that someone else could or could not solve it given certain additional information, but we are not always told just what this additional information is. We may, however, be given partial information about it, which enables the reader to solve the problem [7, p. 91].

Unlike more straightforward puzzles, two types of information compose metapuzzles:

- explicit initial information, insufficient to solve the puzzle;

- additional implicit information deduced from someone's success or generally someone's failure to solve, at least partially, the puzzle. This second type of information characterizes them as metapuzzles.

There are well-known methods for solving more straightforward puzzles[1]. We propose in this paper a general framework for solving metapuzzles.

This framework instantiates a formal structure developed by Paulo Veloso [11], based on a technique for problem solving proposed by George Polya [6]. It accurately models the flow of information. A space of the initially possible states of affairs models the initial information. We infer additional information from propositions, representing someone's success or failure, and eliminating initially possible states of affairs models this additional information.

In the first section, we will present,*in abstrato*, this framework for solving metapuzzles. In the second section, we will present two examples of the application of this framework: one example is positive, because what is asked for is to determine what someone has, and the other is negative because what is asked for is to determine what someone does not have. In the third section, we will present Polya's Problem Solving Technique [6], its formalization by Veloso [11], and the relation of the framework developed here with the formalization of Veloso.

1 A Framework for Solving Metapuzzles

Let us begin with the more straightforward puzzles. Smullyan [9, p.55-66] shows how to solve simple puzzles about knights and knaves. If A_i is a native of the Island of Knights and Knaves and asserts a proposition P, this can be formalized as $k_i \equiv P$.

Let us show how it works in practice.

In his diaries, Lewis Carroll presents the following simple puzzle [1, p.11]:

> The Dodo says that the Hatter tells lies.
> The Hatter says that the March Hare tells lies.
> The March Hare says that both the Dodo and the Hatter tell lies.

[1] We will give some examples in the next section.

$k_{Dodo} \equiv \neg k_{Hatter}$ formalizes the first proposition; $k_{Hatter} \equiv \neg k_{MarchHare}$ formalizes the second one; and $k_{MarchHare} \equiv (\neg k_{Dodo} \wedge \neg k_{Hatter})$ formalizes the third one. It is not difficult to deduce that the Hatter tells the truth and both the Dodo and the March Hare tell lies.

Let us show another example.

Martin Hollis gives another example of an elementary puzzle, but this one is not about knights and knaves, although it can be solved in the same way [3, p.21]:

> George: 'Today is not Thursday.'
> Henry: 'That's true.'
> Ivan: 'We have made one false statement between us.'

$k_{George} \equiv p_1$, where p_1 stands for Georges statement, formalizes the first proposition; $k_{Henry} \equiv p_2$, where p_2 stands for Henry's statement, formalizes the second one; and $k_{Ivan} \equiv [(\neg k_{Ivan} \wedge p_1 \wedge p_2) \vee (\neg p_1 \wedge p_2 \wedge k_{Ivan}) \vee (\neg p_2 \wedge p_1 \wedge k_{Ivan})]$ formalizes the third one. It is also truth that $p_1 \equiv p_2$.

It is not difficult to deduce that all of them lie and today is Thursday!

Nevertheless Smullyan's framework is insufficient to solve metapuzzles, even those only of knights and knaves. So, I propose a general logical framework to solve metapuzzles, consisting of the following steps:

1. Identify the space of the initially possible state of affairs. A first-order predicate P, whose arity is equal to the number of characters in the puzzle, formalizes this space.
2. From the space of the initially possible state of affairs, discard those initially possible state of affairs incompatible with new information available, given as someone's success or failure to solve, at least partially, the puzzle. The first-order predicates D_i, of the same arity as P, one for each time a new relevant information is given, formalize this discard.
3. After all the new relevant information is given, collect those possible state of affairs not discarded into a first-order predicate M, of the same arity as P and the D_i's.

In the second step, the following non-logical rule of inference – Rule of Preservation of Information – is part of the deductive system: from $D_i x_1, \ldots, x_n$ to deduce $D_{i+1} x_1, \ldots, x_n$, for i from 0 (time in which the first step occurs) to the last but one[2].

The *immediate* solution is collected into a set Σ of tuples that satisfy M.

Most of the time, a logical puzzle requires a secondary solution. But it is always deducted from the set Σ.

2 Tahan's Metapuzzle

Our first example to the general approach presented at the previous section is extracted from the fiction book "The man who counted" [10], written by the Brazilian writer Júlio César de Mello e Souza under the Arab pseudonym Malba Tahan[3]. Chapter 31 (In black and white) presents the following logical puzzle, perhaps one of the simplest hat-type puzzles[4],[5]:

> "The three princes were summoned to the palace, and the dervish, showing them five simple wooden disks, said to them,'Here are five disks, two of them black and three of them white. They are all the same size and weight and are different only in color.'
> "Next, a page carefully bound the eyes of the three princes so that they could see nothing. The old dervish then picked three disks at random and fastened one each to the backs of the three suitors, saying as he did so, 'Each one of you has on his back a disk whose color you do not know. You are to be questioned in turn. The one who discovers the color of the disk he is wearing will be declared the winner and will receive the hand of the beautiful Dahize in marriage. The first one questioned can look at the disks of the other two. The second can see only the disk of the third, and the third must make his reply seeing none of the others. The one who

[2] A formalization of Smullyan's "Alice in Puzzle-Land" [8] suggests the use of non-logical rules of inference. Pereira and Peron [5] solved Smullyan's puzzles in that book with the following "Rule of Madness": $Lt \dashv\vdash B_t\alpha \equiv \neg\alpha$, where Lt means that the individual named t is mad, and $B_t\alpha$ means that the individual named t believes in α.

[3] The first edition was published in 1938 with the Portuguese title "O Homem que Calculava".

[4] But, instead of hats, Tahan's speaks of disks.

[5] Hat-type puzzles are very popular. Hartston [2] published recently a book on puzzles whose sixth chapter is entirely devoted to hat-type puzzles.

> gives the correct answer must, in order to prove that he was not simply guessing, justify his answer by clear reasoning. Now, who wants to go first?'
> " 'Let me be first,' said Prince Comozan promptly.
> "The page removed the bandage from his eyes, and Prince Comozan saw the disks on the backs of his two rivals. The dervish took him aside to hear his answer, but it was wrong. Declaring himself beaten, he withdrew. He had seen the two disks on the backs of the other princes and still not been able to determine the color of his own disk.
> " Prince Comozan has failed,' said the king in a loud voice, to inform the other two.
> " Then let me be next,' said Prince Benefir. Once his eyes were uncovered, the second prince saw the disk worn by the third on his back. He motioned to the dervish and whispered his reply to him. The dervish shook his head. The second prince was also mistaken and was given leave to withdraw immediately. Only one was left, Prince Aradin.
> "When the king announced that the second suitor had also failed, he approached with his eyes still bandaged and announced in a loud voice the correct color of the disk on his back."

One way or another, we resort to a version of Dirichlet's Principle to solve all puzzles of this type. The main version of Dirichlet's Principle asserts that if we have n items and m boxes, with $n \geq m$, then at least one box must contain at least k itens, where k is the least integer greater than or equal to $n \div m$. But what if $m \geq n$? If we understand the Dirichlet's Principle as establishing a lower bound on the maximal number of items in a box, then, if $m \geq n$, at least $m - n$ boxes contain no items. For example, we can apply this last version of Dirichlet's Principle to Comozan's reasoning in the following way: there are three boxes (the back of the three suitors) and two items (black disks); so, by Dirichlet's Principle, at least one box (one back of a suitor) has no item (black disk) on it.

One of the best Brazilian textbooks to logic gives a solution to Tahan's Metapuzzle according to the following lines[6][4, p. 8-10]:

Dictionary;

[6]It gives an informal solution; we will give a formal reconstruction of it. It also gives a slightly different wording from the puzzle.

- b: black;
- w: white;
- a: Aradin;
- e: Benefir;
- o: Comozan;
- Ax: The color of Aradin's disk is x;
- Bx: The color of Benefir's disk is x;
- Cx: The color of Comozan's disk is x;
- Kx: x knows the color of his own disk.

Boundary conditions:
$Ab \vee Aw$, $\neg(Ab \wedge Aw)$, $Bb \vee Bw$, $\neg(Bb \wedge Bw)$, $Cb \vee Cw$, $\neg(Cb \wedge Cw)$.

Comozan's failure:
$(Bb \wedge Ab) \supset (Ko \wedge Cw)$ and $\neg Ko$. It follows that $\neg Bb \vee \neg Ab$ and $Bw \vee Aw$.

Benedir's failure:
$Ab \supset (Ke \wedge Bw)$ and $\neg Ke$. It follows that $\neg Ab$ and Aw.

This formalization has, at least, two defects[7]:

- It does not establishes the connection between Comozan's failure and Benefir's one[8].
- It does not minimally models the flow of information.

[7]The puzzle occurs at very beginning of the textbook, so the focus is not on details, but on exemplification of lines of reasoning.

[8]The pair of propositions corresponding to Comozan's failure is incomparable with the pair of propositions corresponding to Benefir's failure, but what happens if we put them into the same language for comparison purposes? $[Ab \supset (Ko \wedge Cw)] \wedge \neg Ko$ is more informative than $[(Bb \wedge Ab) \supset (Ko \wedge Cw)] \wedge \neg Ko$. So, in some sense, Benefir's failure is more informative than Comozan's failure; it benefits from Comozan's failure.

A formalization of the problem and its solution, in the manner proposed in the previous section, does not suffer from these defects. It is the following:

Dictionary:

- b: black (this individual is a type, not a token);
- w: white (this individual is also a type, not a token);
- $Pxyz$: It is initially possible that simultaneously Comozan's disk is of type x, and Benefir's disk is of type y, and Aradin's disk is of type z;
- D_ixyz: It was discarded at time i that simultaneously Comozan's disk is of type x, and Benefir's disk is of type y, and Aradin's disk is of type z, for $0 \leq i \leq n$, where n is the n_{th} time a new information is given;
- $Mxyz$: It is possible, all information considered, that simultaneously Comozan's disk is of type x, and Benefir's disk is of type y, and Aradin's disk is of type z.

Initial situation Γ_0:
$\Delta = \{Pwww, Pwwb, Pwbw, Pbww, Pwbb, Pbwb, Pbbw\}$[9]
$\Gamma_0 = \Delta \cup \{\neg D_0xyz : Pxyz \in \Delta\}$

Situation Γ_1 *after Comozan's failure*:
$\Gamma_1 = \Gamma_0 \cup \{\phi : \Gamma_0 \cup condition_1 \models \phi\}$,
where $condition_1$ is
$\forall x \forall y \forall z\{[Pxyz \wedge \neg D_0xyz \wedge \neg\exists(Pvyz \wedge \neg D_0vyz \wedge v \neq x)] \supset D_1xyz\}$[10]

Situation Γ_2 *after Benefir's failure*:
$\Gamma_2 = \Gamma_1 \cup \{\phi : \Gamma_1 \cup condition_2 \models \phi\}$,
where $condition_2$ is
$\forall x \forall y \forall z\{[Pxyz \wedge \neg D_1xyz \wedge \neg\exists(Pxvz \wedge \neg D_1xvz \wedge v \neq y)] \supset D_2xyz\}$

Remember, from the previous section, that $D_1k_1k_2k_3 \models D_2k_1k_2k_3$, for every individual constants k_1, k_2, and k_3.

[9] It is not difficult to construct an algorithm that produces all the elements of Δ.

[10] The domain of discourse is finite, and we have one individual constant for each individual in the domain of discourse, so there is no need to express the conditions with quantifiers. But they become cumbersome if we do not use them.

Final situation:
$\forall x \forall y \forall z (Mxyz \equiv (Pxyz \wedge \neg D_2xyz))$
The solution is the set $\Sigma = \{< x, y, z >: Mxyz\}$

It is not difficult to show that the following propositions can be deduced:

- D_1wbb and, by the Rule of Preservation of Information, D_2wbb;
- D_2wwb;
- D_2bwb;
- For all others triples $< x, y, z >$ such that $Pxyz$, $\neg D_2xyz$;
- $\Sigma = \{< w, w, w >, < wbw >, < b, w, w >, < b, b, w >\}$.

Σ is the *immediate* solution.

If we want to know which disk each character uses, it is necessary to introduce new predicates and conditions of satisfaction for them:
New dictionary items:

- C_cx: It is know by Comozan (and all the other contenders) that his disk is of type x;
- C_bx: It is known by Benefir (and all the other contenders) that his disk is of type x;
- C_ax: It is known by Aradin (and all the other contenders) that his disk is of type x.

Conditions of satisfaction for the new predicates:

- $\forall x[C_cx \equiv \forall v \forall y \forall z (Mvyz \supset v = x)]$
- $\forall x[C_bx \equiv \forall y \forall v \forall z (Myvz \supset v = x)]$
- $\forall x[C_ax \equiv \forall y \forall z \forall v (Myzv \supset v = x)]$

It is not difficult to prove than only C_aw can be deduced with respect to these predicates. This is the *secondary* solution, although the one required in this story.

3 Smullyan's Metapuzzle

Our second example to the framework presented at the first section is extracted from Smullyan's book "The Lady or the Tiger" [7]. Problem 9 (A New "Colored Hats" Problem) from Chapter 1 presents the following puzzle:

> Three subjects – A, B, and C – were all perfect logicians. Each could instantly deduce all consequences of any set of premises. Also, each was aware that each of the others was a perfect logician. The three were shown seven stamps: two red ones, two yellows ones, and three green ones. They were then blinfolded, and a stamp was pasted on each of their foreheads; the remaining four stamps were placed in a drawer. When the blinfolds were removed, A was asked, "Do you know one color that you definitely do not have?" A replied, "No." The B was asked the same question and replied, "No."
> Is it possible, from this information, to deduce the color of As stamp, or of B's, or of C's?

Tahan's logical puzzle can be seen as a *positive* one, because what is requested is the color of the wooden disk that *is* on the back of each character. On the other hand, Smullyan's logical puzzle can be seen as a *negative* one, because what is requested is a stamp's color that *is not* on the forehead of a character. Smullyan's logical puzzle is, also, a bit more complex than Tahan's one, but both can be solved in a common framework.

A formalization of the problem and its solution, in the manner proposed in the first section, is the following:

Dictionary:

- r: red (a type individual);
- y: yellow (also a type individual);
- g: green (another type individual);
- $Pvwz$: It is initially possible that simultaneously A's forehead stamp is of type v, and B's forehead stamp is of type w, and C's forehead stamp is of type z;

- D_ivwz: It was discarded at time i that simultaneously A's forehead stamp is of type v, and Bs forehead stamp is of type w, and C's forehead stamp is of type z;

- $Mvwz$: It is possible, all information considered, that simultaneously A's forehead stamp is of type v, and B's forehead stamp is of type w, and C's forehead stamp is of type z.

Initial situation Γ_0:
$\Delta = \{Pggg, Pggy, Pggr, Pgyg, Pgrg, Pygg, Prgg, Pyyg, Pyyr, Pygy, Pyry, Pgyy, Pryy, Prrg, Prry, Prgr, Pryr, Pgrr, Pyrr, Pgyr, Pgry, Pygr, Pyrg, Prgy, Pryg\}$[11]
$\Gamma_0 = \Delta \cup \{\neg D_0xyz : Pxyz \in \Delta\}$

Situation Γ_1 *after A's failure*:
$\Gamma_1 = \Gamma_0 \cup \{\phi : \Gamma_0 \cup condition_1 \models \phi\}$,
where $condition_1$ is
$\forall w \forall z\{(\neg Pgwz \vee \neg Pywz \vee \neg Prwz) \supset \forall v[(Pvwz \wedge \neg D_0vwz) \supset D_1vwz]\}$

Situation Γ_2 *after B's failure*:
$\Gamma_2 = \Gamma_1 \cup \{\phi : \Gamma_1 \cup condition_2 \models \phi\}$,
where $condition_2$ is
$\forall v \forall z\{(\neg Pvgz \vee \neg Pvyz \vee \neg Pyrz) \supset \forall w[(Pvwz \wedge \neg D_1vwz) \supset D_2vwz]\}$

Remember, from the first section, that $D_1k_1k_2k_3 \models D_2k_1k_2k_3$, for every individual constants k_1, k_2, and k_3.

Final situation:
$\forall x \forall y \forall z(Mxyz \equiv (Pxyz \wedge \neg D_2xyz))$
The solution is the set $\Sigma = \{< x, y, z >: Mxyz\}$

It is not difficult to show that the following propositions can be deduced:

- D_1gyy, D_1ryy, D_1grr, D_1yrr and, by the Rule of Preservation of Information, D_2gyy, D_2ryy, D_2grr, D_2yrr ;

- D_2ygy, D_2yry, D_2rgr, D_2ryr;

- D_2rry, D_2rgy; D_2yyr, D_2ygr, D_2ggy, D_2gry, D_2ggr, D_2gyr;

[11] It is not difficult to construct an algorithm that produces all the elements of Δ.

- For all others triples $< v, w, z >$ such that $Pvwz$, $\neg D_2vwz$;
- $\Sigma = \{< g, g, g >, < g, y, g >, < g, r, g >, < y, g, g >, < r, g, g >, < y, y, g >, < y, r, g >, < r, y, g >\}$.

Σ is the *elementary* solution.

If we want to know which stamp each character uses, it is necessary to introduce new predicates and conditions of satisfaction for them:

New dictionary items:

- N_Ax: It is know by A (and all the other contenders) that her stamp is *not* of type x;
- N_Bx: It is known by B (and all the other contenders) that her stamp is *not* of type x;
- N_Cx: It is known by C (and all the other contenders) that her stamp is *not* of type x.

Conditions of satisfaction for the new predicates:

- $\forall x[N_Ax \equiv \forall v \forall w \forall z(Mvwz \supset v \neq x)]$
- $\forall x[N_Bx \equiv \forall v \forall w \forall z(Mvwz \supset w \neq x)]$
- $\forall x[N_Cx \equiv \forall v \forall w \forall z(Mvwz \supset z \neq x)]$

It is not difficult to prove than only N_Cy and N_Cr can be deduced with respect to these predicates. This is the *secondary* solution, although the one required in this story.

4 Polya and Veloso on a General Theory of Problems

Our framework comes from Veloso's General Theory of Problems [11], and this is derived from the three questions that, according to Polya [6, p. 2], need to

be answered in order to understand a problem.[12] First: *What is the unkown?* Second: *What are the data?* Last: *What is the condition?* About them, Polya [6, p. 2] says:

> These questions are generally applicable, we can ask them with good effect dealing with all sorts of problems. Their use is no restricted to any subject-matter. Our problem may be algebraic or geometric, mathematical or nonmathematical, theoretical or practical, a serious problem or *a mere puzzle* [my emphasis]; it makes no difference, the questions make sense and might help us to solve the problems.

Despite its universal applicability, there is a particular type of problem for which these questions are particularly relevant. Polya [6, p.154] distinguishes *problems to find* from *problems to prove*. The first ones aim "to find a certain object, the unknown of the problem."[13]. On the other hand, problems to prove aim "to show conclusively that a certain clearly stated assertion is true, or else to show that it is false." Metapuzzles are plainly problems to find[14].

Veloso [11, p.136] transforms these questions into a two-sorted mathematical structure $< D, R, q >$[15], such that:

- D is a nonempty set, called the domain of (input) data,
- R is a nonempty set, called the domain of (output) data,
- q is a binary relation from D to R, called the problem requirement.

He also imposes that "a solution should assign to each input data a result so as to satisfy the problem requirement. So, we define a solution [...] to be a

[12] Polya [6, p.xvi-xvii] divides the task of solving a problem in four phases. First: understanding the problem. Second: Devising a plan. Third: Carrying out the plan. Fourth: looking back. The three questions to be answered are included in the first phase.

[13] Polya [6, p.154] also calls this unknown the "quaesitum", or the thing sought, or the thing required.

[14] By the way, Polya [6, p. 155] says that 'the principal parts of a "problem to find" are the *unknown*, the *data*, and the *cognition*', and he also says [6, p. 155] that 'if you wish to solve a "problem to find" you must know, and know very exactly, its principal parts, the unknown, the data, and the condition.'

[15] Veloso [11, p. 136] calls it a *concrete problem*.

(total) function $f : D \to R$ such that for every d in D one has $(d, f(d))$ in the relation q." [11, p.136]

A metapuzzle is a very particular type of concrete problem. Let $S = \{< x_1, \ldots, x_n >: Px_1 \ldots x_n\}$. In this case $D = \{S\}$, $R = \wp(S)$, and $f : D \to R$ is such that $f(S) = \{e \in \Sigma\}$.

References

[1] L. Carroll. *Lewis Carroll's Games and Puzzles.* Edited by E. Wakeling. Dover, 1992.

[2] W. Hartston. *A Brief History of Puzzles.* Atlantic Books, 2019.

[3] M. Hollis, *Tantalizers:* a book of original logical puzzles. George Allen and Unwin, 1970.

[4] C. Mortari, *Introdução à Lógica* (*Introduction to Logic*, in Portuguese). EDUNESP, 2001.

[5] L.P.Pereira & N.M. Peron (supervisor), *A Lógica da Loucura em "Alice no País dos Enigmas" de Smullyan* (*The Logic of Madness in "Alice in Puzzle-Land" by Smullyan*, in Portuguese). Degree in Philosophy. Federal University of Fronteira Sul (Brazil), 2018.

[6] G. Polya, *How to solve it:* A New Aspect of Mathematical Method, Second Edition, Princeton University Press, 1973.

[7] R. Smullyan, *The Lady or the Tiger:* And Other Logic Puzzles Including a Mathematical Novel That Features Gödel's Great Discovery. Knopf, 1982.

[8] R. Smullyan, *Alice in Puzzle-Land*: A Carrollian Tale for Children Under Eighty. Penguin Books, 1982.

[9] R. Smullyan, *Logical Labyrinths.* A. K. Peters, 2009.

[10] M. Tahan, *The Man Who Counted:* A Collection of Mathematical Adventures. Translated by L. Clark & A. Reid. W.W. Norton & Company, 1993.

[11] P. A. S. Veloso, *On the Concepts of Problem and Problem-Solving Method.* Decision Support Systems 3 (1987), 133 - 139.

Frank Thomas Sautter
Department of Philosophy

Federal University of Santa Maria(UFSM)
Avenida Roraima, 1000, CEP 97105-900, Santa Maria, RS, Brazil
E-mail: ftsautter@ufsm.br