Automated Solution of the Riddle of Dracula and Other Puzzles

László Aszalós

IRIT, Universite Paul Sabatier,
118 route de Narbonne F-31062 Toulouse Cedex 4, France,
aszalos@irit.fr

Abstract. The Door of Freedom is a well-known and a hard puzzle. In R. M. Smullyan’s books on puzzles there are some similar puzzles. In this article we shall show how can we solve automatically such puzzles.

1 Introduction

Raymond M. Smullyan wrote some book on logical puzzles [4, 5]. Adam Kolany [2] classified these puzzles into three categories:

Guess, who I am! In this case we know all the hypothesis, and we need to select a formula from a set of formulae which is a conclusion of the hypothesis.

I’ve forgotten what You said. (Metapuzzles) In this case we know all the hypothesis but one, which is a member of a known set of formulae. We know that an element of another set of formulae is a conclusion of the hypothesis.

What am I to say?/What shall I ask for? In this case we know the conclusion and all the hypothesis but one. We need to find the unknown hypotheses.

The puzzles of the first and the second category can be solved easily, we only need to check systematically that there is a valid consequence between the hypothesis and the conclusion.

To solve the puzzles of the third category, we need to find an unknown formula. As for as we know, only A. Kolany [2] gave a solving method for puzzles of this category, but he didn’t prove the correctness and completeness of his method.

In this article we shall use propositional logic to formulate puzzles, and for this we sometimes get some very complicated formulae. (The author used modal logic languages in [1] to formulate these puzzles, and got simpler formulae.)

We briefly describe Smullyan’s stories: in the puzzles of knights and knaves each inhabitant is either a knight or a knave. The knights always tell the truth and knaves always lie. In the puzzles of knights, knaves and normals, each inhabitant is either a knight or a knave or a normal. The knights and the knaves are same as before, the normals sometimes tell the truth and sometimes lie. Finally each Transylvanian is either a human or a vampire, and is either sane or insane. The sane humans and the insane vampires tell the truth and the others are lie. A small subgroup of Transylvanians instead of using the words “Yes” and “No” use “Bal” and “Da”. The trouble is that we don’t know which of “Bal” or “Da” means “Yes” and which means “No”.

⋆ on leave from University of Debrecen, Hungary, aszalos@math.klte.hu
The paper is organized as follows: we introduce the syntax (Sect. 2) and we review the first solving method (Sect. 3). We solve the Riddle of Dracula with this method (Sect. 4). Then we give another solving method, because the first isn’t general enough, and we cannot solve all the puzzles.

2 Syntax and semantics

The set of formulas is the smallest set which contains the set of propositional letters and closed under negation ($\neg$), conjunction ($\land$), disjunction ($\lor$), implication ($\supset$) and equivalence ($\equiv$).

2.1 Island of knights and knaves

We shall use the propositional letters $T_x$, to express that the person $x$ is a knight. In the puzzles only knights and knaves appear, so somebody is a knave iff he isn’t a knight, hence we shall write $\neg T_x$ to denote that the person $x$ is a knave. The inhabitant $x$ can say the statement $A$ iff he is a knight and the statement $A$ is true or he is a knave and $A$ is false: $$(T_x \land A) \lor (F_x \land \neg A),$$ i.e. $T_x \equiv A$.

2.2 Island of knights, knaves and normals

We shall use the propositional letters $T_x$, $F_x$ and $N_x$ to express, that the person $x$ is a knight, is a knave or is a normal, respectively. For every person $x$ exactly one of $T_x$, $F_x$ and $N_x$ can be true, hence the formula

$$(T_x \land \neg F_x \land \neg N_x) \lor (\neg T_x \land F_x \land \neg N_x) \lor (\neg T_x \land \neg F_x \land N_x) \tag{1}$$

for any $x$ will be a hypothesis of puzzles. The inhabitant $x$ of the island of knights, knaves and normals can say the statement $A$ iff he is a knight and the statement $A$ is true, or he is a knave and $A$ is false, or he is a normal: $$(T_x \land A) \lor (F_x \land \neg A) \lor \neg N_x.$$ 

2.3 Transylvania

We shall use the propositional letters $M_x$ and $V_x$ to express that the person $x$ is an insane or is a vampire, respectively. In Transylvania only the sane humans and insane vampires can say the truth, so one can say the truth if he is an insane if and only if he is a vampire, too. Hence the Transylvanian $x$ can say the statement $A$ iff this formula is true: $$(M_x \equiv V_x) \equiv A.$$ 

3 First method

In Smullyan’s $\alpha$ and $\beta$ tableaux rules [3] the formula written above the rule is equivalent to the formulae written below the rule:

**Lemma 1.** $\alpha \equiv (\alpha_1 \land \alpha_2)$ and $\beta \equiv \beta_1 \lor \beta_2$. 
Corollary 2. If formula $B_{ij}$ is the $j^{th}$ formula of the $i^{th}$ branch of the tableau of formula $A$, then $A \equiv \bigvee_i \bigwedge_j B_{ij}$, moreover from this d.n.f. we can leave out the formulae which are opened while we build the tableaux for $A$.

Let us examine at first a well-known puzzle:

Example 3. Door of Freedom. In a cell there are two doors. If we choose the right one we get free, but in the other case we will die. There are two guards in the cell, one of them always tells the truth, the other always lies. We can put a yes-no question to one of them before we choose a door. What is the right question?

We shall formalize the example as follows: Let the propositional letter $D$ denote that the left door is the right one, and call the guards $a$ and $b$. We would like to get a formula $X$ such that if we ask the guard “Is it true that $X$?" then he answers “yes” if the left door is the right one ($D$), and he answers “no” if the right door is the right one ($\neg D$).

In general if we ask the person $a$ “Is it true that $X$?" where $X$ is a formula, then if he answers “yes” or “yes, $X$ is true” then he can say the formula $X$, and if he answers “no” or “no, $X$ is false” then he can say the formula $\neg X$.

If we realize that the two guards can be the inhabitants of the island of knights and knaves then we can easily formalize the example. The hypothesis of the example is the following: the types of the guards are different, that is $\neg (T_a \equiv T_b)$. By applying this to our example we get the following:

$$(\neg (T_a \equiv T_b) \land (T_a \equiv X) \supset D) \land (\neg (T_a \equiv T_b) \land (T_a \equiv \neg X) \supset \neg D).$$

After some simplification we get

$$(T_a \equiv X) \equiv D$$

The puzzle have a solution, if this formula is valid for some formula $X$. Let assume the this formula is valid so we shall get a closed tableau. (We treat the formula $X$ as atomic at this stage.) Using the Corollary 2, after some simplification we get the

<table>
<thead>
<tr>
<th>$T_a$</th>
<th>$\neg T_a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_a$</td>
<td>$\neg T_a$</td>
</tr>
<tr>
<td>$X$</td>
<td>$\neg X$</td>
</tr>
<tr>
<td>$\neg T_a$</td>
<td>$T_a$</td>
</tr>
<tr>
<td>$T_a$</td>
<td>$\neg T_a$</td>
</tr>
<tr>
<td>$\neg T_a$</td>
<td>$T_a$</td>
</tr>
</tbody>
</table>

Table 1. Analytic tableau of Example 3
following formula (which is equivalent to negation of Formula 2).

\[(X \land T_a \land \neg T_b \land \neg D) \lor (X \land \neg T_a \land T_b \land D) \lor (\neg X \land \neg T_a \land T_b \land \neg D) \lor (\neg X \land T_a \land \neg T_b \land D).\] (3)

We separate the conjunctions containing the unknown formula \(X\) and its negation. By grouping the conjunctions we get the following equivalent form:

\[
\left( X \land ((T_a \land \neg T_b \land \neg D) \lor (\neg T_a \land T_b \land D)) \right) \lor \\
\left( \neg X \land ((\neg T_a \land T_b \land \neg D) \lor (T_a \land \neg T_b \land D)) \right)
\] (4)

The tableau is closed, so the Formula 3 is constant false, hence the Formula 4 is constant false, therefore Formulae 5 and 6 are constant false, too.

\[
\neg X \land ((T_a \land \neg T_b \land \neg D) \lor (\neg T_a \land T_b \land D)) \] (5)

\[
X \land ((\neg T_a \land T_b \land \neg D) \lor (T_a \land \neg T_b \land D)) \] (6)

Since Formulae 5 and 6 are constant false, hence the Formulae 7 and 8 are valid.

\[
\left( (T_a \land \neg T_b \land \neg D) \lor (\neg T_a \land T_b \land D) \right) \supset X \] (7)

\[
X \supset \neg\left( (\neg T_a \land T_b \land \neg D) \lor (T_a \land \neg T_b \land D) \right) \] (8)

In propositional logic Craig’s interpolation theorem holds. Hence if formula \(E \supset \neg F\) is valid, then there exists a solution. Smullyan gave a constructive method [3] to find the interpolant, but we now construct the solutions with truth-tables.

Let us build a truth-table, where the rows belongs to the different valuations of propositional letters of \(E\) and \(F\). Now this propositional letters are \(T_a\), \(T_b\) and \(D\). The truth-table has three columns respectively for \(E\), \(X\) and \(\neg F\). Let us fill the columns of \(E\) and \(\neg F\) according to the valuations. For each row let us do the following:

1) If the column of \(E\) contains 1 then write 1 into the column of \(X\).
2) If the column of $\neg F$ contains 0 then write 0 into the column of $X$.

If these two rules contradict each other then there doesn’t exist a suitable $X$, therefore the puzzle has no solution, we can stop. With these two rules we need to fill four places (marked with arrows). The remaining four place (marked with $a$, $b$, $c$ and $d$) can be filled arbitrary, so we have 16 different solutions.

Let us compose the formula $X$ according to the truth-table in a standard way. At first let us choose the following values: $a = 1$, $b = 0$, $c = 0$ and $d = 1$ Then the synthesize form of the formula $X$ will be the following:

$$(T_a \land T_b \land D) \lor (T_a \land \neg T_b \land \neg D) \lor (\neg T_a \land T_b \land D) \lor (\neg T_a \land \neg T_b \land \neg D).$$

This formula can be simplified in different ways, and each variant correspond to an English sentence:

$– \neg (T_b \equiv D)$: “It isn’t true, that the other guard is a knight iff the left door is the right one?”

$– \neg T_b \equiv D$: “The other guard is a knave iff the left door is the right one?” “The other guard could say the left door isn’t the right one?” or “The other guard could say the right door is the right one?”

If we set $a = 0$, $b = 1$, $c = 1$ and $d = 0$, then the formula $X$ will be the following:

$$(T_a \land T_b \land \neg D) \lor (T_a \land \neg T_b \land \neg D) \lor (\neg T_a \land T_b \land D) \lor (\neg T_a \land \neg T_b \land D).$$

Hence we have two more solutions:

$– T_a \equiv D$: “Are you a knight iff the left door is the right one?” or “Could you answer ‘yes’ if I ask you ‘the left door is the right one’?”

Sometimes not all the hypotheses of the puzzle can be formalized with valid formulae. In this case we use the first method only for the valid formulae. And we reject the solutions of the method not satisfying some of the other hypotheses.

4 Riddle of Dracula

In this section we shall abbreviate formula $(M_a \equiv V_a) \equiv X$ as $C_a X$ to make formulae more readable. As we wrote in the introduction the Transylvanian elite uses the words “Bal” and “Da” instead of “Yes” and “No” but we don’t know which is which. Let $B$ be a propositional letter, which is true iff “Bal” means “Yes”. When do we get for the question “It is true that $X$?” the answer “Bal” from the person $a$? There are two cases:

1) if his answer would be “Yes” and “Bal” means “Yes” and
2) if his answer would be “No” and “Bal” means “No”:

$$(C_a X \land B) \lor (C_a \neg X \land \neg B). \quad (9)$$

The formula $C_a \neg X \equiv \neg C_a X$ is valid, so the formula (9) can be shortened to $C_a X \equiv B$. In a similar way we can get that $a$ answers “Da” to the same question iff $C_a X \equiv \neg B$. Now we are ready to solve the famous riddle of Dracula [4, Puzzle 196]:
Example 4. There is one sentence $S$ having the almost magical property that given any sentence $X$ whose you wish to ascertain, all you would have to do is ask anyone in this castle, “Is $S$ equivalent to $X$?” If you get “Bal” for an answer, $X$ must be true; If you get “Da” for an answer, $X$ must be false. What is the sentence $S$?

If we believe in Smullyan (and why not), the Count Dracula who said the riddle is an insane vampire, so the whole sentence is true. If person $a$ answers “Bal” to the question “It is true that $S \equiv X$?” then $X$ is true:

$$(Ca(S \equiv X) \equiv B) \supset X.$$  

If he answers “Da” to the question “It is true that $S \equiv X$?” then $X$ is false:

$$(Ca(S \equiv X)) \equiv \neg B \supset \neg X.$$  

So the formula of the puzzle is

$$((Ca(S \equiv X)) \equiv B) \equiv X.$$  

If we would make the tableau of the negation of this formula, then it wouldn’t fit on the page. So we need to apply some dirty trick, for example the associativity of equivalence: $(A \equiv B) \equiv C \sim (A \equiv C) \equiv B$. Now we are ready to build the tableau (Fig. 3). This tableau isn’t finished yet, but enough to get the solution. Based on this

<table>
<thead>
<tr>
<th>$\neg((M_a \equiv V_a) \equiv ((S \equiv B) \equiv X)) \equiv X$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(M_a \equiv V_a) \equiv ((S \equiv B) \equiv X)$</td>
</tr>
<tr>
<td>$\neg X$</td>
</tr>
<tr>
<td>$(V_a \equiv M_a)$</td>
</tr>
<tr>
<td>$(S \equiv B) \equiv X$</td>
</tr>
<tr>
<td>$\neg(V_a \equiv M_a)$</td>
</tr>
<tr>
<td>$(S \equiv B) \equiv X$</td>
</tr>
<tr>
<td>$(S \equiv B)$</td>
</tr>
<tr>
<td>$(S \equiv B)$</td>
</tr>
<tr>
<td>$\neg(S \equiv B) \equiv X$</td>
</tr>
<tr>
<td>$(S \equiv B)$</td>
</tr>
<tr>
<td>$(S \equiv B)$</td>
</tr>
<tr>
<td>$\neg(S \equiv B)$</td>
</tr>
<tr>
<td>$X$</td>
</tr>
<tr>
<td>$\neg X$</td>
</tr>
<tr>
<td>$S$</td>
</tr>
<tr>
<td>$\neg S$</td>
</tr>
<tr>
<td>$\neg S$</td>
</tr>
<tr>
<td>$\neg B$</td>
</tr>
<tr>
<td>$B$</td>
</tr>
<tr>
<td>$\neg B$</td>
</tr>
</tbody>
</table>

Table 3. Analytic tableau of Example 4

tableaux we get that the formulae $(B \equiv (V_a \equiv M_a)) \supset S$ and $S \supset (B \equiv (V_a \equiv M_a))$ are valid, so formula $S$ is equivalent to

$$(B \equiv (V_a \equiv M_a)). \quad (10)$$

We can transform and interpret the formula (10) in several ways. For example $S$ is equivalent to $(B \equiv \top \equiv (V_a \equiv M_a))$ so the question can be the following: “Would You answer ‘Bal’ to the question ‘2 plus 2 is equal 4?’?"
5 Second method

The first method has some restrictions:

– the original formula must be valid,
– we can have exactly one unknown formula, or (as we see at the riddle) the other unknown formulas need to drop out at simplification.

Some of Smullyan’s puzzles contain more unknown sentences. By formalizing we get more unknown formulae, so the former method is unusable. To solve these puzzles we need a new method based on the fact that the unknown formulae are logical combinations of the propositional letters. If the propositional letters we need to formulate the puzzle are \( P_1, P_2, \ldots, P_n \), and the unknown formulae are \( X_1, X_2, \ldots, X_m \), then we can assume that \( X_i = \phi_i(P_1, \ldots, P_n) \) for all \( i \). The formula of the puzzle (\( \psi \)) can be treated as a logical combination of the propositional letters and the unknown formulae: \( \psi(P_1, \ldots, P_n, \phi_1(P_1, \ldots, P_n), \ldots, \phi_m(P_1, \ldots, P_n)) \). We build a special truth-table of formula \( \psi \) which has \( 2^n \) rows and \( 2^m \) columns according to the propositional letters and unknown formulae. Formula \( \psi \) is a valid formula, hence in each row of its truth-table there must to be at least one 1, because formula \( \psi \) can be treated as a logical combination of the propositional letters: \( \psi'(P_1, \ldots, P_n) \).

Let us construct this truth-table of Example 3 (Fig. 4). We can see that if \( T_a, \neg T_b \) and \( D \) are true in the valuation \( \vartheta \) then the formula \( X \) must be true, too (third row). To get the solution we need to choose exactly one “1” from each row. If one row doesn’t contain “1”, then the original formula cannot be valid, so there is no solution. If there are more “1” in one row, then we can choose arbitrarily, and we have several solutions. If we choose the ones marked with ‘ then we get back our solution of Example 3.

The synthesis of the unknown formula is similar as before, we need to take the ones we choosed and are in the column where the unknown formula has value 1 and construct a d.n.f. where each conjunction corresponds to the values of the atomic variables.

By choosing the denoted ones, the formula \( X \) will be

\[
(T_a \land T_b \land \neg D) \lor (T_a \land \neg T_b \land D) \lor (\neg T_a \land T_b \land \neg D) \lor (\neg T_a \land \neg T_b \land D).
\]
Let us examine a puzzle ([4, puzzle 173]) where there are more unknown variables!

**Example 5.** The converse of a statement “If $P$ then $Q$” is the statement “If $Q$ then $P$”. Now, there exist two statements $X$, $Y$, which are converses of each other and such that:

1) Neither statement is deducible from the other.
2) If a Transylvanian makes either one of the statements, it follows that the other one must be true.

Can you supply two such statement?

We prefer the unrelated unknown formulae, so we shall denote by $X$ and $Y$ the antecedent and the succedent part of the unknown sentences. We cannot express the Restriction 1 by a logical law, so we solve the puzzle without it, and we shall check afterwards whether the solution fulfills this restriction or not.

\[
\begin{array}{|c|c|c|c|}
\hline
& (C_a(X \supset Y) \supset (Y \supset X)) & (C_a(Y \supset X) \supset (X \supset Y)) \\
\hline
(X = 1) & (Y = 1) & (Y = 0) & (X = 0) \\
\hline
V_a & M_a & 1' & 0 & 0 & 1 \\
\hline
\neg V_a & M_a & 1 & 1' & 1 & 1 \\
\hline
V_a & \neg M_a & 1 & 1 & 1' & 1 \\
\hline
\neg V_a & \neg M_a & 1 & 0 & 0 & 1' \\
\hline
\end{array}
\]

The are 2, 4, 4 and 2 “1” in the rows so we have $2 \times 4 \times 4 \times 2 = 64$ solution candidates. We need to check whether they fulfills the Restriction 1 or not. We left it to the reader to check them but we marked one solution. To get the formula $X$ we need to collect the marked numbers from the columns where $X = 1$, they are in the first and the second row. Hence $X \sim (V_a \land M_a) \lor (\neg V_a \land \neg M_a) \sim M_a$. To get the formula $Y$ we need to collect the marked numbers from the columns where $Y = 1$, they are in the first and the third row. Hence $Y \sim (V_a \land M_a) \lor (V_a \land \neg M_a) \sim V_a$. Therefore our solution is the following: “If I am a vampire then I am sane.” and “If I am sane then I am a vampire.”

The Puzzle 125 is a different type puzzle, and it is hard enough to work with it:

**Example 6.** Another time I was visiting a different island of knights, knaves and normals. I was rumored that there was gold on the island, and I wanted to find out whether there was. The King of the island, who was knight, graciously introduced me to three of the natives, $A$, $B$ and $C$, and told me, that at most one of them was normal. I was allowed to ask two yes-no questions to whichever ones I wished.

Is there a way to find out in two question whether there is gold on the island?

If somebody gets here in the book he knows that the normals are totally unreliable. Hence with the first question we need to find a non-normal native, and ask him. Our method isn’t so clever, we don’t use any database about solved puzzles, just the hypotheses of the current puzzle.

When we have more than one question it is very important who are the person we ask. This puzzle is symmetric, so without loss of generality at first we ask the person
A. Based on his answer we need to choose the second person (maybe \(A\) again), and ask the second question. Attention, if the first answer is “yes” the second question can be different when the first answer is “no”. If the first answer is “yes” we can ask secondly any of them, and the same holds for the answer “no”, therefore we have nine cases to test.

With our first method we can find out, that only one question isn’t enough to solve the puzzle. So with the second question we can decide (if there exists solution of this puzzle) that there is gold on the island, or not. Let be the second question such that, the answer will be yes iff there is gold on the island. Let \(X\) denote the formula of the first question, and abbreviate formula \((T_u \land X) \lor (F_u \land \neg X) \lor N_u\) with \(C_uX\). If the first answer was “yes” then the formula \(C_AX\) is true. If the second answer (for the question with formula \(Y\)) is “yes” then there is gold on the island \((G)\). This can be expressed as \(C_AX \land C_Y \supset G\). If the second answer is “no” then there is no gold on the island: \(C_AX \land C_\neg Y \supset \neg G\). In a similar way we can express that after the first “no” answer we get “yes” or “no” answer, but in this formulas we need to use \(Z\) instead of \(Y\) because the question can be different. To get the whole formula of the puzzle we need to add formula

\[(\neg N_A \land \neg N_B) \lor (\neg N_B \land \neg N_C) \lor (\neg N_C \land \neg N_A)\]

as hypothesis, that is maximum one of them is a normal.

Similarly we need to add Formula 1 for each person as hypothesis. We shall abbreviate the conjunction of hypotheses with \(M\). So the formula of the puzzle after some simplification is the following:

\[(M \land C_AX \supset (C_uY \equiv G)) \land (M \land C_{\neg X} \supset (C_vZ \equiv G)).\]

We must replace \(u\) and \(v\) with \(A\), \(B\) or \(C\). We left again to the reader to check all the nine cases, and build a table for each case. (The simplest truth-table has 54 rows and 8 columns.) If \(u = B\) and \(v = C\), then there is no row without “1” in the truth-table, so if \(A\) answered “yes” then we need to ask \(B\), else \(C\). We are interested of course in the questions, too. We have no restriction, so at arbitrary choosing we get solution, but maybe an extremely complicated. The simplest one is the following solution:

\[
\begin{align*}
X \sim T_A &\equiv N_c: “Are you knight iff \(C\) is a normal?” \\
Y \sim T_B &\equiv G: “Are you knight iff there is gold on this island?” \\
Z \sim T_C &\equiv G: “Are you knight iff there is gold on this island?”
\end{align*}
\]

We left to reader to check the solution.

6 Future

As we have seen usually there are several solutions of a puzzle. Sometimes it is enough to find only one solution, preferably the simplest one. In the future we want to find a method to search such solutions. If we have only one unknown variable then some generalization of the Karnough-map would be enough. Another research direction could be the automated translation of puzzles (from a very restricted ‘puzzle English’) to formulae and translation back of the solutions.
7 Acknowledgements

I would like to thank Fabio Massacci and Andreas Herzig for their help, and for trying valiantly to make this a more readable paper.

References