A total probability theorem for non-additive probabilities

Fabio Cuzzolin [a]

[a] *Perception group*
INRIA Rhône-Alpes
655, avenue de l'Europe
38334 SAINT ISMIER CEDEX, France

Abstract

In this paper...

Key words:

1 Introduction

Since the theory of evidence is relatively young, a large number of well known tools present in the classical theory are still missing, partly due to the tendency of many researcher of focusing more on philosophical debates about the intrinsic nature of probability rather than developing the mathematical potentials of imprecise probabilities in general and belief functions in particular.

However, it often happens that the need for the solution of interesting theoretical problems only arise after dealing with difficult applications, as we have seen when discussing the motivations of the geometric approach to the ToE. Of course, since we are more familiar with computer vision applications, we have often taken hints from vision problems.

In many situations, for instance, clouds of points are detected by a camera, generating a sequence of images of this cloud in each instant of its evolution. This points are often *unlabelled*, i.e. they cannot be distinguished (think about moving aircrafts on a controller's screen). In this context, the *data association*

Email address: `Fabio.Cuzzolin@inrialpes.fr` (Fabio Cuzzolin).
URL: `www.cs.ucla.edu/~cuzzolin` (Fabio Cuzzolin).

problem [?] consists on reconstructing the association between moving points appearing in consecutive images of the sequence.
This task is very well studied. The more popular approach is based on a battery of Kalman filters, each one associated to a single point (target), whose position is assumed independent. However, when additional information is available it is natural to search for a formalism in which all the data can be combined to achieve the best estimate of the targets' position.
When one supposes the targets to belong to an articulated body of known topological model, then the rigid motion constraint can be exploited improve the robustness of the estimation process. In Chapter **??** we will see a framework based on the theory of evidence, in which both Kalman estimates an rigid motion constraints concur to form the predicted position.

Since this constraint must be expressed in a conditional way, the task of combining *conditional belief functions* (see Chapter **??**) in a filtering-like process has to be addressed.
Both these questions, as Chapter **??**'s review has shown, still require a satisfactory solution. The second, in particular, has not even been given a correct formulation.

2 Combining conditional functions

Let us now abstract from the data association problem, and state the conditions the overall function must obey, given a set of conditional functions $s_i : 2^{\Pi_i} \rightarrow [0,1]$ over N of the elements Π_i of a partition $\Pi = \{\Pi_1, ..., \Pi_{|\Omega|}\}$ of a frame Θ induced by a coarsening Ω.

(1) *a-priori constraint*: the restriction of the candidate total function s must coincide to an *a-priori* b.f. s_0 defined on a coarsening Ω of the frame Θ;

in the data association problem the *a-priori* constraint is the belief function representing the estimate of the past association $M \leftrightarrow m(k-1)$, defined over Θ_k^{k-1} (see Figure **??** again). It ensures that the total function is compatible with the last available estimate.

(2) *Conditional constraint*: the restriction $s \oplus s_{\Pi_i}$ of the total function s to each sub-domain must coincide with the corresponding conditional belief function s_i

$$s \oplus s_{\Pi_i} = s_i \ \forall i = 1, ..., N$$

where

$$m_{\Pi_i}(A) = \begin{cases} 1 \;\; A = \Pi_i \\ \\ 0 \;\; A \neq \Pi_i \end{cases}$$

By exploiting the *a-priori* constraint, we can state an interesting condition over the focal elements of s.

Lemma 1 *The inner reduction $\bar{\theta}(e_{(.)})$ of every focal element of the candidate total belief function s must be a focal element of s_0. In other words, there must exist a focal element $E_k \in \mathcal{E}_{s_0}$ of the* a-priori *function such that $e_{(.)}$ is a subset of $\omega(E_k)$ and all the projections $\omega(\sigma)$, $\sigma \in E_k$, have non-empty intersection with $e_{(.)}$:*

$$\forall\, e_{(.)} \in \mathcal{E}_s \;\exists\, E_k \in \mathcal{E}_{s_0} \quad s.t. \quad e_{(.)} \subset \omega(\mathcal{E}_k) \quad \bigwedge \quad e_{(.)} \cap \omega(\sigma) \neq \emptyset \;\forall\, \sigma \in E_k.$$

The proof can be found in [1].
On the other hand, the conditional constraint gives the structure the candidate f.e. must respect. If we denote with $e^k_{(.)}$ any candidate f.e. included in $\omega(E_k)$,

Lemma 2 *Every focal element $e^k_{(.)}$ of the candidate total belief function s is the union of* exactly one *focal element of each conditional function whose domain Π_i is included in $\omega(E_k)$, where E_k is the smallest focal elements of s_0 s.t. $e^k_{(.)} \subset \omega(E_k)$. Namely,*

$$e^k_{(.)} = \bigcup_{\Pi_i \subset \omega(E_k)} e^i_{j_i}$$

where $e^i_{j_i} \in \mathcal{E}_{s_i} \forall\, i$.

Since $s \oplus s_{\Pi_i} = s_i$, with $m_{\Pi_i}(\Pi_i) = 1$, from the nature of Dempster's rule it necessarily follows that

$$e^k_{(.)} \cap \Pi_i = e^i_{j_i}$$

for some focal element $e^i_{j_i}$ of s_i. Furthermore $e^k_{(.)} \cap \Pi_i$ must be non-empty, for if there exists $e^k_{(.)} \cap \Pi_l = \emptyset$ for some Π_l then

$$\bar{\theta}(e^k_{(.)}) \subsetneq E_k$$

that is a contradiction.

This result has a fundamental consequence.

Lemma 3 *The minimum number of focal elements of the total function s belonging to $\omega(E_k)$ is*

$$n = \sum_{i=1,\ldots,|E_k|} (n_i - 1) + 1$$

where $n_i = |\mathcal{C}_{s_i}|$.

Called a_i^j the j-th focal element of s_i, the focal elements of s must satisfy for each s_i the $n_i - 1$ constraints

$$(\sum_{e^k_{(.)} \supset a_i^j} m_s(e^k_{(.)})) = \frac{m_{s_i}(a_i^j)}{m_{s_i}(a_i^1)} \cdot (\sum_{e^k_{(.)} \supset a_i^1} m_s(e^k_{(.)}))$$

Adding the normalization constraint we have the thesis.

3 The total belief theorem

We are now ready to formulate the analogous of the total probability theorem in the theory of evidence.

Theorem 1 *Suppose* Θ *and* Ω *are two frames of discernment, and* $\omega : 2^\Omega \rightarrow 2^\Theta$ *a refining. Let* s_0 *be a belief function defined over* Ω *and* $\sigma_1, ..., \sigma_N$ *the elements of the core* $\mathcal{C}_{s_0}$ *of* s_0*, and there exists a collection of n belief functions* s_i *defined over* Θ *whose cores* $\mathcal{C}_i$ *are included respectively into* $\Pi_i = \omega(\{\sigma_i\})$*, where* $\Pi = \{\Pi_1, ..., \Pi_{|\Omega|}\}$ *is the partition of* Θ *induced by the coarsening* Ω*. Then there exists a belief function over* $s : 2^\Theta \rightarrow [0, 1]$ *such that:*

(1) s_0 *is the restriction of* s *to* Ω, $s_0 = s|_\Omega$ *(see Equation* **??**, *Chapter* **??***);*

(2) the belief function obtained by combining s *with*

$$s_{\Pi_i} : m(\Pi_i) = 1, m(A) = 0 \; for \; A \neq \Pi_i$$

coincide with s_i *for all i:* $s \oplus s_{\Pi_i} = s_i \; \forall i = 1, ..., N$

with the minimum number of focal elements

$$n = \sum_{i=1,...,N} (n_i - 1) + 1,$$

where $n_i = |\mathcal{E}_i|$ *is the number of focal elements of* s_i.

If the belief function s_0 plays the role of *a-priori probability* (i.e. s_0 is *Bayesian*), the solution of this problem is trivial. The collection $\mathcal{E}_s$ of focal elements of s will be the union of all the collections of focal elements of the family $\{s_i\}$,

$$\mathcal{E}_s = \bigcup_{i=1}^{|\mathcal{C}_{s_0}|} \mathcal{E}_i$$

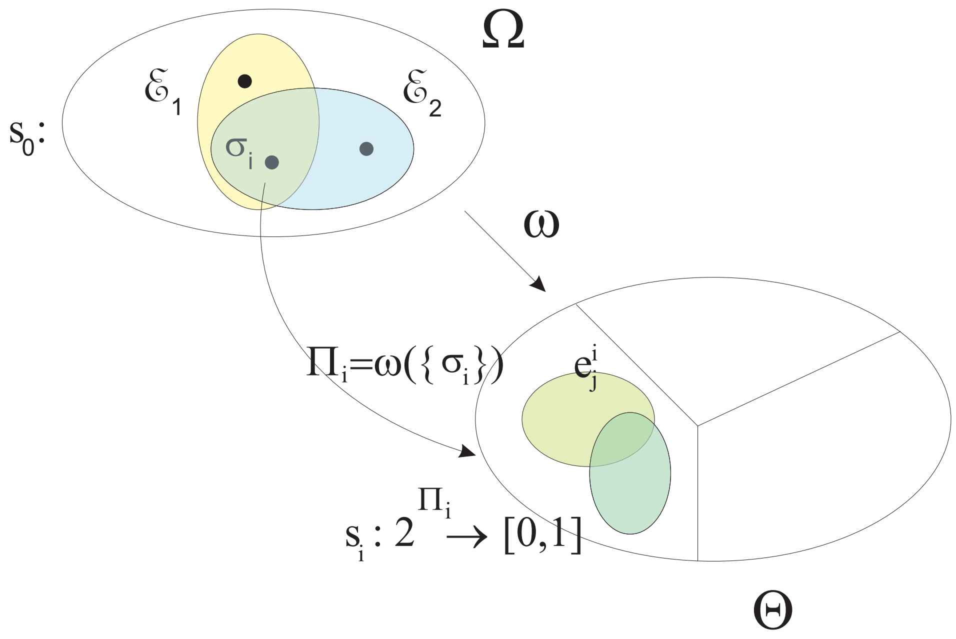

Fig. 1. Pictorial representation of the total belief theorem hypotheses.

with basic probability assignment normalized by the value $m_{s_0}(\sigma_i)$ associated to the element of Ω $\sigma_i \in \mathcal{C}_{s_0}$ "covering" s_i ($\mathcal{C}_i \subset \omega(\{\sigma_i\}) = \Pi_i$):

$$m_s(A) = m_{s_i}(A) * m_{s_0}(\sigma_i), \quad A \subset \Pi_i.$$

3.1 The restricted total belief theorem

If we force the *a-priori* function s_0 to have only *disjoint* focal elements (i.e. s_0 is the vacuous extension of a Bayesian function defined on some coarsening of Ω), we have what we call the *restricted total belief theorem*. This is the case of the data association problem, when the *a-priori* belief function is usually a simple support function with core containing a few elements.
In this situation, it suffices to solve $\mathcal{K} = |\mathcal{E}_{s_0}|$ simplified problems by considering each focal element E_k of s_0 *separately*, and then combine the sub-solutions by simply normalizing the resulting basic probability assignments by means of $m_0(E_k)$.

It is easy to see that:

Theorem 2 *The above procedure produces an optimal solution if and only if the* a-priori *function has no intersecting focal elements, otherwise the solution is* not optimal.

The number of constraints of the general problem (see Figure 1) is

$$n_g = \sum_{i=1,...,N} (n_i - 1) + \mathcal{K};$$

on the other side, by combining the $\mathcal{K}$ subproblems related to each $E_k \in \mathcal{E}_{s_0}$ we have

$$n_r = \sum_{k=1}^{\mathcal{K}} \sum_{i=1}^{N_k} (n_i - 1) + \mathcal{K} = \sum_{i=1}^{N} \mu_i (n_i - 1) + \mathcal{K}$$

where $\mu_i \geq 1$ is the *multiplicity* of σ_i, i.e. the number of focal elements E_k of s_0 such that $\sigma_i \in E_k$.
Hence $n_r \geq n_g$ and

$$n_r = n_g \;\Leftrightarrow \mu_i = 1 \;\forall\; i = 1, ..., N$$

i.e the focal elements E_k are disjoint.

4 Solution systems and transformable columns

The task of finding a suitable candidate for the total belief theorem (1) transforms into a linear algebra problem, given we are in the special case.

4.1 Solution systems

Let N be the number of elements of a given focal element E of s_0. The proof of Lemma 3 teaches us that a solution of the problem corresponds to a linear system with n equations and n variables

$$\Sigma: \quad Ax = b \tag{1}$$

where each column a_j of A represents a focal element e_j of the candidate total belief function, $x = [m_s(e_1)...m_s(e_n)]$ and n is the number of constrains given by the N belief function to combine. The first $n-1$ equations of the system have the following form

$$(\sum_{e_{(.)} \supset a_i^j} m_s(e_{(.)})) = \frac{m_{s_i}(a_i^j)}{m_{s_i}(a_i^1)} \cdot (\sum_{e_{(.)} \supset a_i^1} m_s(e_{(.)}))$$

(where a_i^j is again the j-th focal element of s_i), or alternatively

$$m_{s_i}(a_i^1) \cdot \sum_{e_{(.)} \supset a_i^j} m_s(e_{(.)}) - m_{s_i}(a_i^j) \cdot \sum_{e_{(.)} \supset a_i^1} m_s(e_{(.)}) = 0, \tag{2}$$

while the last row of the system is simply the normalization constraint

$$\sum_{e_{(.)}} m_s(e_{(.)}) = 1.$$

Hence, the problem can be reformulated. We have to choose n focal elements in the collection of the unions of exactly one focal element for each conditional function s_i,

$$\{e_{(.)} = \bigcup_{i=1}^{N} a_i^j,\ j \in [1, ..., n_i]\} \tag{3}$$

to form a matrix A of the linear system (1) such that the unique solution of the system has all positive components. For simplicity, these focal elements can be denoted by the formal vector of their components:

$$e_{(.)} = [a_1^{j_1} a_2^{j_2} ... a_N^{j_N}]'.$$

4.2 Solution space and transformable columns

If we *fix the instance* of the problem, each candidate solution system is associated to a vector $y_\sigma \in \Re^n$, the solution of the corresponding linear system:

$$\begin{aligned} \bar{x} : \Sigma &\longrightarrow \quad \mathcal{Y} \subset \Re^n \\ \sigma &\mapsto \ y_\sigma = A_\sigma^{-1}(\bar{x}) \cdot b \end{aligned}$$

we call $\mathcal{Y}$ *solution space.*

Now, let us consider an arbitrary solution system, obtained by choosing n random elements in the collection (3). The kind of columns that can be associated to *negative* components of the solution is easily detectable: since the columns of the system must have positive sum for every focal element $a_i^{j_i}$ of s_i (from the constraints (2) expressed by the rows of the system), a critical column e_k must have *at least one companion for every* i, i.e.

$$\forall i = 1, ..., N \quad \exists e_{l_i}\ s.t.\ e_k(i) = e_{l_i}(i).$$

In other words, a companion coincides to e_k over Π_i.
To this types of columns, that we call *transformable*, we can apply the following transformation:

Definition 1 *The class $\mathcal{T}$ of transformations acts on a transformable column*

e of a solution system by means of the formal sum

$$e \mapsto e' = -e + \sum_{i \in \mathcal{C}} e_i - \sum_{j \in \mathcal{S}} e_j \tag{4}$$

where $\mathcal{C}$ is a covering set of companions of e (i.e. at least one of them cover every component of e, $|\mathcal{C}| < N$) and the selection *columns $\mathcal{S}$, $|\mathcal{S}| = |\mathcal{C}| - 2$, are used to eliminate $|\mathcal{C}| - 2$ addenda from the formal sum and yield an admissible formal column.*

We call the elements of $\mathcal{T}$ column substitutions*.*

4.2.1 Effect of a transformation on a point of the solution space

A sequence of transitions can then be seen as a discrete path in the solution space: the values of the solution components associated to each column vary, and in a predictable way. In fact, called $s < 0$ the solution component of the old column e,

(1) the new column e' has solution component $-s > 0$;
(2) each companion column receives as new value $p' = p + s$, where p is the old one;
(3) each selection column has as new value $n' = n - s$, where n is the old one;
(4) the other columns maintain the old value of their solution components.

If we choose the column with the most negative solution component, the overall effect is that the most negative component is changed into a positive one, components associated to selection columns are increased, while even if one or more companion columns can receive a negative solution component, these will have an absolute value smaller then $|e|$, since $p > 0$. Hence

Proposition 1 *Column transformations of the class $\mathcal{T}$ reduce the absolute value of the most negative solution component, i.e. if Σ is transformed into Σ' by the substitution $e_{\max} \mapsto \mathcal{T}(e_{\max})$, then*

$$|e'_{\max}| < |e_{\max}|.$$

The *length of a transition* in the solution space can be easily computed from the above remarks, getting

$$\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} = \sqrt{4s^2 + \sum_{i\in\mathcal{C}} s^2 + \sum_{i\in\mathcal{S}} s^2} = \sqrt{s^2(4 + |C| + |S|)}$$

$$= \sqrt{s^2(4 + |S| + |S| + 2)} = \sqrt{2}|s| \cdot \sqrt{|S| + 3}.$$

Remark. Simple counterexamples show that the shortest path to an admissible system *is not necessarily composed by longest (greedy) steps.* This means that algorithms based on greedy choices or dynamic programming cannot work, for the problem does not have the *optimal substructure* property.

The length of the longest transition in the solution space, corresponding to choosing the most negative variable t and all the selection columns positive, can be computed in the same way. We get

$$d_{s'} - d_s = \sum_{i=1} |C|x_i(x_i + 2t) + (|C| - 1)t^2, \;\; 2 \leq |C| \leq N.$$

4.3 Hint of an existence proof

We are finally able to draw the sketch of an existence proof for the restricted total belief theorem, thanks to the effects of transformations of type $\mathcal{T}$ on the solution components. In fact,

(1) at each transformation $|e_{max}|$ *decreases*;
(2) by transforming the most negative variable we always obtain *different* systems, for each transformed column receive a positive solution component, that *cannot be changed back to a negative one* by applying the $\mathcal{T}$-class transformations;
(3) the number of solution systems is obviously finite, hence the procedure must terminate.

Now, if every column with companions on every focal element of s_i could be $\mathcal{T}$-transformable the procedure could not terminate at a system with negative variables, for they have one or more positive companions on each component i.
Unfortunately, there are "transformable" columns that do not admit a transformation of the class (4), for even if they have companions on every Π_i there not exists a collection of selection columns.

5 Solution graphs

Let us define an *adjacency* relation between solution systems. We say that $\Sigma_i \sim \Sigma_j$ if Σ_i can be obtained from Σ_j by substituting a column by means of transformation (4) (and vice-versa). This allows us to rearrange the solution systems related to a problem of a given size $\{n_i,\ i = 1, ..., N\}$ in a *solution graph.*

5.1 Significant examples of solution graphs

Let us see some significant examples, and infer some properties of the general structure of a solution graph.

5.1.1 3x2 graph

Figure 2 shows the solution graph formed by all the candidate solution systems for the problem of size $N = 2$, $n_1 = 3$, $n_2 = 2$. The twelve possible solution systems can be arranged in a matrix, whose rows and columns are labeled respectively by the numbers $(i_1, ..., i_{n_i})$ of focal elements (columns) of the associate candidate total function containing each of the n_i focal elements of s_i.
These solution systems can be arranged in two classes, according to their "external" behavior, i.e. the number of transformable columns and the number of admissible $\mathcal{T}$ transformations (edges) for each of these columns. Type II systems (central row) have two transformable columns, each admitting only one $\mathcal{T}$ transformations, while type I systems (top and bottom rows) have only one t.c. that can be replaced in two different ways.

It is interesting to note that the graph of Figure 2 can be rearranged to form a *chain* of solution systems: its edges form a unique, closed loop. The chain is composed by "rings" whose central node is a type I system connected to a couple of type II systems. Two consecutive rings are linked by a type II system.

5.1.2 2x2x2 graph

Figure 3, instead, shows a more complicated example, where the overall symmetry of the graph, and the influence of the position within the 3D matrix emerge.

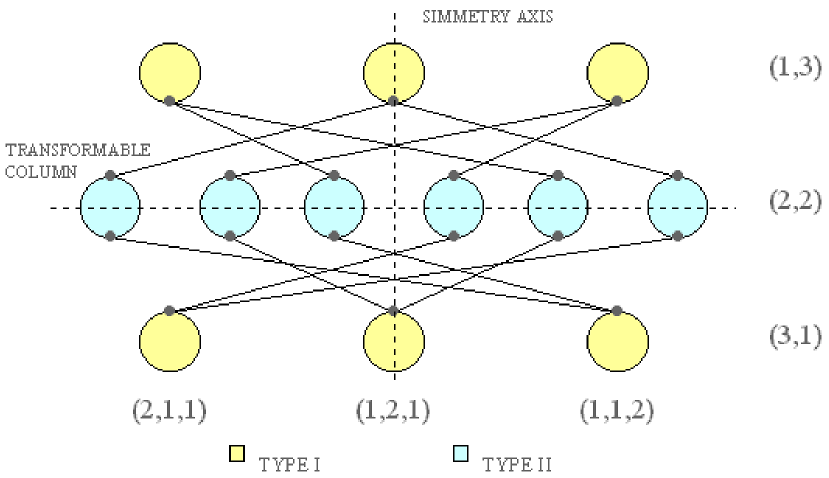

Fig. 2. Solution graph related to the problem of size $n_1 = 3$, $n_2 = 2$.

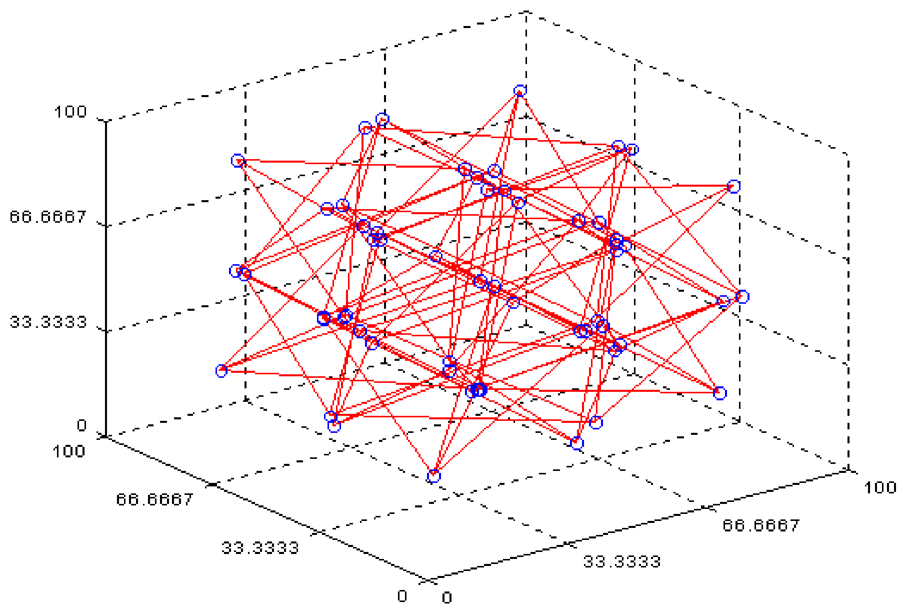

Fig. 3. Solution graph related to the problem $N = 3$, $n_1 = n_2 = n_3 = 2$.

5.2 Symmetries and solution systems

Its position in the hypermatrix influences the properties of a solution system, even if it can be shown that in the general case the same entry can contain systems of different type. Hence, the type of a solution system must be induced by some other global property of the graph.

Let us call

$$G = S(n_1) \cdot \dots \cdot S(n_N)$$

the group of the *permutations of focal elements* of the conditional belief functions s_i. The study of several solution graphs has made us formulate the following conjecture.

Conjecture. The orbits of G coincide with the types of solution systems.

This is reasonable, for the behavior of a system in terms of transformable columns depends only on the number and size of the groups containing each focal element of s_i, and not on what is the particular e.f. assigned to each group. The orbits of G are disjoint (from group theory), so they form a partition of the set of graph nodes, as it should be if they represent solution system types.

6 Future work

Even if the idea of transformable column seems to point towards the right direction, we still do not have a complete proof for the restricted total belief theorem. If our guesses are correct, we need to find a more general class of linear transformations $\mathcal{T}'$, which is applicable to *any* column with negative solution component.
The algorithm of Section 4.3 can be interpreted as proof of existence of an optimal path within a graph: chosen an arbitrary node Σ of the graph, there exists *at least* a path to another node corresponding to a system with positive solution. An investigation of the properties of such optimal paths will be necessary, together with a global analysis of the structure of our solution graphs and their mutual relationships.
In fact, intuition suggest that each graph must contain a number of "copies" of solution graphs related to lower size problems. This is confirmed by the fact that, for instance, the graph for the problem $N = 2$, $n_1 = 4$, $n_2 = 2$ is composed by 32 systems arranged in 8 chains resembling the graph of Figure 2: each system of the graph is covered by 3 of these chains. That could be extremely useful for the computation of the number of solutions of the total belief theorem, by exploiting the superposition of optimal paths for lower size graphs.

On the other side, our conjecture about the action of the group of permutations G must be proved, and its relation with the global symmetry of the graph investigated.
Our opinion is that the results obtained for the restricted total belief problem could be useful to infer a line of investigation for the solution of the general issue.

References

[1] Glenn Shafer, *A mathematical theory of evidence*, Princeton University Press, 1976.