



HAL
open science

Dynamics of trust building: models of information cross-checking in a multivalued logic framework

Adrien Revault d'Allonnes, Marie-Jeanne Lesot

► **To cite this version:**

Adrien Revault d'Allonnes, Marie-Jeanne Lesot. Dynamics of trust building: models of information cross-checking in a multivalued logic framework. The 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2015), Aug 2015, Istanbul, Turkey. hal-01213315

HAL Id: hal-01213315

<https://hal.science/hal-01213315v1>

Submitted on 18 Oct 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Dynamics of Trust Building: Models of Information Cross-Checking in a Multivalued Logic Framework

Adrien Revault d'Allonnes
Université Paris 8, EA 4383, LIASD,
FR-93526, Saint-Denis, France
Email: Allonnes@ai.univ-paris8.fr

Marie-Jeanne Lesot
Sorbonne Universités, LIP6,
UPMC Univ Paris 06, CNRS, UMR 7606
F-75005, Paris, France
Email: Marie-Jeanne.Lesot@lip6.fr

Abstract—Information cross-checking is an essential step of the trust building process that grants it its dynamics: it assesses the credibility dimension, finding confirmations or invalidations that respectively increase or weaken the current trust level of a considered piece of information and whose order influences its final value. This paper proposes a model of credibility integration that realistically takes into account even dubious confirmations and invalidations, allowing to represent a wide range of dynamic credulity stances when faced with contradictory information streams. It is formalised in an extended multivalued logic framework and illustrated with several examples to highlight the variety of behaviours it captures.

I. INTRODUCTION

Information evaluation ranges from fact-checking to source verification and spans various other topics along the way, among which information quality, search query ranking and algorithm performance analysis are but a few. One of the common objectives of these domains, however, is to determine how much faith to put into a piece of information [1]–[7]. One way of looking at this problem is to focus on trust building, the process through which belief is garnered. Numerous criteria can be considered in this task [8], such as characteristics of the source providing the considered piece of information, e.g. reliability, competence or sincerity, or of the contents, e.g. plausibility with respect to the rater's background knowledge or credibility, understood here as cross-validation by other sources.

This crucial credibility dimension usually corresponds to the final step of the trust building process and aims at cross-checking the considered piece of information, i.e. finding confirmations or invalidations by other sources [1], [4], [6], possibly independently from the actual truth of the underlying fact. Such *homologous information* lead the rater to update his current trust level: confirmations tend to increase it whereas invalidations lower it.

Trust building is linked to the art of persuasion or Aristotle's rhetoric. One particular aspect of interest in this system is the arrangement of arguments and its influence on persuasion. Indeed, the dynamics of trust building result from cross-checking: two confirmations followed by an invalidation may lead to a different level of trust than a confirmation followed by a contradiction and another confirmation might. This dependence on the order of the information stream is the source of many a theory on argumentation, some of which study the ideal moment to contradict an opponent in a political debate or in court.

Trust building in general and sensitivity to argumentation organisation are, of course, highly subjective processes. As such, different attitudes when confronted to an information stream may be observed. One rater may, for instance, be much more sensitive to contradictions than to confirmations, thus exhibiting a distrustful behaviour. Another may not favour invalidations over corroborations but be more unwilling to change his mind, showing a more mulish disposition. An efficient trust building model should be able to represent these different attitudes, called *credulity stances* hereafter.

This paper extends the multivalued logic formalisation of the trust building process presented in [7] by focusing on the influence of identified homologous information on the ensuing level of trust. It proposes a generic approach that makes it possible to exploit all available pieces of information, taking into account even dubious ones. It also discusses variants, thus offering a rich modelling framework which allows the representation of a wide range of realistic credulity stances, when faced with contradictory information stream.

The paper is organised as follows: Section II reviews existing models of credibility and, in particular, details the main features of the approach based on a multivalued logic framework. Section III describes the proposed extension of this approach, called cumulative credibility, and Section IV discusses some variants that further enrich its modelling capacities. Section V illustrates the richness of the proposed models and the variety of credulity stances they allow to represent, commenting on their respective properties.

II. EXISTING CREDIBILITY MODELS

This section discusses the notion of credibility, which occurs in several existing information scoring models, detailing and enriching with further comments the approach based on an extended multivalued logic that this paper proposes to extend.

A. Credibility Definition and Integration

In the framework of information scoring and trust building, credibility is understood as a degree of confirmation, resulting from an information cross-checking step [1], [4], [6]: this step aims at finding possible corroboration, identifying complementary information backing up or undermining a considered piece of information to be rated. The trust put in it is indeed influenced by these possible confirmations and invalidations.

The importance of cross-checking for information evaluation is underlined from the original military model [1]. Indeed,

it scores any piece of information as a bi-gram, defined as the concatenation of two symbols measured on two discrete graded scales associated to linguistic labels. The first symbol, called reliability, characterises the source, the second one, called credibility, the information content. If this credibility dimension can be estimated, it is measured on a 5-level scale, associated to the linguistic labels *improbable*, *doubtful*, *possibly true*, *probably true*, *confirmed by other sources* [1]. These labels mainly describe the information certainty; however the detailed description they are accompanied by relate them to a degree of confirmation of the considered piece of information, as also indicated by the maximal degree ‘confirmed by other sources’. This underlines the crucial role of the cross-checking step in the evaluation of a piece of information: its result is directly represented in the final global score. However, this model does not define a practical procedure to measure credibility, leaving it to the subjectivity of the rater. A fortiori, the model gives no indication regarding the effect of confirmation/invalidation order on the trust level evolution and its dynamics.

The information scoring model proposed in [4], formalised in possibility theory, aims at estimating whether a considered event occurred or not, based on the evaluation of pieces of information that confirm or invalidate it. It includes a final corroboration step that combines the possibility distributions inferred from individual pieces of information, to update the a priori event estimation taking into account the homologous information. Among other aggregation operations, a temporal effect is considered, making the model sensitive to the information order, through a weighting scheme that allows to take into account information obsolescence to reduce the effect of older and possibly out-of-date pieces of information.

The trust building model proposed in [6], [7] focuses on the process by which a rater builds his conviction on a hitherto unknown piece of information. Although it does not aim at measuring the certainty of a piece of information as the models mentioned above do, but how convinced one is by it, it shares with them the final cross-checking step that looks for corroboration to update a current trust level. Since this paper proposes to extend it, as described in Sections III and IV, the following subsections detail the formal framework it considers, its general characteristics and an implementation, enriching it with further comments.

B. Extended Multivalued Logic

The formal framework considered in [7] to represent and manipulate trust scores is an extended multivalued logic framework. This model has trust defined on a single discrete graded scale, clarified with linguistic labels, thus improving the legibility of a unique degree with a semantic interpretation.

Formally, the multivalued logic paradigm [9] that this model extends is based on a symbolic, discrete truth scale defined as a totally ordered set $\mathcal{L}_M = \{\tau_0, \dots, \tau_{M-1}\}$, where $\tau_\alpha \leq \tau_\beta \Leftrightarrow \alpha \leq \beta$. These degrees span, at varying granularity with M , the different levels of truth from τ_0 , meaning ‘false’, to τ_{M-1} , for ‘true’. The change from positive to negative truth appears around the middle value $\tau_{\frac{M-1}{2}}$, usually forced into the chosen scale by choosing an odd M . Multivalued logics are equipped with formal tools to combine the truth degrees through logical operators, that generalise conjunction,

TABLE I. EXAMPLE OF LINGUISTIC LABELS ASSOCIATED TO TRUST LEVEL AND CREDIBILITY IN \mathcal{L}_5^e

\mathcal{L}_5^e	Trust label	Credibility label
τ_4	extremely likely	totally confirmed
τ_3	likely	partially confirmed
τ_2	possible	insufficiently confirmed
τ_1	doubtful	partially contradicted
τ_0	unlikely	totally contradicted
$\tau_?$	confidence cannot be estimated	credibility cannot be estimated

disjunction or implication [9], as well as arithmetical ones, defining symbolic addition and subtraction [10].

The extended multivalued logic proposed in [7] introduces an additional degree, denoted $\tau_?$. It aims at allowing to distinguish between facts that are ‘neither true nor false’, i.e. that have a neutral truth value, and facts whose truth value cannot be evaluated. The neutral, or indifferent, value can then be mapped to $\tau_{\frac{M-1}{2}}$, as usual, whereas the unknown one can be represented by $\tau_?$, i.e. a distinct value.

Formally, $\tau_?$ is defined by the following properties [7], where \odot denotes any multivalued operator:

$$\begin{aligned} (P_1) \tau_? \notin \mathcal{L}_M & \quad (P_2) \forall \tau_\alpha \in \mathcal{L}_M, \tau_? \odot \tau_\alpha = \tau_\alpha \odot \tau_? = \tau_\alpha \\ (P_3) \neg \tau_? = \tau_? & \quad (P_4) \tau_? \odot \tau_? = \tau_? \end{aligned}$$

Property (P_1) takes $\tau_?$ out of the ordering constraints imposed on all τ_α in \mathcal{L}_M . This is justified by the fact that even a very doubtful element cannot be compared to one whose truth value is unknowable. For similar reasons, property (P_2) turns $\tau_?$ into a neutral element for all multivalued operators. Properties (P_3) and (P_4) describe the rules for combining $\tau_?$ with itself.

The extended multivalued logic is then defined as $\mathcal{L}_M^e = \mathcal{L}_M \cup \{\tau_?\}$, where the usual degree manipulation operations, both logical and arithmetical, are extended according to properties (P_2) to (P_4) to handle $\tau_?$. Characteristics and properties of \mathcal{L}_M^e are discussed further in [7].

C. Extended Multivalued Credibility Model

The model proposed in [7] defines the trust building process as the sequential integration, into an initially indeterminate level of trust, of relevant dimensions, namely reliability, competence, plausibility and credibility. It is formalised in the extended multivalued logic framework briefly described in the previous subsection: the trust level and the activation of all involved dimensions are represented and manipulated in \mathcal{L}_M^e .

Table I illustrates, in the case of \mathcal{L}_5^e , possible linguistic labels associated to trust and credibility [6]. As in the models briefly reviewed in Section II-A, credibility is understood as a degree of confirmation of the rated piece of information. As a consequence, in the case where the homologous piece of information it is compared to is an invalidation, credibility has a degree in the bottom half of the scale, $\tau < \tau_{\frac{M-1}{2}}$; if it is a confirmation, the degree is greater than $\tau_{\frac{M-1}{2}}$.

Given a current evaluation of the trust score, denoted τ_α , the integration of the credibility dimension aims at updating τ_α with respect to a homologous piece of information: the output trust level depends on the degree of confirmation it provides, i.e. activation of the credibility dimension, denoted τ_β , as well as its current evaluation, denoted τ_γ . Indeed, if a homologous

piece of information completely confirms the studied element, it obviously increases the trust level. However, if this absolute confirmation is doubtful, its effect should be less than it would be, were it nearly certain.

Therefore the cross-checking operator is formalised as a mapping $G : \mathcal{L}_M^e \times \mathcal{L}_M^e \times \mathcal{L}_M^e \rightarrow \mathcal{L}_M^e$, that satisfies the following properties [7]:

- G is increasing in its first 2 arguments
- G is:
 - increasing in its third argument if $\tau_\beta \geq \frac{\tau_{M-1}}{2}$
 - decreasing in its third argument if $\tau_\beta < \frac{\tau_{M-1}}{2}$
- $\forall \tau_\alpha, \tau_\beta, \tau_\gamma \in \mathcal{L}_M \quad G(\tau_\alpha, \tau_\beta, \tau_\gamma) \geq \tau_\alpha$ if $\tau_\beta \geq \frac{\tau_{M-1}}{2}$
 $G(\tau_\alpha, \tau_\beta, \tau_\gamma) \leq \tau_\alpha$ if $\tau_\beta < \frac{\tau_{M-1}}{2}$
- τ_γ is a neutral element for all 3 arguments

The first property corresponds to a double statement: an increase in the first argument τ_α imposes that the higher the initial trust level, the higher the updated one, whatever the correction. Similarly, for any current score, the updated score increases with how much the homologous piece of information confirms the current evidence.

The second property emanates from this two-sided scale. Indeed, for a fixed current level of trust and credibility, two cases appear. The case of a confirmation is briefly discussed above: the higher the trust level τ_γ of the compared piece of information, the higher the updated score. The opposite case of a contradiction has a similar effect, that is a certain invalidation lowers the score further than a doubtful one would.

The third property describes the respective effects of confirmations, when $\tau_\beta \geq \frac{\tau_{M-1}}{2}$, for which G shows an emphasising effect on the trust level, and invalidations which have an abating effect: for confirmations, the initial trust score is a lower bound of the updated score, whereas it is an upper bound in the case of invalidations.

The final property deals with undetermined values: if the current score τ_α is unknown, then the updated score is determined by the degree of confirmation and the trust level of the compared piece of information. If the degree of confirmation τ_β or the trust level of the compared piece of information τ_γ is unknown, the current score is not modified.

D. Implementation as Immediate Credibility

Many functions can be considered in order to satisfy the desired properties above. We propose to name *immediate credibility* the one described in [7] and briefly recalled in this section, as opposed to the model proposed in the next section, called *cumulative credibility*.

1) *Principle based on Transition Thresholds*: In the general case where $\tau_\beta \neq \tau_\gamma$, i.e. when the degree of confirmation offered by the homologous piece of information has a determined value, the cross-checking operator is based on the definition of transition thresholds, specified as parameters $\kappa_\alpha^\gamma \in \mathcal{L}_M$: they express constraints for the transition between a current score τ_α and an updated value τ_γ . More precisely, they allow to update the current score from τ_α to τ_γ if the influence of the considered homologous information is beyond κ_α^γ . The notion of influence, defined as weighted credibility, is described in the next subsection.

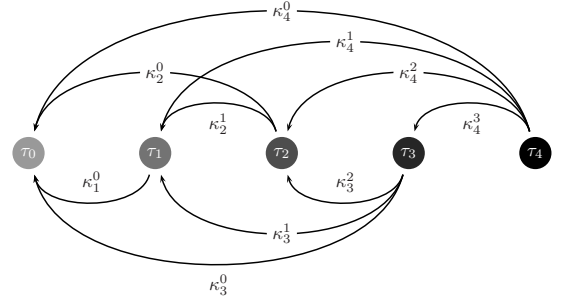


Fig. 1. Graphical representation of $\tilde{F}_c(\tau_\alpha, \tau_\beta)$ ignoring τ_γ [7]

This principle is illustrated on Figure 1 in the case of invalidation: the shaded disks represent the current level of trust in \mathcal{L}_M^e and the arrows show the transitions from a current score to an updated one. All arrows are oriented towards the left, as invalidations can only lower the trust level. A decrease of the trust level from τ_3 to τ_1 can for instance only occur if the homologous piece of information has an influence above κ_3^1 , as formalised below.

2) *Weighted Credibility*: The influence that a given homologous piece of information has on the amount of change the rated piece of information undergoes both depends on the degree of confirmation it provides, as measured by the credibility dimension, and on its own trust level. As previously stated, for a given level of corroboration, a dubious homologous piece of information should influence the trust update to a lesser extent than a highly trusted one; similarly, an absolute confirmation should have a higher impact than a partial one.

The weighted credibility operator, written $g(\tau_\beta, \tau_\gamma)$, that implements this principle is a mapping $g : \mathcal{L}_M^e \times \mathcal{L}_M^e \rightarrow \mathcal{L}_M^e$ satisfying the following properties [7]: (i) g is increasing (resp. decreasing) in its first argument for confirmations (resp. invalidations) and increasing in its second argument, (ii) τ_γ is an absorbing element for both arguments, and (iii) g is bounded above by τ_β for confirmations and below by $\neg\tau_\beta$ for invalidations. The negation operator serves to allow a symmetric behaviour.

As a consequence, a homologous piece of information that is dubious or has a low cross-checking effect has a low influence on the evolution of the trust level.

3) *Abating and Emphasising Functions*: Two functions, denoted \tilde{F}_c and \tilde{F}_i , are defined to represent the emphasising and abating influence [7], respectively required for invalidating and confirming pieces of information. Both are mappings $\mathcal{L}_M^e \times \mathcal{L}_M^e \rightarrow \mathcal{L}_M^e$ that take as arguments the score to be updated, denoted τ_α , and the weighted credibility, i.e. the influence strength, $\tau_\omega = g(\tau_\beta, \tau_\gamma)$; they output the updated score. Both are increasing in their first argument, but, due to their abating and emphasising influences, \tilde{F}_i is decreasing in the second one whereas \tilde{F}_c is increasing. Moreover, \tilde{F}_i is bounded above by its first argument, i.e. $\tilde{F}_i(\tau_\alpha, \tau_\omega) \leq \tau_\alpha$, whereas \tilde{F}_c is bounded below.

Using the transition thresholds, \tilde{F}_i and \tilde{F}_c can for instance

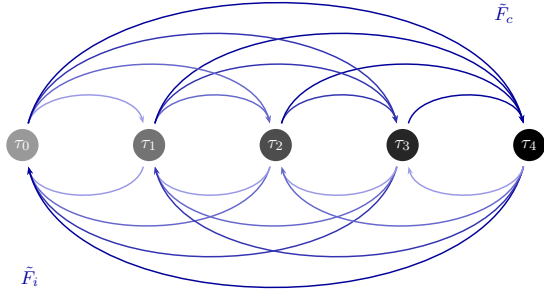


Fig. 2. Graphical representation of the immediate credibility operator G , in \mathcal{L}_5 [7]

be defined as [7]:

$$\tilde{F}_i(\tau_\alpha, \tau_\omega) = \begin{cases} \min\{\tau_\gamma \in \mathcal{L}_M | \tau_\omega \geq \kappa_\alpha^\gamma\} & \text{if } \tau_\alpha, \tau_\omega \in \mathcal{L}_M \\ \tau_\alpha & \text{if } \tau_\omega = \tau_\gamma \\ \tau_\omega & \text{if } \tau_\alpha = \tau_\gamma \end{cases}$$

$$\tilde{F}_c(\tau_\alpha, \tau_\omega) = \begin{cases} \max\{\tau_\gamma \in \mathcal{L}_M | \tau_\omega \geq \kappa_\alpha^\gamma\} & \text{if } \tau_\alpha, \tau_\omega \in \mathcal{L}_M \\ \tau_\alpha & \text{if } \tau_\omega = \tau_\gamma \\ \tau_\omega & \text{if } \tau_\alpha = \tau_\gamma \end{cases}$$

Both functions follow the most extreme transition that can be triggered by the considered influence τ_ω according to the transition thresholds, respectively in an abating or emphasising way. Also, τ_γ is a neutral value for all arguments. Figure 1 illustrates the general architecture of \tilde{F}_i omitting the cases where τ_γ .

Note that the monotonicity property of \tilde{F}_i and \tilde{F}_c imposes constraints on the parameters κ_α^γ : in the case of \tilde{F}_i for instance, they must be ordered so that if $\tau_\gamma < \tau_\delta$, then $\kappa_\alpha^\gamma \geq \kappa_\alpha^\delta$. In general, the second inequality is strict, so as to avoid multiple outcomes for one τ_ω .

4) *Immediate Credibility Operator:* The global cross-checking operator satisfying the desired behaviour stated in Section II-C is finally defined as the combination of these two functions, whose application depends whether the homologous information is a confirmation or an invalidation [7]:

$$G(\tau_\alpha, \tau_\beta, \tau_\gamma) = \begin{cases} \tau_\alpha & \text{if } \tau_\beta = \tau_\gamma \\ \tilde{F}_i(\tau_\alpha, g(\tau_\beta, \tau_\gamma)) & \text{if } \tau_\beta < \tau_{\frac{M-1}{2}} \\ \tilde{F}_c(\tau_\alpha, g(\tau_\beta, \tau_\gamma)) & \text{otherwise} \end{cases}$$

It is illustrated in Figure 2, where the lower part represents the \tilde{F}_i function, similar to the one shown on Figure 1, and the upper part \tilde{F}_c , omitting the transition thresholds for clarity's sake. Specific examples are given and commented in Section V-B.

5) *Transition Threshold Setting and Resulting Properties:* The proposed immediate credibility implementation of the cross-checking operator is defined by transition thresholds, allowing to model both conviction dynamics and a large variety of credulity stances.

Indeed, the ordering constraints on κ_α^γ are imposed for a given value of α but constraints of the form $\kappa_\alpha^\beta = \kappa_\alpha^\gamma \oplus \kappa_\gamma^\beta$, where τ_γ is an intermediate value between τ_α and τ_β and \oplus a symbolic addition operator [10], are not required. As a consequence, going directly from τ_α to τ_β need not be

equivalent to successive transitions from τ_α to τ_γ and from τ_γ to τ_β .

Moreover the definitions of the \tilde{F}_i and \tilde{F}_c functions are independent: in particular, it is not required that $\kappa_\alpha^\gamma = \kappa_\gamma^\alpha$. This property allows the immediate cross-checking operator G to represent different credulity stances, modelling different sensitivities to contradictory arguments: a dubious invalidation may for instance lower the trust level to a lesser extent than a dubious confirmation may increase it.

Besides, except in the special case where the considered transition parameters make the operators associative, the proposed function does not lead to the same result for an identical set of homologous information presented in a different order: integrating two confirmations followed by an invalidation does not necessarily provide the same result as a confirmation followed by an invalidation in turn followed by a confirmation. This property grants the proposed model the desired ability to model the dynamics of conviction.

III. CUMULATIVE CREDIBILITY

This section describes the proposed extension of the immediate cross-checking operator, offering an alternative implementation of credibility integration that satisfies the desired properties as well as a refined usage of all available homologous information.

A. Motivation

The immediate model for cross-checking given above only takes into account pieces of information whose current weighted impact exceeds transition thresholds. What this means is that any piece of information whose level of trust combined to how comparable it is is lower than these thresholds has no influence over the evaluation and the final score is the same as had they not occurred.

One may consider that this behaviour is too radical. Indeed, an abundant succession of confirmations, even if they are doubtful, may have an impact, such as that of persistent rumours. Even when what they report seems doubtful, their incessant stream of confirmations usually has an impact.

Consider, for instance, a mildly believable piece of information, with score τ_3 . Suppose it is confirmed by a succession of highly doubtful pieces of information, each of weighted credibility τ_1 . If the strategy is such that only confirmations with weighted credibilities above τ_2 have a positive influence, that is $\kappa_3^4 = \tau_2$, applying G does not increase trust, be there dozens such -possibly independent- confirmations.

Even though strategies aim at limiting drastic behaviours, it seems worthwhile to keep track of these confirmations, or invalidations, to fully exploit all information. The cumulative strategy proposed in this section aims at offering such a behaviour.

B. Trust Updating Algorithm

1) *General Idea:* The proposed method enriches operator G by allowing it to keep track of past homologous information. In order to achieve this, two accumulators, written g_+ and g_- , are introduced: the former stores the cumulative weighted

```

1: if  $\tau_\beta = \tau_\gamma$  then
2:    $\tau \leftarrow \tau$ 
3:    $g_- \leftarrow g_-$ 
4:    $g_+ \leftarrow g_+$ 
5: else if  $\tau_\beta < \tau_{\frac{M-1}{2}}$  then
6:    $g_- \leftarrow g_- \oplus g(\tau_\beta, \tau_\gamma)$ 
7:   if  $\tilde{F}_i(\tau, g_-) \neq \tau$  then
8:      $\tau \leftarrow \tilde{F}_i(\tau, g_-)$ 
9:      $g_- \leftarrow \tau$ 
10:     $g_+ \leftarrow \tau$ 
11:   end if
12: else if  $\tau_\beta \geq \tau_{\frac{M-1}{2}}$  then
13:    $g_+ \leftarrow g_+ \oplus g(\tau_\beta, \tau_\gamma)$ 
14:   if  $\tilde{F}_c(\tau, g_+) \neq \tau$  then
15:      $\tau \leftarrow \tilde{F}_c(\tau, g_+)$ 
16:      $g_- \leftarrow \tau$ 
17:      $g_+ \leftarrow \tau$ 
18:   end if
19: end if

```

Fig. 3. Cumulative cross-checking algorithm G^*

credibility of confirmations, the latter that of invalidations. Functions \tilde{F}_c and \tilde{F}_i then apply to the current score and, respectively, g_+ for confirmations and g_- for invalidations, instead of just the latest g value.

The persistence of the accumulators g_+ and g_- no longer allows credibility integration to be written as a function. It is, thus, presented below as a score updating algorithm G^* which can return the current score value at any time.

2) *Description:* The algorithm G^* , whose complete description is given in Fig. 3, is initially called on a piece of information -and its current trust score τ_α - when it encounters its first homologous counterpart. As long as new instances emerge, the level of trust τ , initialised by τ_α , is updated. At any time, the current value of τ can be returned to update τ_α .

On the initial call, the two accumulators g_+ and g_- , internal parameters of the algorithm, are initialised at τ_γ since, before the first confirmation and invalidation are observed, their value is undetermined.

For each subsequent homologous information, of credibility τ_β and trust value τ_γ , the appropriate accumulator is updated: g_- (line 6) for an invalidation, i.e. if $\tau_\beta < \tau_{\frac{M-1}{2}}$, g_+ (line 13) for a confirmation. In both cases, the weighted credibility, representing the influence the considered homologous piece of information must play, is added to the current accumulator value. If either the trust score or the degree of confirmation is still unknown, i.e. if either $\tau_\beta = \tau_\gamma$ or $\tau_\gamma = \tau_\gamma$, no internal variable (τ , g_+ or g_-) is modified: for τ_β , this is explicitly stated in lines 1 to 4; for τ_γ , it results from τ_γ being an absorbing element for g (see Section II-D2).

Any time the accumulated baggage influence the dynamics of conviction, that is when either $\tilde{F}_i(\tau, g_-)$ or $\tilde{F}_c(\tau, g_+)$ modify the current trust value (lines 7 and 14 for invalidation and confirmation respectively), τ is updated and the two accumulators are reset.

3) *Properties:* It can be shown easily that the cumulative cross-checking algorithm G^* satisfies all desired properties

described in Section II-C, namely the monotonicity constraints, the behaviour of τ_γ and the boundary conditions. The proofs, informally sketched below, rely on the properties of functions g , \tilde{F}_c and \tilde{F}_i and on that of the symbolic addition operator \oplus , which is increasing in both its arguments [10].

The fact that G^* is monotonically increasing in its first argument τ_α , the current evaluation before credibility analysis begins, comes down to it only serving as an initialisation for its internal placeholder τ which, in turn, is used either with \tilde{F}_c or \tilde{F}_i , lines 8 and 15, both increasing in their first argument.

To prove G^* 's monotonicity in its second argument, one must consider two values τ_β and τ'_β , such that $\tau_\beta \leq \tau'_\beta$. Suppose both are less than $\tau_{\frac{M-1}{2}}$, then they are parameters of g and participate in g_- 's update, line 6, and thus potentially in τ 's, line 8. Since g is decreasing in its first argument for invalidations and \oplus is increasing, this leads to the inequality $g_- \geq g'_-$. Finally, as \tilde{F}_i is decreasing in its second argument, it holds that $\tilde{F}_i(\tau, g_-) \leq \tilde{F}_i(\tau, g'_-)$. If either both or neither updates are triggered, the required increase follows. If only one update is performed, the property is justified by \tilde{F}_i 's boundary property. The case where both τ_β and τ'_β are greater than $\tau_{\frac{M-1}{2}}$ is proved similarly. Finally, if $\tau_\beta < \tau_{\frac{M-1}{2}} \leq \tau'_\beta$, the required property derives from the boundary constraints on \tilde{F}_i and \tilde{F}_c : the update value based on τ_β , defined in line 8 and denoted τ_{new} , satisfies $\tau_{new} \leq \tau$ and the one based on τ'_β , denoted τ'_{new} and defined in line 15, is such that $\tau \leq \tau'_{new}$.

The monotonicity constraint on τ_γ can be proved distinguishing by between confirmations and invalidations. If $\tau_\beta \geq \tau_{\frac{M-1}{2}}$, the increasing monotonicity in τ_γ can be proved using the fact that g , \oplus and \tilde{F}_c are increasing in their second argument. The case of invalidations is proved similarly using \tilde{F}_i 's monotonicity.

The emphasising (resp. abating) behaviour for confirmations (resp. invalidations) emanates from operator definitions. Suppose, e.g., $\tau_\beta \geq \tau_{\frac{M-1}{2}}$, the possible update for τ is defined as $\tilde{F}_c(\tau, g_+)$ (line 8) which is greater than or equal to τ according to the boundary condition on \tilde{F}_c . If τ is not updated, the inequality trivially holds. The case of invalidations is similar, using the boundary condition on \tilde{F}_i .

Finally, the neutrality of τ_γ for the first argument comes from its neutrality for the two functions \tilde{F}_c and \tilde{F}_i . For the second and third arguments, it is shown by the absence of modification $\tau_\beta = \tau_\gamma$ and $\tau_\gamma = \tau_\gamma$ lead to, discussed above.

IV. CUMULATIVE CREDIBILITY VARIANTS

This section discusses variants of the proposed cumulative cross-checking algorithm G^* , considering alternatives for accumulator re-initialisation and trust score updating tests.

A. Trust Momentum Preservation

In the cross-checking algorithm above, every time the current trust level is updated the accumulators are reset (lines 9:10 or 16:17, depending on $\tau_{\frac{M-1}{2}}$). When modelling the dynamics of trust building, such a drastic change limits expressiveness. Indeed, if, instead of having g_+ and g_- go back to the undetermined τ_γ value, some trace of the previous directions

TABLE II. ALTERNATIVE RE-INITIALISATION SCHEMES TO PRESERVE TRUST MOMENTUM, κ_α^β IS THE TRIGGERED THRESHOLD

Line #	Enthusiastic	Mistrustful	Eager sceptic
9:	$g_- \leftarrow \tau_?$	$g_- \leftarrow g_- \ominus \kappa_\alpha^\beta$	$g_- \leftarrow g_- \ominus \kappa_\alpha^\beta$
10:	$g_+ \leftarrow g_+$	$g_+ \leftarrow \tau_?$	$g_+ \leftarrow g_+$
16:	$g_- \leftarrow \tau_?$	$g_- \leftarrow g_-$	$g_- \leftarrow g_-$
17:	$g_+ \leftarrow g_+ \ominus \kappa_\alpha^\beta$	$g_+ \leftarrow \tau_?$	$g_- \leftarrow g_+ \ominus \kappa_\alpha^\beta$

were to be kept, then the trust building momentum could be preserved.

Since any time either one exceeds a threshold κ_α^β both accumulators are reset, two options exist. Suppose that the trust level increases because $g_+ \geq \kappa_\alpha^\beta$. Instead of resetting g_+ one may wish to hang on to some potential surplus confirmation history by choosing $g_+ \ominus \kappa_\alpha^\beta$ as its next value, where \ominus denotes a symbolic subtraction operator [10]. Another possibility to consider is that, even though a sufficient number of confirmations has been reached, one may wish to keep a memory of one's previous doubts. To do this, g_- may need to stay unchanged.

These options therefore offer three alternative variants to the algorithm above: an enthusiastic attitude, where only positive momentum is preserved; a mistrustful one, in which only the doubts remain; and an eager sceptic, where both traces are kept. Algorithmically they respectively correspond to replacing lines 16 and 17 with the lines indicated in the lower part of Table II.

Obviously, had the transition been reversed (i.e. $g_- \geq \kappa_\alpha^\beta$), all opposite options would have been considered, as indicated by the first two lines of the same Table II.

B. Order Independent Credibility

Even though the momentum preservation and the rest of the dynamics of trust modelling presented above increase the expressiveness of the model, there are some cases where they are not acceptable. Suppose, for instance, that one wishes to evaluate trust in the production of an automatic knowledge extraction system. Suppose further that the extraction system is non-deterministic in the order it produces knowledge. In such a case the integration of credibility can neither be immediate nor cumulative. The stability of the evaluation supersedes the benefits of expressiveness.

One way of circumventing this problem is to accumulate all available homologous information before actually updating the trust value. Explicitly, this comes down to updating g_+ and g_- , without evaluating either \tilde{F}_c nor \tilde{F}_i until all homologous information have been taken into account.

The last remaining option becomes whether to inspect invalidations or confirmations first, i.e. $\tilde{F}_c(\tilde{F}_i(\tau_\alpha, g_-), g_+)$ and $\tilde{F}_i(\tilde{F}_c(\tau_\alpha, g_+), g_-)$, either along credulity stance or relative importance of the dynamics. As long as these settings are fixed for a user, the output remains stable.

V. ILLUSTRATIVE EXAMPLE

This section illustrates the richness of the model by comparing the credulity stances of two users on a common stream of information. All evaluations (trust, credibility, weighted credibility and transition thresholds) are in \mathcal{L}_5^ϵ .

TABLE III. EXAMPLE INFORMATION STREAM

id.	τ_β	$g(\tau_\beta, \tau_\gamma)$
i_1	τ_4	τ_1
i_2	τ_0	τ_2
i_3	τ_2	τ_2
i_4	τ_1	τ_2
i_5	τ_3	τ_3
i_6	τ_0	τ_2

TABLE IV. TRANSITION THRESHOLDS FOR USER 1 OF MULISH STANCE AND A DISTRUSTFUL USER 2

User 1				User 2			
↓		↑		↓		↑	
κ_1^0	τ_3	κ_0^1	τ_3	κ_1^0	τ_2	κ_0^1	τ_4
κ_2^0	τ_4	κ_0^2	τ_4	κ_2^0	τ_3	κ_0^2	<i>und</i>
κ_2^1	τ_3	κ_0^3	<i>und</i>	κ_2^1	τ_2	κ_0^3	<i>und</i>
κ_3^0	<i>und</i>	κ_0^4	<i>und</i>	κ_3^0	τ_3	κ_0^4	<i>und</i>
κ_3^1	τ_4	κ_1^2	τ_3	κ_3^1	τ_2	κ_1^2	τ_3
κ_3^2	τ_3	κ_1^3	τ_4	κ_3^2	τ_1	κ_1^3	τ_4
κ_4^0	<i>und</i>	κ_1^4	<i>und</i>	κ_4^0	τ_4	κ_1^4	<i>und</i>
κ_4^1	<i>und</i>	κ_2^3	τ_3	κ_4^1	τ_3	κ_2^3	τ_2
κ_4^2	τ_4	κ_2^4	τ_4	κ_4^2	τ_2	κ_2^4	τ_3
κ_4^3	τ_3	κ_3^4	τ_3	κ_4^3	τ_1	κ_3^4	τ_3

This section first describes the considered example, presenting the given information stream and credulity stances. It then discusses the results obtained by three cross-checking operators: the immediate, the cumulative and the order independent ones.

A. Information Stream

The example we use is defined only in terms of the model: only evaluations are given, the actual facts and story are left out. The evaluation is that of a piece of information which, before the credulity step, is rated τ_2 , i.e. *possible*. A contradictory stream containing six homologous pieces of information is then observed, as indicated in Table III. Each comes with its credulity score, that is its confirmation degree to the original piece of information, named τ_β and its own level of trust, written τ_γ . To simplify further, Table III does not give τ_γ , only the result of its aggregation with τ_β , i.e. the weighted credulity value $g(\tau_\beta, \tau_\gamma)$.

The information stream is made of six pieces of information: the first a confirmation ($\tau_\beta \geq \tau_{\frac{M-1}{2}}$) with a low influence ($g(\tau_\beta, \tau_\gamma) = \tau_1$). The rest of stream alternates between invalidations and confirmations.

B. Considered User Credulity Stances

Table IV describes the credulity stances of the two users. For each of them, the thresholds for both abating and increasing operators, respectively \tilde{F}_i , in the columns headed with ↓, and \tilde{F}_c , in the ↑ columns, are given.

Due to the constraints imposed to κ_α^β , some transitions are impossible. For instance, for user 1, the threshold to get from τ_4 to τ_2 is $\kappa_4^2 = \tau_4$. The constraints impose that $\kappa_4^2 < \kappa_4^1$ (see Section II-D3). Since $\tau_4 = \tau_{M-1}$, the maximal possible value, trust cannot directly go from τ_4 to τ_1 . Such cases with undefined transition are represented in the table as *und*.

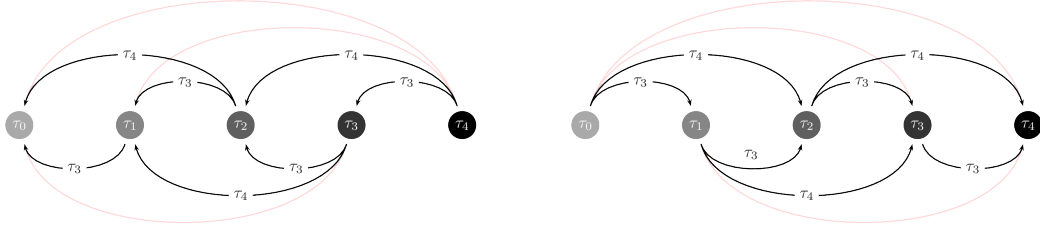


Fig. 4. Graphical representation of user 1's credulity stance, \tilde{F}_i on the left, \tilde{F}_c (right)

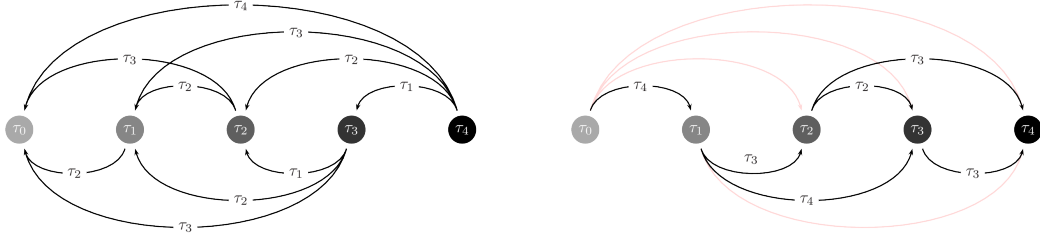


Fig. 5. Graphical representation of user 2's credulity stance, \tilde{F}_i on the left, \tilde{F}_c (right)

A more intuitive representation of the user credulity stances is provided by Fig. 4 and 5. In order to preserve the expressiveness of the overall view of the underlying operator, undefined transitions are represented, with light arrows without thresholds. The two operators must be aligned for a full definition of G .

For user 1, the transition thresholds only take large values, either τ_3 or τ_4 , making it hard for him to change his mind. Only homologous information that either provide a clear confirmation or invalidation of the considered piece of information, i.e. with a high credibility activation, or with a high trust level, can lead him to update his trust level. Because his thresholds are symmetrical, i.e. $\kappa_\alpha^\gamma = \kappa_\gamma^\alpha$, he exhibits no inclination either way. For these traits, we call him *mulish*.

For user 2, the transition thresholds are not symmetrical in their increasing and abating versions. For instance, $\kappa_3^4 = \tau_3$, whereas $\kappa_4^3 = \tau_1$. The induced conviction dynamics can be characterised as *distrustful*: the increasing thresholds are all greater than the abating ones, meaning that user 2 requests higher confirmations, as measured by the weighted credibility, to grant trust, than to withdraw it. In other words, he is more easily discouraged than he is convinced.

C. Immediate Credibility

The left columns in Table V show the evolution of the information score for both users during the progressive integration of the information stream using the immediate credibility approach. The first line corresponds to the original considered piece of information, each following line is numbered with the piece of homologous information's identifier, from Table III.

For user 1, the first four pieces of information have too low an influence and are thus insufficient to trigger a modification in his trust level. Indeed, their weighted credibility is lower than all transition thresholds starting from τ_2 , i.e. κ_2^γ for all γ . Only the confirmation i_5 reaches the minimal value, leading the user to increase his evaluation to τ_3 , i.e. *likely*.

TABLE V. SCORE EVOLUTION USING BOTH CREDIBILITY APPROACHES

id.	Immediate		Cumulative					
	User 1	User 2	User 1			User 2		
	τ	g	τ	$g+$	$g-$	τ	$g+$	$g-$
	τ_2	τ_2	τ_2	$\tau_?$	$\tau_?$	τ_2	$\tau_?$	$\tau_?$
i_1	τ_2	τ_2	τ_2	τ_1	$\tau_?$	τ_2	τ_1	$\tau_?$
i_2	τ_2	τ_1	τ_2	τ_1	τ_2	τ_1	$\tau_?$	$\tau_?$
i_3	τ_2	τ_1	τ_3	$\tau_?$	$\tau_?$	τ_1	τ_2	$\tau_?$
i_4	τ_2	τ_0	τ_3	$\tau_?$	τ_2	τ_0	$\tau_?$	$\tau_?$
i_5	τ_3	τ_0	τ_4	$\tau_?$	$\tau_?$	τ_0	τ_3	$\tau_?$
i_6	τ_3	τ_0	τ_4	$\tau_?$	τ_2	τ_0	τ_3	τ_2

User 2 is more sensitive and reacts to the invalidations i_2 and i_4 . Contrary to user 1, as he is easily put off and disinclined to trust, he does not take into account the confirmation i_5 . As a consequence, his final evaluation is the minimal value, i.e. τ_0 or *unlikely*.

D. Cumulative Credibility

The right columns in Table V show the evolution of the information score for both users during the progressive integration of the information stream with a cumulative credibility approach. Two additional columns for each user show the evolution of the accumulators.

Contrary to the immediate approach, even pieces of information with low credibility play a role in the evolution of the score. For this reason, user 1 does not wait for confirmation i_5 to update his level of trust, but starts after the second confirmation, i_3 . Since the considered implementation of the cumulative credibility resets accumulators at each evolution, the memory of the τ_2 doubts is lost. The next homologous information to have an impact being i_5 , the final score is the maximum, τ_4 , a higher score than in the case of immediate credibility. User 1 is less sceptical than previously.

User 2, more amenable to a change of heart than user 1 but harder to convince reproduces the same behaviour as in the immediate case. In order for him to exhibit a different

behaviour, either the information stream has to be reordered or a momentum preserving variant chosen.

E. Total Cumulative Credibility

In the total cumulative case, described above as order independent credibility, the decision is taken once the entire stream has been considered, that is only once all 6 pieces of information have been processed. In the example above, since the accumulators are not reinitialised, adding the weighted credibility of confirmations and invalidations respectively, they end up with the same value $g_+ = g_- = \tau_4$. Note that, since the decision is postponed and the accumulators augmented, transitions may skip more trust levels than with previous operators.

Because user 1 has a symmetrical strategy, the order in which \bar{F}_c and \bar{F}_i are applied does not influence the final result: in both cases, the output score is τ_2 , i.e. identical to the initial score, as $g_+ = g_-$.

User 2, more distrustful than user 1, is discouraged more easily than he is convinced. In the present case, where arguments in favour are as convincing as those against, the order in which they are taken into account determines his final trust level. If contradictions are integrated first, he finds the piece of information *doubtful*, with a final score of τ_1 . If, however, confirmations are considered first and invalidations revealed afterwards, he does not believe the information at all and attributes it the minimal score τ_0 .

F. Impact of the Order on Trust Building

This section highlights the influence of the order in which the information are considered on the evolution of trust, thereby underlining the dynamic component of the trust building process. The homologous pieces of information considered are the same as in the original stream, described in Table III, only their order differs. This reordering can be seen in the first column of Table VI, the other columns group all evolutions of trust, for both users and both credibility approaches.

Even though both users end with the same final evaluations, for the immediate credibility, than they did before, the reordering changes the levels of trust visited during the evaluation process. This hints at the possibility of different results if another piece of information had been inserted in the stream.

The impact of the reordering becomes clearer when looking at the cumulative credibility. Indeed, neither user then finishes with same trust level as he did before. User 1, whom we refer to as mulish, is initially favourable to this new order, reaching the maximum level earlier. However, the final increasing lot of contradictions, and the ensuing necessity for a change of mind, have a larger negative impact than ever before, leaving him thinking the piece of information just *possible*. Distrustful user 2 is less sensitive to the final contradictory closing arguments. Interestingly, he never thinks the piece of information *unlikely*, this time, but twice finds it maximally *extremely likely*. He changes his mind more often than user 1 does, yet manages to have a higher score than before.

As a conclusion of the various studied cases, User 1 appears to be persuaded, to degrees varying between τ_2 and τ_4 , whereas

TABLE VI. SCORE EVOLUTION USING BOTH CREDIBILITY APPROACHES ON A REORDERED INFORMATION STREAM

id.	Immediate		Cumulative					
	User 1	User 2	User 1			User 2		
	τ_2	τ_2	τ	g_+	g_-	τ	g_+	g_-
i_5	τ_3	τ_4	τ_3	$\tau_?$	$\tau_?$	τ_4	$\tau_?$	$\tau_?$
i_2	τ_3	τ_2	τ_3	$\tau_?$	τ_2	τ_2	$\tau_?$	$\tau_?$
i_1	τ_3	τ_2	τ_3	τ_1	τ_2	τ_2	τ_1	$\tau_?$
i_3	τ_3	τ_3	τ_4	$\tau_?$	$\tau_?$	τ_4	$\tau_?$	$\tau_?$
i_4	τ_3	τ_1	τ_4	$\tau_?$	τ_2	τ_2	$\tau_?$	$\tau_?$
i_6	τ_3	τ_0	τ_2	$\tau_?$	$\tau_?$	τ_1	$\tau_?$	$\tau_?$

User 2 remains unconvinced: this example shows how the model makes it possible to represent distinct credulity stances, providing a global formalism allowing to express very different behaviours and different adjustments to sensitivities.

VI. CONCLUSION AND FUTURE WORKS

This paper offers a multivalued formalisation of the cross-checking step of a trust building process. Various operators are proposed which allow to model the dynamics of the process, different stances with regards to persuasion and sensitivities to the order arguments are delivered. Different variants of the operators offer different tools for information evaluation, to take into account all levels of persuasion momentum.

Future works will include an experimental comparison of the model and its variants, needing to address the difficult task of making an evaluation protocol. Another perspective lies in the study of the proposed model in the framework of argumentation theory, from the point of view of debaters. Such a study would look at the ideal moment to take part in a debate and provide contradictory arguments, in the aim of maximising the trust level of the audience.

REFERENCES

- [1] NATO, "Annex to STANAG2022," Information handling system, 1997.
- [2] R. Demolombe, "Reasoning about trust: a formal logical framework," in *Int. Conf. on iTrust*, 2004.
- [3] J. Besombes and A. Revault d'Allonnes, "An extension of STANAG2022 for information scoring," in *Proc. of the Int. Conf. on Information Fusion*, 2008, pp. 1635–1641.
- [4] M.-J. Lesot, T. Delavallade, F. Pichon, H. Akdag, and B. Bouchon-Meunier, "Proposition of a semi-automatic possibilistic information scoring process," in *Proc. of EUSFLAT'11*, 2011, pp. 949–956.
- [5] F. Pichon, D. Dubois, and T. Denœux, "Relevance and truthfulness in information correction and fusion," *IJAR*, vol. 53, no. 2, pp. 159–175, 2011.
- [6] A. Revault d'Allonnes, "An architecture for the evolution of trust: Definition and impact of the necessary dimensions of opinion making," in *Information Evaluation*, P. Capet and T. Delavallade, Eds. Wiley, 2014, ch. 9, pp. 261–292.
- [7] A. Revault d'Allonnes and M.-J. Lesot, "Formalising information scoring in a multivalued logic framework," in *Proc. of IPMU'14*, vol. 1, no. CCIS442. Springer, 2014, pp. 314–323.
- [8] P. Capet and A. Revault d'Allonnes, "Information evaluation in the military domain: Doctrines, practices and shortcomings," in *Information Evaluation*, P. Capet and T. Delavallade, Eds. Wiley, 2014, ch. 9, pp. 103–128.
- [9] H. Akdag, M. De Glas, and D. Pacholczyk, "A qualitative theory of uncertainty," *Fundamenta Informaticae*, vol. 17, no. 4, pp. 333–362, 1992.
- [10] H. Seridi and H. Akdag, "Approximate Reasoning for Processing Uncertainty," *J. of Advanced Comp. Intell. and Intell. Informatics*, vol. 5, no. 2, pp. 110–118, 2001.