Implementing Explanation-Based Argumentation using Answer Set Programming

Giovanni Sileno, Alexander Boer, and Tom van Engers

Leibniz Center for Law,
University of Amsterdam, the Netherlands
{g.sileno,a.w.f.boer,t.m.vanengers}@uva.nl

Abstract. This paper presents an implementation for an explanation-based theory of argumentation. Instead of referring to attack/support relationships between arguments, as in traditional argumentation theories, we focus on the relation of messages with the space of hypothetical explanations. The consequences of this choice are two-fold. First, attack and support relationships become derivative measures. Second, we unveil a natural integration with probabilistic reasoning. The proposed operationalization is based on stable models semantics for logic programming.

Keywords: Argumentation, Explanation-based argumentation, Justification, Answer set programming

1 Introduction

Argumentation is traditionally perceived as operating at a *meta-level*, concerned with *support* and *attack* relationships between *claims* uttered by participants in a conversation. Although absolutely not bound by such practical perspective, formal theories follow, if not strengthen, this meta-level interpretation. Dung's seminal work [1] reduces argumentation to an abstract setting, which consists of a set of atomic components called *arguments* and *attack* relations between them. In this context, an argument can be for instance an atomic proposition, a (defeasible) rule, or even an argument scheme. In order to interpret such *argument systems*, e.g. so as to evaluate conflicts between arguments, Dung and following authors have proposed many formal *argumentation semantics* (for an overview, see [2]), used as a basis for deriving the *justification state* of each argument. In *extension-based semantics*, for instance, the key role is given to *extensions*, i.e. subsets of arguments of the argumentation framework, collectively *acceptable* according to a given semantic. The *justification* of an argument is then defined in terms of its membership to extensions.[1]

A practical application of this abstract framework would consist in three steps: (a) the *observation* of the argumentation process between certain parties[2], in a certain applied domain, (b) the *reduction* of the observation to a

[1] An argument is *skeptically justified* if it belongs to all extensions, it is *credulously justified* if there is at least one extension which contains it.

[2] Not necessarly different persons, parties may belong to the same person, assuming different perspectives.

system of arguments and attacks between arguments, (c) the *analysis*, using a certain argumentation semantic, of the resulting argumentation framework, so as to assign a certain justification state to arguments. Each task may be conventionally associated to a different role: the *observer*, the *modeler* and the *analyst.* Unfortunately, few but important issues haunt this operational chain.

Inside and Outside of Argument Systems First, the extraction of relations between distinct utterances may be problematic. Claims are often not explicitly directed against other claims (i.e. the *syntaxic* definition of attack). The step (b) externalizes this problem to the modeler. As a consequence, different modelers may produce alternative results, because the underlying process depends on cognitive abilities and background knowledge of the modeler. Despite of being abstracted as *systemically external* to the whole process, the construction of the argumentation meta-level is intrinsic to the argumentation process as well.

In order to solve this issue, many authors connect argumentation to default reasoning and other non-monotonic logics. For instance, in *assumption-based argumentation* (ABA) [3], arguments and attacks are not any more primitive components. Arguments are derived via backward reasoning (from conclusions to assumptions) using a given set of inference rules. Attacks on a target argument are defined if the "contrary" of the assumptions of this argument can be inferred. Other approaches [4, 5] count explicitly also the *rebuttal* attack, related to the deduction of the negation of the conclusion.[3] In both cases, the externalization of (b) is now placed at the level of the *support* relationships, defined via *defeasible rules*, and based on *assumptions.* Unfortunately, potential problems still exist, as *zombie arguments* and the *consilience effect*, which will be discussed hereafter.

The main objective of *explanation-based argumentation* is to push the limit of the externalization further (or equivalently, to not consider a meta-level for argumentation). Relying on a *deep model* of the domain, the relationships of attack and support become derivative measures of the impact of the observation on the space of explanatory hypotheses.

Strength of Truth Second, justification is defined only in discrete terms: an argument is justified or not, and if justified, it can be skeptically or credulously justified. A more fine-grained determination is however necessary in most practical cases. When there is no skeptically justified conclusion, how to decide the strength of a certain credulously justified interpretation over another? Intuitively, counting the number of arguments present in the different extensions would be a measure of their strength — as proposed for instance in [7]. But other solutions are possible as well. According to the subjective interpretation of Bayesian probability, probability counts as a measure of the *strength of belief.* In this line of thought, a certain probability assigned to arguments can be considered as a proxy for their strength. Some authors, as for instance [8], propose to integrate probability to Dung's abstract framework; others target more applied contexts, as evidential reasoning [9, 10], in the legal domain [11, 12]. We share part of their

[3] Dung argues this case can be easily converted to the previous one [6].

objectives: as this contribution shows, explanation-based argumentation unveils a natural integration with probabilistic theory. However, while those works generally insist on the computation of *posterior* probabilities, we will consider a *confirmation measure* over explanations, making the role of subjective commitment more explicit.

The paper proceeds as follows. In section 2 we present a puzzle given by Pollock concerning argumentation and probabilistic reasoning. In sections 3 and 4 we present the main characteristics of an explanation-based theory of argumentation. In section 5, we operationalize it using *answer set programming.* In section 6, we report and comment our results. The paper ends with a note on further developments.

2 An Interesting Puzzle

Pollock presents in [13] a lucid philosophical critique on defeasible reasoning and how probabilistic methods approach the problem of *justification*, in the form of some interesting puzzles. He gives the following case: *Jones says that the gunman had a moustache. Paul says that Jones was looking the other way and did not see what happened. Jacob says that Jones was watching carefully and had a clear view of the gunman.*

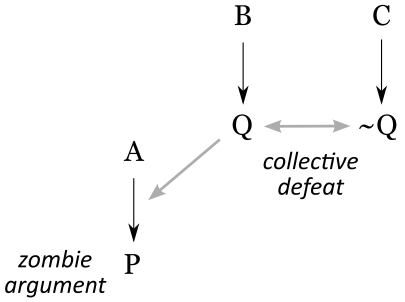

Fig. 1. Argumentation scheme of Pollock's puzzle

The associated argument scheme is illustrated in Fig. 1. A (Jones' claim) *supports* P (gunman had a moustache), B (Paul's) supports Q (Jones was not watching the gunman), C (Jacob's) supports $\sim Q$ (Jones was watching him). Evidently, Q *attacks* the relation between A and P, while Q and $\sim Q$ attack each other. In terms of argumentation, this is an example of *collective defeat* (Q vs $\sim Q$), which results in a *zombie argument* (P) [14]. Although formal semantics usually allow the presence of zombie arguments, it is not clear — Pollock admits — whether they should. Therefore, he targets some intuitive properties, easy to be agreed upon from a common-sense perspective:

1. given the conflict, we should not believe to Jones' claim carelessly;

2. if we consider Paul more trustworthy than Jacob, Paul's claim should be justified, but to a lesser *degree*;
3. if Jacob had confirmed Paul's claim, its *degree of justification* should have increased.[4]

Pollock gives then a preliminary, elaborated proposal based on "probable probabilities", which however does not differ in the idea of solving the issue within the meta-level of the argumentation framework.

3 Informal Presentation

Considering an applied perspective, argumentation can be seen as a *dialectic process.*[5] Parties produce and receive others' *messages*, interpreting and evaluating them. Sometimes these messages are collected by a third-party adjudicator, entitled to interpret the *case* from a neutral position. The set of collected messages forms an **observation**.

The presumption of conflict between parties is naturally associated to the *epistemic function* of argumentation. However, weaker definitions of conflict may include even a simple *assertion.* If an agent shares something with another agent who is ignorant about it, the second agent usually performs some evaluation on what the first said before believing in it. A similar process occurs during a *persuasion dialogue.* In our daily experience, we know that such evaluation does not concern only the content of such message, but also the context in which it has been provided. Related common questions are "Is what he says plausible?", "Is he reliable?", "Why is he telling that?", "Why now?", etc. More structured taxonomies of *critical questions*, relevant in specific domains of expertise, are available in the literature [16].

Generalizing this, we observe that argumentation does not involve only a certain *story* —the case which is matter of debate— but also the *meta-story* related to the "construction" of such a story. Therefore, in our perspective, given a disputed case, an **explanation** is a possible *scenario* compatible with the content of the provided messages *and* with the generation process of the messages as well. An explanation is *valid* if it reproduces the messages collected in the observation. Evidently, the nature of such scenarios is that of a *multi-representation* model [17]: they may integrate physical, intentional, socio-institutional, and abstract domains.

Traditionally, AI research relates *story understanding* to *abduction* [18]. This connection holds also here. There may be several valid explanations associated to the same observation. However, when interpreting a story, and still more when adjudicating a case, we are interested in determining what *is* the case. Valid explanations compete, and such competition is a matter of (epistemic) *justification.*

[4] We slightly changed the third one, in order to make use of the same story.

[5] For an overview of contributions focusing on the dialogue aspect of argumentation you may refer to [15, 2.1].

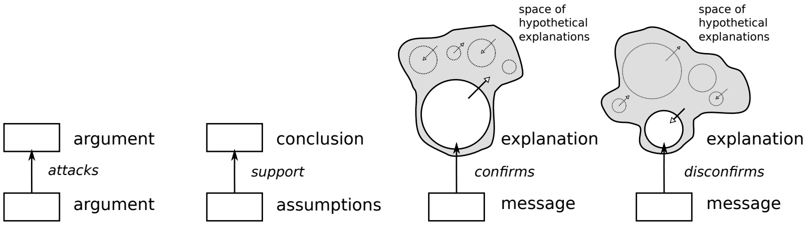

Fig. 2. A visual synthesis of attack in Dung's argumentation framework, support in assumption-based argumentation, and confirmation/disconfirmation in explanation-based argumentation

Furthermore, agents make claim also in order manipulate the explanatory search space ascribed to the recipient of the message. When the dissent opinion in *Pierson vs Post*[6] raises the argument concerning the ferocity of the fox, a new allocation rule is presented —or better, its applicability is claimed—, and new arguments are generated from the assumptions shared in the court.

Instead of being a static entity, the space of hypothetical explanation is incrementally constructed along with the observation (as *foreground*), integrated on top of common and expert knowledge about the world (as *background*), determining *factors*, *constraints* and *strengths of epistemic commitment.* We recognize therefore three operational steps in an explanation-based argumentation:

(a) *generation*: relevant factors are grounded into *scenarios*;
(b) *deletion*: (I) impossible scenarios are removed, leaving a set of *hypothetical explanations*; (II) hypothetical explanations fitting the observation are selected as *explanations*;
(c) *justification*: the relative position of explanations is evaluated according some measures of epistemic commitment.

Usually, (a) and (b) go together, resulting in a set of explanations.[7] The main difference between (b) and (c) is matter of certainty (in a way similar to the difference between *strict* and *defeasible* rules): they both represent a refinement over (a). In practice, hypothetical explanations can be associated to a certain *likelihood.* After some *confirming* message, the relative weights of explanations will change, and, consequently, stronger explanations will emerge from the set of hypothesis. Using a metaphor, *instead of being drawn, the best explanation is sculpted.* Conversely, when messages transport conflicting claims, they are actually *disconfirming* certain explanations, equilibrating their relative weight. The matter becomes raw again. A visual synthesis is in Fig. 2.

[6] This is a well-known case for the AI & Law community, see for instance [19].

[7] Argumentation frameworks based on default reasoning may be seen as covering these steps as well, but insisting on the *inferential* aspect of the problem, rather than the selection of an adequate search space.

Following this approach, not only attack, but also support between arguments becomes a derivative measure. In this way, we are not vulnerable to the *consilience* effect, occurring when a claim merely supports another claim because they both confirm the same explanation, while otherwise being uncorrelated.

Prior Probabilities as Likelihood Prior probabilities derived from statistics are problematic, especially in law. A more "just" approach would be in considering a *neutral perspective*, i.e. all hypothetical explanations having the same initial likelihood. On the other side, the *selection* of hypothetical explanations (and, more in general, the choice of relevant factors and background theory) hides already a certain commitment. This is evident in default reasoning. For instance, in the third scenario of the puzzle, where Jacob confirms what Paul says, a conflict-free extension is given by the two arguments brought by Jacob and by Paul. With "no-evidence-to-the-contrary" [20], we are neglecting the case in which they are both lying.

4 Formal Presentation

We present hereby a limited version of our framework. In its full form, explanations and observations would consist of *multi-agent systems* based on *agent-roles* [21, 22], so as to take into account intentional and institutional layers, and multiple figures for *speech acts*. For simplicity, we will consider here only a generic *emission* of propositional content, and neglect all intentional and causal components, leaving them to future extensions.

4.1 Fundamental Concepts

Definition 1. *A message M is a tuple $\langle E, R, C\rangle$, where E is the emitter entity, R the receiver entity, and C the message content. Hereafter it will be represented as $M = [C]_E^R$. λ_E is the function labelling messages with their emitters.*

Definition 2. *The content C of a message M is a proposition. Φ_C is the function mapping C to its relevant factors $\{f_1, ..., f_n\}$, i.e. the variables to be evaluated in order to assess its truth value.*

In general, when someone reports something, he may tell the truth or not, i.e. what he says may hold or not, considering truth as a successful *word-to-world* alignment. There is therefore an additional implicit factor involved in the evaluation of an assertion. In general, the *reliability* attributed to a source is a sufficient condition to consider the content of the message as holding.[8] Nevertheless, the fact that the agent is not reliable does not imply that he is necessarily lying[9], but, if he is *not* reliable, it is perfectly acceptable that what

[8] Consider the case of direct perceptions. In general we are practically certain of the reliability of our senses. In certain situations, however, we may doubt them.

[9] Note that the relation "being reliable" has no direct reference with the real intent of the emitter/assertor. Lying refers hereby only to a *word-to-world* misalignment.

he says does not hold. We call Φ'_C the extended version of the mapping function, including the reliability condition.

Definition 3. *Given an observer S_0 and a set of sources $\Sigma_S = \{S_1, ..., S_s\}$, an observation $O = \{M_1, ..., M_m\} = \{[C_1]^{S_0}_{\lambda_E(M_1)}, ..., [C_m]^{S_0}_{\lambda_E(M_m)}\}$ is a set of messages received by the observer from the sources.*

Definition 4. *An assumption A is a proposition.*

Because Φ_C applies on propositions, we can use it on assumptions as well. A concept similar to *reliability* can be introduced here as well, so as to make explicit a general principle of *conditional encapsulation.* There are factors related to the propositional content, and factors "outside" the propositional content, which may be relevant in certain conditions in determining its truth value. In the case of belief, truth is associated to a *mind-to-world* alignment, and the condition would be the *commitment to belief.*[10] In the case of rules, such external condition is *applicability.* Here as well, we call Φ'_C the extended versions of the mapping function.

Definition 5 (generation). *Given an observation O, a background theory B consisting of a set of assumptions $\{A_1, ..., A_s\}$, a scenario is an allocation of relevant factors of B and O.*

Definition 6 (deletion I). *Given a background theory B, a possible scenario or explanatory hypothesis is a scenario satisfying the constraints given by B.*

These definitions rely on a more generic *operational assumption*, i.e. given an observation, the modeler should be able to generate a set of *hypothetical explanations*, using factors adequately relevant to the observation.

Definition 7 (deletion II). *Given an observation O, an explanation is a possible scenario which fits with O.*

In general, in order to apply this methodology, explanations should be specific enough to entail the occurrence of a certain message. We name this as the *informative assumption*: an observation O either fits an explanation E or it doesn't. Note that in the present contribution, because we are neglecting the causal component, occurrence is translated with holding.

4.2 Evaluation of Explanations

The subsequent problem is to decide how to evaluate explanations, or, equivalently, how to measure their *degree of justification*, in respect to some prior assumptions about the world. We consider as mathematical framework of reference Bayesian probability. Given an explanation E, the likelihood of E on O $\mathcal{L}(E|O)$ is equivalent to $P(O|E)$ —the conditional probability of observing O given the hypothesis E— and, on the basis of the informative assumption, it is one of $\{0, 1\}$ for any E or O.

[10] Informally, an assumption is usually considered less strong than a belief. We can translate that as "This assumption certainly holds, if I believe in this assumption".

Definition 8. *If all likelihoods of explanations on a given observation are the same, then this observation is irrelevant.*

Going further, we assume, in Bayesian terms, that an explanatory problem space is *well-known*:

$$P(E_1) + P(E_2) + .. + P(E_n) \sim 1$$

If we are not willing to commit to prior probabilities, we may assume *prior indifference* between explanatory hypotheses $E_1, .., E_n$. In this case, we would have $P(E) = \frac{1}{n}$.

At this point, in order to determine the relative value of an explanation E, given O, we may calculate the likelihood of O on E. Considering the *Bayesian confirmation constraint* [23], we may say that O confirms E if $P(E|O) - P(E) > 0$, and disconfirms E if $P(E|O) - P(E) < 0$. Unfortunately, this measure does not permit ordinal comparisons of explanations. Therefore, we decide to calculate the confirmation value of O for explanation E with an alternative measure, which permits ordinal judgments.

Definition 9. *Given an observation O, and an explanation E, the confirmation value c of O on E is defined as:*

$$c(O, E) = \frac{P(O|E) - P(O|\neg E)}{P(O|E) + P(O|\neg E)} \tag{1}$$

Put in words, an observation confirms an explanation if it is predicted by the explanation *and* discriminates the explanation from its alternatives.[11] If $c(O, E)$ approaches 1 (-1), the observation O confirms (disconfirms) the explanation E. If c is equal to 0, O is irrelevant.

Once calculated for all explanations, confirmation values can be used to order them. The final ordering depends on:

- the effective capacity of generating adequate scenarios (i.e. the operational assumption), which contain fitting explanations (i.e. the informative assumption); if these assumptions are not adequately satisfied, relevant explanations may be missing;
- the set of *prior probabilities* associated to the generated explanations;
- the confidence in the previous structure as representation of the world (related to the assumption of well-known problem space).

Leaving apart the third point, it is difficult to qualify the second as related to “objective” measures. Statistics describe aggregates, not individuals. The case which is discussed has already occurred and we are in the realm of *epistemic uncertainty.* Prior probabilities play the role of prior assumptions towards the facts, and then become a matter of *belief.* If belief is involved, then the ordering of explanations expresses their relative (epistemic) *justification.*

[11] Tentori et al. [23] suggest that the above is the psychologically most plausible confirmation measure of those proposed in the literature.

Definition 10 (justification). *Given an observation O, and an explanation E, the degree or strength of justification of E is a measure relative to his position in the set of all explanations, ordered by their confirmation value.*

We preferred to leave this generic definition, because further research is required in order to propose a specific analytic expression.

5 Operationalization

In this section, we will translate the previous concepts in a concrete computational setting. After a brief overview on *answer set programming*, we describe how it can be integrated in an explanation-based argumentation framework. Then we provide some modeling guidelines, and, at the end of the section, we analyse more in detail the computation of the confirmation values.

5.1 Answer Set Programming: an Introduction

Answer set programming is a declarative programming paradigm [24] based on the *stable-model semantic*[12] [27], oriented towards difficult (NP-hard) search problems. In the literature, ASP is used to model and solve problems belonging to a wide range of applications. For instance, [28, 29] apply ASP to compute extension-based semantics on argument systems. In ASP, similarly to Prolog, the programmer models a problem in terms of rules and facts, instead of specifying an algorithm. The resulting code is given as input to a solver, which returns multiple *answer sets* or *stable models* satisfying the problem. The main operational difference to Prolog is that all variables are grounded before performing search, and unlike SLDNF resolution, ASP solvers algorithms always terminate.

5.2 Integration with Explanation-Based Argumentation

Basically, our idea is to take advantage of the search capabilities of ASP solvers, in order to effectively perform the *generation* and *deletion* steps at once. An ASP program related to an explanation-based argumentation consists of 3 parts:

1. *allocation choices*, grounding all permutations of relevant factors,
2. *world properties* and *ground facts*, modeling shared assumptions,
3. *observation*, modeling the collected messages.

The execution of a program with only (1) would give scenarios (possible and impossible); with (1) and (2) the hypothetical explanations; with the complete code the explanations. (2) can be interpreted as the *deep model* of the domain.

At this point, hypothetical explanations and explanations can be parsed. Assigning a prior probability to hypothetical explanations, and analysing the resulting final explanations we can calculate their confirmation values. In our prototype, this is performed via an external script.

[12] Stable-model semantics apply ideas of *auto-epistemic* logic of Moore [25] and default logic of Reiter [26].

5.3 Modeling Guidelines

Facts are written as in Prolog.[13] For instance, "it is raining" can be written as:

```
rain.
```

If the rule is given as a property of the world (and then as a constraint on possible worlds) it is coded similarly to Prolog rules as well. Prolog rules consist of a *head* (conclusion) and a *body* (premises). For instance, a rule R_1 described by $rain \to wet$ can be written as:

```
wet :- rain.
```

Differently to the usual use of rules, however, we want to *ground* all relevant factors, so as to generate all scenarios, possible and impossible. The pruning of impossible ones will occur subsequently, with the application of the rules.

In ASP, a factor f can be allocated using the *choice* operator[14], so as to translate the principle of non-contradiction (f holds or it doesn't: $f \oplus \neg f$):

```
1{f, -f}1.
```

We consider then the following *allocation principle*:

Proposition 1. *For each assumption A belonging to a background theory B, we allocate each factor $f \in \phi'_C(A)$. For each message M part of an observation O, we allocate each factor $f \in \phi'_C(M)$.*

Applying this principle to the previous rule we have:

```
1{rain, -rain}1.
1{wet, -wet}1.
wet :- rain.
```

A rule can be also activated at a second level, i.e. introducing an external condition which triggers the constraint that the rule *posits* to the world. Using the material implication ($a \to b \Leftrightarrow \neg a \vee b$), we translate R_1 as $\neg rain \vee wet$. Anchoring the rule to this factor $r1$ (to be read as *the rule R_1 applies*, or R_1 *is applicable*), we can model it as:

```
1{-rain, wet} :- r1.
r1.
```

If $r1$ is not grounded, i.e. the applicability of R_1 is matter of debate, we may count it as a relevant factor as well:

[13] The code excerpts presented here refer in particular to the syntax of the ASP solver `lparse+smodels`.

[14] A brief summary of the syntax of ASP logic operators: OR: $a_1 \vee .. \vee a_N \Leftrightarrow$ `1{a1, .., aN}` — XOR: $a_1 \oplus .. \oplus a_N \Leftrightarrow$ `1{a1, .., aN}1` — AND: $a_1 \wedge .. \wedge a_N \Leftrightarrow$ `a1, .., aN` (only in the body of rules) or `N{a1, .., aN}N` (body and head).

```
1{-rain, wet} :- r1.
1{r1, -r1}1.
```

We can use this artifice also to create meta-rules concerning the priority between rules. In general, each rule may hold or not. If a certain rule holds then the conflicting rules with lower priority do not hold. If we want to model $R_2 > R_1$, we can write:

```
1{r1, -r1}1.
1{r2, -r2}1.
-r2 :- r1.
```

We can also rewrite the rule in the last line introducing another fact $r1r2$ (to be read as *the meta-rule* $R_1 > R_2$ *holds*)

```
1{-r1, -r2} :- r1r2.
r1r2.
```

Messages As we observed in section 4.1, an assertion, as individual message, may be generalized taking into account a *reliability* condition:

- what an agent says may hold or not (*allocation choice*),
- an agent may be reliable or not (*allocation choice*),
- when he is reliable, what he says is what it holds (*constraint rule*).

In our puzzle, Paul says Jones was not seeing the gunman. Writing “Paul is reliable” as `paul` and “Jones was seeing” as `eye`, we have:

```
1{eye, -eye}1.
1{paul, -paul}1.
-eye :- paul.
```

We can proceed similarly in case of reported rules and meta-rules.

5.4 Computation of Confirmation Values

When integrated in a explanation-based framework, the outputs of the ASP solver are sets of (hypothetical) explanations. For all E_i, $c(O, E_i)$ depends on two parameters. $P(O|E_i)$ is equal to 1 if E_i belongs to the output of the execution of the complete code, 0 otherwise. If it is 0, we directly infer that $c(O, E_i) = -1$. In the remaining cases, $P(O|\neg E_i)$ can be calculated as:

$$P(O|\neg E_i) = \frac{P(O \cap \neg E_i)}{P(\neg E_i)} = \frac{\sum_{i \neq j} P(O \cap E_j)}{1 - P(E_i)} = \frac{\sum_{i \neq j} P(O|E_j) \cdot P(E_j)}{1 - P(E_i)} \quad (2)$$

Then, in order to compute c, we need a *prior probability* $P(E_i)$ for each E_i. In general, we could assign it directly at this point. However, we can also take advantage of the generative construction associated to the methodology. We know which factors are relevant in the construction of an explanation, therefore, if we have prior assumptions in their respect, we can calculate $P(E_i)$ starting from these components. Given the set of all relevant factors $\phi = \{f_1, f_2, ..., f_n\}$, assuming they are all *independent*, we have:

$$P(E_i) = P(f_1) \cdot P(f_2) \cdot ... \cdot P(f_n)$$

To obtain a neutral perspective toward explanations, all $P(f_i) = 0.5$. More complex figures may be obtained integrating *Bayesian networks* to model the probabilistic relationships between factors. In all cases, however, it is important to underline that this counts as a subjective measure of belief. If $P(f_i) = 0.5$, we are neutral towards that factor f_i. It may equivalently hold or not in a specific case. If $P(f_i) > 0.5$, we assign more likelihood for its presence, and vice-versa for $P(f_i) < 0.5$. If all factors are neutrally positioned, we are not able to discriminate explanations, because they will have the same degree of justification.

6 Results

Puzzle code In order to apply our methodology on Pollock's puzzle, we start by translating the proposed story. We have evidently three messages. The only world property that we consider is that, in order to be reliable, Jones has at least seen the gunman (`eye` is a necessary condition to Jones' reliability). In code:

```
%% allocation
1{moustache, -moustache}1.
1{eye, -eye}1.
1{jones, -jones}1.
1{paul, -paul}1.
1{jacob, -jacob}1.

%% world property
eye :- jones.

%% observation
moustache :- jones.
-eye :- paul.
eye :- jacob.
```

ASP solver output For analysis purposes, we write down the code to generate hypothetical explanations (*before* observation) and the code for explanations (*after* observation) for each step of the observation ($O_1 = \{M_1\}, O_2 = \{M_1, M_2\}$, etc.), and we have in total 6 ASP programs. The following table resumes the numbers at stake:

#	O_1	O_2	O_3
relevant factors	3	4	5
scenarios	8	16	32
hypothetical explanations	6	12	24
explanations	5	7	10

As the table shows, the introduction of new factors entails an explosion of the number of scenarios (given n factors, 2^n), and similarly the number of hypothetical explanations and of explanations.[15] The following table illustrates all resulting explanations:

`moustache`	T						F			
`eye`	T				F		T		F	
`jones`	T		F		F		F		F	
`paul`	F		F		T	F	F		T	F
`jacob`	T	F	T	F	T	F	T	F	F	F

Probabilistic evaluation Before proceeding in the analysis we introduce a synthesized visualization:

Definition 11. *Given a set of explanations $\{E_1, ..E_n\}$, for each allocation factor f, if f is true (false) in all explanations, then f is true (false) in the union explanation E_U, otherwise f is undecided.*

Basically, we refer to the *union explanation* of the set of explanations with the *maximum* confirmation value to have a synthesis of the shared common points of the best explanations. The following table synthetizes the incremental results ($O_1, \ldots, O_3$ columns), in different probabilistic settings (P columns):

	E_U (Jacob attacks Paul)											
	P	O_1	O_2	O_3	P	O_1	O_2	O_3	P	O_1	O_2	O_3
`moustache`	.50	U	U	U	.50	U	U	U	.50	T	U	T
`eye`	.50	U	U	U	.50	U	F	F	.50	T	U	T
`jones`	.50	U	U	U	.50	U	F	F	**.55**	T	U	T
`paul`	.50	-	U	U	**.55**	-	T	T	**.55**	-	U	F
`jacob`	.50	-	-	U	.50	-	-	F	**.55**	-	-	T
$c(O\|E_U)$		.27	.43	.55		.27	.45	.58		.31	.48	.61

Let us analyze the table according to the properties targeted by Pollock, as reported in section 2. (1) Assuming indifference toward hypotheses (group of columns on the left), our approach confirms *to the same degree* hypotheses in which the gunman has a moustache, and not. (2) Using for instance $P(\texttt{paul}) = 0.55 > P(\texttt{jacob}) = 0.5$ (in the middle of the table), the hypothesis in which Paul is telling the truth is the one confirmed to the greater degree. Note that we still cannot say anything about the moustache.

In the third setting (on the right), we consider the case in which all witnesses are assumed relatively reliable. When a conflict arises, then things become unclear again. When Jacob supports Jones however, the first scenario wins again, producing a kind of *majority opinion* effect. This is an acceptable choice when the epistemic motivation of the agents is taken for granted (e.g. judges, experts).

[15] On the contrary, if we add messages concerning only already known factors, the number of explanations will decrease, or remain the same. This is not the case in our puzzle: each message comes with its own reliability condition.

However, a best practice would be to consider as much as possible a position of indifference towards prior assumptions. This would help for instance to take into account organized crime scenarios.

Story variations Following the puzzle, we modify the story, in a way that Jacob confirms what said by Paul. The resulting explanations are:

`moustache`	T						F	F			
`eye`	T		F				T	F			
`jones`	T	F	F				F	F			
`paul`	F	F	T		F		F	T		F	
`jacob`	F	F	T	F	T	F	F	T	F	T	F

As before, we calculate the confirmation values and the union explanations for a few probabilistic settings:

	E_U (Jacob supports Paul)							
	P	O_1	O_2	O_3	P	O_1	O_2	O_3
`moustache`	.50	U	U	U	.50	T	U	U
`eye`	.50	U	U	U	.50	T	U	F
`jones`	.50	U	U	U	**.55**	T	U	F
`paul`	.50	-	U	U	**.55**	-	U	T
`jacob`	.50	-	-	U	**.55**	-	-	T
$c(O\|E_U)$		.27	.43	.51		.31	.48	.57

(3) In conditions of indifference (on the left), we observe that the hypothesis they both support are confirmed *just as much* as the hypotheses in which they are lying. While, assuming *bona fide* (on the right), we are again in a majority opinion case.

For completeness, we consider also the case in which they all support each other (at least partially — Jones remains the only one able to the see the gunman).

`moustache`	T									F				
`eye`	T								F	T				F
`jones`	T				F				F	F				F
`paul`	T		F		T		F		F	T		F		F
`jacob`	T	F	T	F	T	F	T	F	F	T	F	T	F	F

The number of scenarios have increased. The confirmation values are in general lower then before, showing an underlying principle similar to Shannon's theory of communication: less conflict, less "information".

	E_U (no conflict)							
	P	O_1	O_2	O_3	P	O_1	O_2	O_3
`moustache`	.50	U	U	U	.50	T	T	T
`eye`	.50	U	U	U	.50	T	T	T
`jones`	.50	U	U	U	**.55**	T	T	T
`paul`	.50	-	U	U	**.55**	-	T	T
`jacob`	.50	-	-	U	**.55**	-	-	T
$c(O\|E_U)$		.27	.36	.41		.31	.41	.45

Supports and attacks Analyzing the confirmation values, we can draw some pictures about attack/support relationships. We can consider as target arguments both assumptions (`paul`, `jacob`, `jones`), than conclusions (`eye`, `moustache`). We compute the confirmation values before and after the observation, and, calling $c_{max}(f)$ the maximum confirmation value of explanations in which f is satisfied, we have (in respect to the original puzzle):

		$c_{max}(f)/c_{max}(-f)$					
		O_1		O_2		O_3	
	P	pre	post	pre	post	pre	post
`moustache`	.50	.19/.19	.31/.30	.18/.18	.48/.48	.17/.17	.61/.60
`eye`	.50	.19/.18	.31/.30	.18/.17	.48/.48	.17/.16	.61/.60
`jones`	.55	.19/.18	.31/.30	.18/.17	.48/.48	.17/.16	.61/.60
`paul`	.55			.18/.17	.48/.48	.17/.16	.60/.61
`jacob`	.55					.17/.16	.61/.60

Thus, we can refer to two dimensions of change, for each observation O_i: from after to before the observation (post O_i $-$ pre O_i), and from the current observation to the previous one (post O_i $-$ post O_{i-1}). The following table illustrates the incremental impact of observations in respect to the ongoing dialogue, according to such dimensions:[16]

	O_1	O_2	O_3
`moustache`	$+/+$	$=/-$	$+/+$
`eye`	$=/+$	$-/-$	$=/+$
`jones`	$=/+$	$-/-$	$=/+$
`paul`		$-/=$	$-/-$
`jacob`			$=/+$

7 Further developments

The paper presents a first implementation of an explanation-based argumentation framework. Despite its concrete operational result, many points require further research. For instance, we are working on finding adequate analytical

[16] We take an undecided position when confirmation values are not defined (e.g. O_0).

expressions to measure the (relative) justification of an explanation, and to derive attack and support from confirmation/discorfimation. As we can see in the pictures, confirmation values tend to always increase introducing a new message. Intuitively, the relative ratio of the increase depends on a kind of informational value, relative to the *clarification* of the case. Our objective is to extract this informational measure from the confirmation value, so as to be independent from the number of messages taken into account. Related to this objective, we plan to quantify the strength of attack/support, given a certain message in a contextual dialogue. Future extensions will consider the integration of game-theory analysis.

Evidently, the crucial point of this methodology is on the ability of constructing an adequate *deep model*. From a wider perspective, our research aims to integrate background theories and allocation factors described in terms of *agent-roles* [21, 22], acknowledged in specific social settings. In this respect, the configuration investigated here is limited: no treatment of events/causation, only assertion, no intentional/institutional components, etc. This was functional to present a global picture of the methodology, and operationalize it with Pollock's puzzle. Further developments will investigate the encapsulation of fundamental components (as the "emitter" seen here) in higher-abstraction models. Observations and explanations would become explicitly structured on multi-agent systems. As a relevant consequence, we will need to evaluate how Answer Set Programming (or other computational tools) respond to the integration of such more complex models.

References

1. Dung, P.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artificial intelligence 77 (1995) 321–357
2. Baroni, P., Giacomin, M.: Semantics of Abstract Argument Systems. In Simari, G., Rahwan, I., eds.: Argumentation in Artificial Intelligence. (2009) 25–44
3. Bondarenko, A., Dung, P., Kowalski, R., Toni, F.: An abstract, argumentation-theoretic approach to default reasoning. Artificial intelligence 93 (1997) 63–101
4. Pollock, J.: Defeasible reasoning. Cognitive science (1987) 1–31
5. Prakken, H., Sartor, G.: The role of logic in computational models of legal argumet - a critical survey. Computational Logic: Logic Programming and Beyond. Essays In Honour of Robert A. Kowalski 2 (2002) 342–380
6. Dung, P., Kowalski, R., Toni, F.: Assumption-based argumentation. Argumentation in Artificial Intelligence (2009) 1–20
7. Baroni, P., Dunne, P.E., Giacomin, M.: On Extension Counting Problems in Argumentation Frameworks. Proceedings of the COMMA 2010: Conference on Computational Models of Argument (2010) 63–74
8. Li, H., Oren, N., Norman, T.: Probabilistic argumentation frameworks. Theory and Applications of Formal Argumentation (2012) 1–16
9. Williams, M., Williamson, J.: Combining argumentation and Bayesian nets for breast cancer prognosis. Journal of Logic, Language and Information (2006) 1–23

10. Keppens, J.: On extracting arguments from Bayesian network representations of evidential reasoning. In: Proceedings of the ICAIL 2011: 13th International Conference on Artificial Intelligence and Law (2011) 141–150
11. Fenton, N., Neil, M., Lagnado, D.a.: A general structure for legal arguments about evidence using Bayesian networks. Cognitive science 37(1) (2012) 61–102
12. Timmer, S.T., Meyer, J.J.C., Prakken, H., Renooij, S., Verheij, B.: Inference and Attack in Bayesian Networks. In: Proceedings of the BNAIC 2013: 25th Benelux Conference Artificial Intelligence (2013)
13. John L. Pollock: Reasoning and probability. Law, Probability and Risk 6(1-4) (2007) 43–58
14. Makinson, D., Schlechta, K.: Floating conclusions and zombie paths: Two deep difficulties in the directly skeptical approach to defeasible inheritance nets. Artificial Intelligence 48(2) (1991) 199–209
15. Bench-Capon, T.J., Dunne, P.E.: Argumentation in artificial intelligence. Artificial Intelligence 171(10-15) (2007) 619–641
16. Walton, D., Reed, C., Macagno, F.: Argumentation Schemes. Cambridge University Press (2008)
17. Mueller, E.T.: Story understanding through multi-representation model construction. In: Text Meaning: Proceedings of the HLT-NAACL 2003 Workshop (2003) 46–53
18. Charniak, E., McDermott, D.: Introduction to Artificial Intelligence. Addison-Wesley (1985)
19. Atkinson, K., Bench-capon, T., Prakken, H., Wyner, A.: Argumentation Schemes for Reasoning about Factors with Dimensions. Proceedings of the JURIX 2013: 26th International Conference Legal Knowledge and Information Systems (2013) 39–48
20. Dung, P., Mancarella, P., Toni, F.: Computing ideal sceptical argumentation. Artificial Intelligence 171(10-15) (2007) 642–674
21. Boer, A., van Engers, T.: Diagnosis of Multi-Agent Systems and Its Application to Public Administration. In: Business Information Systems Workshops. Volume 97 of Lecture Notes in Business Information Processing, (2011) 258–269
22. Boer, A., van Engers, T.: An Agent-based Legal Knowledge Acquisition Methodology for Agile Public Administration. Proceedings of the ICAIL 2011: 13th International Conference on Artificial Intelligence and Law (2011) 171–180
23. Tentori, K., Crupi, V., Bonini, N., Osherson, D.: Comparison of confirmation measures. Cognition 103(1) (2007) 107–119
24. Lifschitz, V.: What Is Answer Set Programming? Proceedings of the AAAI Conference on Artificial Intelligence (2008)
25. Moore, R.C.: Semantical Considerations on Nonmonotonic Logic. Artificial Intelligence 25(1) (1985) 75–94
26. Reiter, R.: A logic for default reasoning. Artificial Intelligence 13(1-2) (1980) 81–132
27. Gelfond, M., Lifschitz, V.: The stable model semantics for logic programming. Proceedings of the of International Logic Programming Conference and Symposium (1988) 1070–1080
28. Egly, U., Gaggl, S., Woltran, S.: Aspartix: Implementing argumentation frameworks using answer-set programming. Logic Programming (2008) 734–738
29. Osorio, M., Zepeda, C.: Inferring acceptable arguments with answer set programming. Proceedings of the EMC 2005: 6th Mexican International Conference on Computer Science (2005) 198–205