Proceedings of the 22nd International Conference on Principles of Knowledge Representation and Reasoning
Main Track

An Axiomatic Study of a Modular Evaluation of Enthymeme Decoding in Weighted Structured Argumentation

Jonathan Ben-Naim[1] , **Victor David**[2*] , **Anthony Hunter**[3]
[1]Université Paul Sabatier, CNRS, IRIT, France
[2]Université Côte d'Azur, Inria, CNRS, I3S, France
[3]University College London, UK
jonathan.ben-naim@irit.fr, victor.david@inria.fr, anthony.hunter@ucl.ac.uk

Abstract

An argument can be seen as a pair of premises and a claim they support. Human arguments are often approximate, with some premises left implicit, leading to an implicit inference of the claim, i.e., forming enthymemes. To better understand and use them, we must decode these approximate enthymemes, typically by identifying missing premises to make the inference explicit, and, as we propose, by also removing irrelevant content to improve argument quality in specific contexts. Often, multiple decodings of an enthymeme are possible. However, no formal method has yet been proposed for *identifying higher-quality decodings*. To pave the way, we introduce six types of criteria for evaluating aspects of decodings. Then, we introduce the concept of a *criterion measure*, designed to evaluate decodings based on a specific criterion. In parallel, we define desirable properties for criterion measures, referred to as *axioms*, and we systematically evaluate our criterion measures with respect to them. Finally, we introduce the notion of quality measure that combine specific criterion measures to give an overall evaluation of the quality of decodings.

1 Introduction

In the literature on logic-based argumentation, a deductive argument is usually defined as a premise-claim pair where the claim is inferred (according to a logic) from the premises. However, when studying human debates (i.e., real world argumentation), it is common to find incomplete arguments, called enthymemes, for which the premises are insufficient for implying the claim. The reason for this incompleteness is varied, for example it may result from imprecision or error, e.g., a human may argue without knowing all the necessary information, or it may be intentional, e.g., one may presuppose that some information is commonly known and therefore does not need to be stated, or that it is being deployed as a rhetorical device (employment of enthymemes is a well-known instrument since Aristotle (Faure 2010) for persuasion of an audience).

1.1 Novel Framework for Decoding Enthymemes

There are studies in the literature on understanding enthymemes in argumentation, using natural language processing (Habernal et al. 2017; Singh et al. 2022; Wei et al. 2022), but these do not identify logic-based arguments. There are also symbolic approaches for decoding enthymemes in structured argumentation (Hunter 2007; Dupin de Saint-Cyr 2011; Black and Hunter 2012; Hosseini, Modgil, and Rodrigues 2014; Xydis et al. 2020; Panisson, McBurney, and Bordini 2022; Hunter 2022; Leiva, Gottifredi, and García 2023; Ben-Naim, David, and Hunter 2024; David and Hunter 2025), which frame the task as identifying a *completion*, i.e., a set of additional premises that, when combined with the original ones, render the inference of the claim explicit. A simple example of an enthymeme is the following where $r = \texttt{weather report predicts rain}$ and $u = \texttt{take an umbrella}$ are atoms: The premises $\{r\}$ are insufficient to entail the claim u (i.e. $\{r\} \not\vdash u$), but if the formula $r \rightarrow u$ is added to the premises, we have a completion since the claim now follows (i.e. $\{r, r \rightarrow u\} \vdash u$). The added premise bridges the gap between the initial information and the claim, yielding a fully explicit and logically valid inference.

For real-world argumentation, *decoding* offers a broader framework than completion. Unlike previous approaches that treat enthymemes as incomplete arguments requiring only additional premises to restore inference, our framework allows both the addition and removal of information, adapting the decoding to the needs of a specific application or user. This generalization is not just about covering more cases, it introduces flexibility that allows for decodings to have qualities beyond just the premises being consistent, or the premises implying the claim.

For instance, aspects like Granularity, which evaluates the amount of information in a decoding (assessing its level of conciseness or detail) may favor simplifying content rather than expanding it. A simple example illustrating the conciseness dimension of the Granularity criterion is the following: where $s = \texttt{weather report predicts sun}$, and $t = \texttt{report predicts high temperature}$, and $h = \texttt{take a sun hat}$ are atoms. The premises $\{s, t\}$ are insufficient to entail the claim h (i.e., $\{s, t\} \not\vdash h$). However, if both $s \rightarrow h$ and $s \wedge t \rightarrow h$ are available, the decoding $\langle \{s, s \rightarrow h\}, h \rangle$ may be preferred over $\langle \{s, t, s \wedge t \rightarrow h\}, h \rangle$ due to its greater conciseness.

Thus most appropriate decoding of an enthymeme can be undertaken across various dimensions of quality. When multiple decodings of an enthymeme are possible, a key chal-

*corresponding author

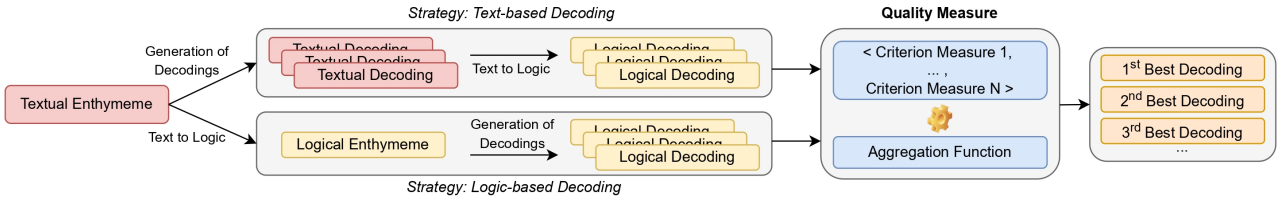

Figure 1: Possible pipeline for decoding a textual enthymeme into its best logical decoding(s), optionally back-translated into text, incorporating one of the two strategies to generate candidates in logic, which are then evaluated using our quality measures based on criterion measures and aggregation. In this paper, we only investigate quality measures and leave the other aspects of the pipeline to future work.

lenge is to compare and evaluate them systematically. The current state of the art lacks a formal evaluation framework for this. There is therefore a need for specific measures of the quality of decodings to be formalized, and for methods to aggregate different dimensions of quality, in order to select the more appropriate decoding for a specific context or objective.

Consider the following example (our running example), illustrating an enthymeme explaining why Bob is happy, with three possible decodings based on different reasons.

- Enthymeme E: Bob is wealthy, he is a researcher, he makes people happy, and he has people around him who seem to love him, then Bob is often happy.
- Decoding D_1: Bob is a researcher and researchers are often happy, so Bob is often happy.
- Decoding D_2: Bob makes people happy and has people who love him, and because giving and receiving love usually makes people happy, Bob is often happy.
- Decoding D_3: Bob is wealthy but not a researcher, wealthy people are generally happy, Bob is often happy.

To assess the quality of a decoding, e.g., to determine whether D_1, D_2 or D_3 is a better decoding for E, we represent the information using formal logic. This formalization makes explicit what is stated in the premises and the claim, allowing precise comparisons between candidates. Furthermore, we introduce a set of criteria for evaluating decoding quality, each associated with a measure designed to satisfy logical axioms that ensure desirable behavior. This formal foundation clarifies the interpretation of each criterion and enables automated evaluation. Importantly, our framework is highly general: it can be instantiated with any logical representation and any set of criteria.

1.2 Using Our Framework in a Pipeline

We see our framework for decoding enthymemes as an important component for addressing the implicit nature of argumentation found in natural language. In the field of argument mining, a growing body of research focuses on the extraction and structuring of argumentative content from natural language texts, including claims, premises, and the relations between them (Lippi and Torroni 2016; Lawrence and Reed 2020). However, real-world argumentation, particularly in debates, social media, or opinionated discourse, is often incomplete or implicit, posing a major challenge for both human interpretation and automated reasoning (Boltužić and Šnajder 2016; Habernal et al. 2017).

To tackle this, our framework may be used in a pipeline for processing natural language argumentation by evaluating multiple candidate decodings. We envision a pipeline with one or more decoding strategies such as illustrated in Figure 1. We explain these decoding strategies as follows.

(a) **Text-based Decoding.** In this strategy, textual candidates for decoding an enthymeme are first generated (e.g., by LLMs), then translated into logic, and finally assessed using our criterion measures to identify the best decodings.

Other authors, such as (Al Khatib et al. 2021), have shown that plausible argumentative completions can be generated from sentential claim inputs using a fine-tuned GPT-2 model enriched with argumentation knowledge graphs. To evaluate the generated outputs, human annotators labeled a subset of examples along five dimensions. These annotations were then used to train automatic classifiers, enabling large-scale evaluation of completions. While such evaluation is informative, it relies on supervised learning from human-labeled data, which limits scalability to new domains or tasks. Moreover, relying on classifiers trained on human judgments introduces an additional limitation: the evaluation process becomes opaque and difficult to interpret, reducing the explainability of the quality assessment.

In contrast, our approach defines quality measures in a formal framework, where decodings are unambiguous and evaluated without supervision (see Section 4). Logical representations allow for transparent, interpretable, and computable measures, supported by formal guarantees.

That said, it is important to acknowledge that translating natural language into formal logic remains a very challenging and unresolved task. While recent advances, such as those in (Han et al. 2022; Lu et al. 2022; Yang et al. 2023; Lalwani et al. 2024; Ryu et al. 2024), are encouraging, this remains an open research area. In this context, we do not assume such translation to be fully solved; rather, we view it as a promising direction that can benefit from our formal evaluation layer once candidate decodings (whether generated by humans or machines) are available in logical form.

(b) **Logic-based Decoding.** In this strategy, the enthymeme is first translated into a logical representation, then candidate decodings are generated directly with symbolic approach and evaluated using our quality measures.

Some existing approaches in the literature apply abductive reasoning to infer missing information and restore incomplete arguments or reasoning (Hunter 2007; Arieli et al. 2022; Hunter 2022). These methods assume access to a sufficiently complete knowledge base, such as a knowledge graph, from which implicit content can be inferred.

However, most existing methods, whether text-based or logic-based, focus solely on completion, adding information to fill gaps, without considering removal to satisfy alternative criteria like Granularity. This highlights the need for new decoding generation methods that go beyond completion. In both text-based and logic-based strategies, guiding the generation process with quality criteria could lead to more efficient and goal-oriented decoding.

By incorporating our framework into a pipeline such as in Figure 1, we enable a modular and explainable evaluation layer that assesses the quality of each decoding based on multiple criteria (see Section 4). This supports not only the correction of noisy or incomplete argument graphs, but also the development of richer, more interpretable reasoning systems that bridge symbolic and neural methods. Combining them through hybrid pipelines may lead to more robust and interpretable end-to-end argument understanding.

2 Weighted Logics

While our framework is designed to be logic-agnostic, it does not depend on any specific formalism and can, in principle, be instantiated with various argumentation systems such as ABA (Toni 2014) or ASPIC+ (Modgil and Prakken 2014). Weighted logics are commonly used in AI for representing and reasoning with imperfect or uncertain knowledge (Zimanyi 1992), and we argue that they offer a useful abstraction for qualifying different types of uncertainty involved in enthymeme decoding. Our goal is not to commit to a specific logic, but to define a flexible framework where the logic acts as a parameter. While existing weighted logic-based argumentation, such as (Alsinet et al. 2008), could be used, our contribution focuses on the criteria-based evaluation of decoding rather than the underlying logic. Therefore, we adopt a broad definition of weighted logic and illustrate it using a simple weighted propositional logic. Furthermore, approaches such as deductive argumentation (Besnard and Hunter 2001) can be captured in this framework by setting all weights to 1.

Definition 1. A **weighted language** is a set $\mathtt{W}$ such that:

- every element of $\mathtt{W}$ is a pair of the form $\alpha = \langle f, w\rangle$ such that f is a *formula* and w a *weight* in $[0, 1]$;
- if $\langle f, w\rangle \in \mathtt{W}$, then, $\forall\, v \in [0, 1]$, $\langle f, v\rangle \in \mathtt{W}$;
- $\forall\, w \in [0, 1]$, $\langle \bot, w\rangle \in \mathtt{W}$ ($\bot$ means *contradiction*).

We see weights as confidence scores of the reliability of formulas. For example, in fact-checking, automated methods can assess information reliability (Rashkin et al. 2017; Nakov et al. 2021).

Definition 2. A **weighted logic** is a triple $\mathbf{L} = \langle \mathtt{W}, \mathrel{|\!\sim}, t\rangle$ s.t.:

- $\mathtt{W}$ is a weighted language;
- $\mathrel{|\!\sim}$ is a **weighted consequence relation** on $\mathtt{W}$, i.e., a relation from $2^{\mathtt{W}}$ to $\mathtt{W}$;
- t is a *consistency threshold* belonging to $[0, 1]$.

We say that $\Gamma \subseteq \mathtt{W}$ is **inconsistent** on $\mathbf{L}$ iff there exists $w \geq t$ s.t. $\Gamma \mathrel{|\!\sim} \langle \bot, w\rangle$, and the set of all inconsistent set of formulae in $\mathbf{L}$ is denoted by $\mathtt{inc}_{\mathbf{L}}$. In the following, when the weighted logic $\mathbf{L}$ is clear, we will omit it (e.g., $\mathtt{inc}$ for $\mathtt{inc}_{\mathbf{L}}$). Otherwise, Γ is said to be consistent.

Next, we present an instance of weighted logic that will be used in examples. As a preliminary, we need two operators that extract the flat formulae or the weights from weighted formulae.

Definition 3. Let $\mathtt{W}$ be a weighted language and $\Gamma \subseteq \mathtt{W}$. We denote by $\mathtt{flat}(\Gamma)$ the set of every **flat formula** appearing in Γ, i.e., $\mathtt{flat}(\Gamma) = \{f : \exists\, w, \langle f, w\rangle \in \Gamma\}$.

We denote by $\mathtt{weight}(\Gamma)$ the set of every **weight** appearing in Γ, i.e., $\mathtt{weight}(\Gamma) = \{w : \exists\, f, \langle f, w\rangle \in \Gamma\}$.

From now on, for functions taking a set of weighted formulae as parameter, we simplify the notation for single-formula cases by removing the brackets: e.g., for $\alpha \in \mathtt{W}$, we write $\mathtt{flat}(\alpha)$ instead of $\mathtt{flat}(\{\alpha\})$.

As another preliminary, we recall the notion of classical propositional language.

Definition 4. We denote by $\mathtt{Lan}$ the set of **classical propositional formula** built from a given non-empty finite set of atomic formulae, denoted by $\mathtt{A}$, and the usual connectives $\neg$, $\vee$, $\wedge$, $\rightarrow$, and $\leftrightarrow$. A **literal** is either an element of $\mathtt{A}$ or the negation of it. For any flat formula $f \in \mathtt{Lan}$ we denote by $\mathtt{lit}(f)$ the set of literals occurring in f (as defined in (Lang, Liberatore, and Marquis 2003)), and $\forall F \subseteq \mathtt{Lan}$, $\mathtt{lit}(F) = \{l : l \in \mathtt{lit}(f) \text{ and } f \in F\}$.

We are ready to introduce our specific weighted logic that we will be used in examples.

Definition 5. We denote by $\mathtt{wLan}$ the **weighted propositional language**, i.e., $\mathtt{wLan}$ is the set of every pair $\langle f, w\rangle$ such that f in $\mathtt{Lan}$ and $w \in [0, 1]$.

We denote by $\mathbf{wLog}$ the **weighted propositional logic**, i.e., $\mathbf{wLog} = \langle \mathtt{W}, \mathrel{|\!\sim}, t\rangle$ s.t. the following holds:

- $\mathtt{W} = \mathtt{wLan}$;
- $\forall\, \Gamma \subseteq \mathtt{wLan}$, $\forall\, \alpha = \langle f, w\rangle \in \mathtt{wLan}$, $\Gamma \mathrel{|\!\sim} \alpha$ iff $\big(f$ is a tautology and $w = 1\big)$ or $\big(f$ is not a tautology, f classically follows from $\mathtt{flat}(\Gamma)$, i.e., $\mathtt{flat}(\Gamma) \vdash f$, and $w = \mathtt{min}[\mathtt{weight}(\Gamma)]\big)$;
- $t = 0.5$.

Following examples 1 and 2 illustrate this definition. From now on, whenever we work with a weighted logic $\mathbf{L}$, the typical instance we have in mind is $\mathbf{wLog}$.

Later in the paper, we count the number of elements in a set of formulae Γ. Thus, we need first to normalize the syntactic form of Γ. To achieve this goal, we assume the notion of a *normalization method*.

Definition 6. Let $\mathtt{W}$ be a weighted language. A **normalization method** on $\mathtt{W}$ is a function $\mathtt{n}$ normalizing the syntactic form of the formulae, i.e., $\mathtt{n}$ is a function from $2^{\mathtt{W}}$ to $2^{\mathtt{W}}$.

These normalizations aim to restrict the language to a single formula per equivalence class.

Example 1. Let $\mathtt{n}$ be the normalization of sets of formulae into a set of clauses by removing unnecessary literals, on $\mathbf{wLog}$, and $\Phi = \{\langle \neg(p \rightarrow q \vee \neg r), 0.7\rangle\}, \Psi = \{\langle (p \vee q) \wedge (p \vee \neg q), 0.7\rangle, \langle \neg q \wedge \neg\neg r, 0.7\rangle\} \subseteq \mathbf{wLog}$. Hence we may obtain $\mathtt{n}(\Phi) = \mathtt{n}(\Psi) = \{\langle p, 0.7\rangle, \langle \neg q, 0.7\rangle, \langle r, 0.7\rangle\}$.

This normalization in the argumentation literature is also sometimes called *compilation* and examples include (Amgoud and Doder 2019; Amgoud and David 2021). In the rest of the article, we assume a normalized language for all definitions and examples.

3 Weighted Structured Argumentation

An argument can be seen as a pair consisting of a set of premises and a claim implied by them. Some constraints on the premises and claim are usually considered (Besnard and Hunter 2001). The goal of this section is to extend the notion of argument to a weighted logic.

Definition 7. Let $\mathbf{L} = \langle \mathtt{W}, \vdash, t\rangle$ be a weighted logic. A **weighted argument** on $\mathbf{L}$ is a pair $\langle \Gamma, \alpha\rangle$ such that Γ is a finite subset of $\mathtt{W}$ and $\alpha \in \mathtt{W}$, Γ is consistent, $\Gamma \vdash \alpha$, $\forall\, \Gamma' \subset \Gamma$, $\Gamma' \nvdash \alpha$. Let $\mathsf{Arg}_{\mathbf{L}}$ be the set of all weighted arguments on $\mathbf{L}$.

However, such ideal arguments, whether weighted or not, are rarely seen. In general, humans use enthymemes, i.e., incomplete arguments in which part of the premises is missing, to logically infer the claim. The formal definition and the task of handling enthymemes is investigated in e.g., (Hunter 2007; 2022).

In what follows, we introduce the notion of an approximate weighted argument, which is subject to no constraints other than the structure of its premises/claims. Thus, an enthymeme is a special case of this type of argument, where it is guaranteed that the inference between the premises and the claim does not logically hold.

Definition 8. Let $\mathbf{L} = \langle \mathtt{W}, \vdash, t\rangle$ be a weighted logic. An **approximate weighted argument** on $\mathbf{L}$ is a pair $A = \langle \Gamma, \alpha\rangle$ such that Γ is a finite subset of $\mathtt{W}$ and $\alpha \in \mathtt{W}$. We denote by $\mathsf{aArg}_{\mathbf{L}}$ the set of all approximate weighted arguments on $\mathbf{L}$. An **enthymeme** on $\mathbf{L}$ is an element $E = \langle \Gamma, \alpha\rangle \in \mathsf{aArg}_{\mathbf{L}}$ such that $\Gamma \nvdash \alpha$. We denote by $\mathtt{Enth}_{\mathbf{L}}$ the set of all enthymemes on $\mathbf{L}$.

We now present the running example. In the logical atoms, we retain the core concepts while abstracting away uncertainty, which will instead be captured by associated weights. In this example, the weights are manually assigned to illustrate interesting cases, but they could be automatically estimated in practice based on contextual or external information. This opens the door to learning-based approaches for weight assignment, such as confidence scoring from knowledge extraction systems (Dong et al. 2014; Lajus, Galárraga, and Suchanek 2020).

Example 2. Assuming that: h = "Bob is happy", w = "Bob is wealthy", r = "Bob is a researcher", p = "Bob gives love to people", l = "Bob receives love". Then,

- $E = \langle \{\langle w, 0.7\rangle, \langle r, 0.7\rangle, \langle p, 1\rangle, \langle l, 1\rangle\}, \langle h, 0.7\rangle\rangle$;
- $D_1 = \langle \{\langle r, 0.7\rangle, \langle \neg r \vee h, 0.8\rangle\}, \langle h, 0.7\rangle\rangle$;
- $D_2 = \langle \{\langle p, 1\rangle, \langle l, 1\rangle, \langle \neg p \vee \neg l \vee h, 0.95\rangle\}, \langle h, 0.7\rangle\rangle$;
- $D_3 = \langle \{\langle \neg r, 0.7\rangle, \langle w, 0.7\rangle, \langle \neg w \vee h, 0.8\rangle\}, \langle h, 0.7\rangle\rangle$.

Where $E, D_2 \in \mathtt{Enth}$ are enthymemes, while $D_1 \in \mathsf{Arg}$ is a weighted argument, and $D_3 \in \mathsf{aArg}$ is just an approximate weighted argument (i.e., D_3 is not an enthymeme).

We now formally define an enthymeme decoding as a pair containing the enthymeme and another approximate weighted argument that we refer to as a decoding.

Definition 9. $\mathbf{L} = \langle \mathtt{W}, \vdash, t\rangle$ be a weighted logic. An **enthymeme decoding** on $\mathbf{L}$ is a pair $\langle E, D\rangle \in \mathtt{Enth} \times \mathsf{aArg}$. Intuitively, D is a decoding of the enthymeme E.

In Example 2, we give an enthymeme, and three examples of a decoding. Note, the decoding D_2 is actually an enthymeme because the premises imply h with value 0.95, and the claim is h with value 0.7, and so this mismatch means the premises do not imply exactly the claim. In contrast, the decoding D_3 has premises that imply h with value 0.7, and the claim is h with value 0.7, but the premises are not minimal, and so the D_3 is an approximate weighted argument.

Decodings are intentionally defined without constraints to accommodate real-world scenarios in which imperfect decodings should at least be provided when no perfect decodings are possible. For example, decodings generated by humans or derived from automatically retrieved information may only be approximately coherent (e.g., in decoding D_2, the combined weight of the premises differs from the weight of the claim by 0.25). Our evaluation criteria below are specifically designed to measure the quality of a given enthymeme decoding, i.e., its degree of perfection.

4 Axioms and Criterion Measures

Some decodings of enthymemes may be unreasonable. To identify reasonable ones, six criteria and the concept of criterion measure are introduced.

Definition 10. Let $\mathbf{L} = \langle \mathtt{W}, \vdash, t\rangle$ be a weighted logic. A **criterion measure** on $\mathbf{L}$ is a measure of the quality of an enthymeme decoding with regard to one criterion, i.e., it is a function $\sigma : \mathtt{Enth} \times \mathsf{aArg} \rightarrow [0, 1]$.

We propose 6 criteria for evaluating enthymeme decodings: the flat *inference* of the claim from the premises, the sound *weighting* between the premises and the claim, the *coherence* of the premises, their *minimality*, the *similarity* between the enthymeme and the decoding, and the *granularity* of the decoded premises.

All these criteria are inspired by criteria defined in argumentation (Simari and Loui 1992), or informally discussed in explainable AI (XAI) (Sokol and Flach 2020) or in philosophy (Grice 1975), as elucidated in Figure 2. It is useful also to recall that the notions of argument and explanation are close (Hahn and Tešić 2023), and that XAI's informal properties are originally based on social science research to make algorithmic explanations more natural for users, which is very relevant in the case of enthymeme decoding (context- and user-dependent).

To ensure that the resulting measures exhibit desirable behavior, we use an axiomatic approach. For each criterion,

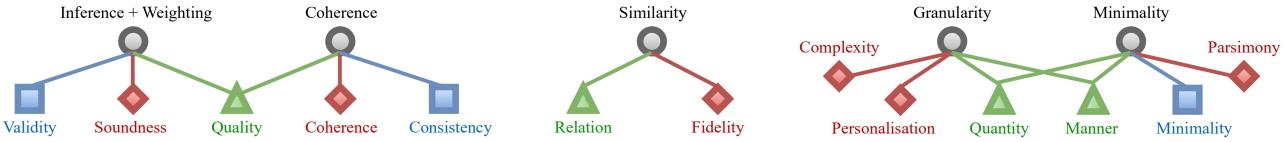

Figure 2: Criteria from argumentation (□), XAI (◇), philosophy (△) which have inspired our decoding criteria (◯).

we take two steps: 1. Define a set of axioms that a criterion measure should satisfy to achieve desirable properties. 2. Propose a measure or family of measures that satisfy these axioms, ensuring their soundness. This proposal to analyze and evaluate each criterion modularly, (rather than attempting to define a single function that incorporates all the criteria), enables a more robust study of the various possible configurations of a quality measure. Indeed, each criterion (i.e., axioms and measures) is examined individually, allowing for simplified composition when constructing the quality measure (see Section 5). This modular approach also facilitates the incorporation of new criteria in our quality measure.

As depicted in Figure 2, to analyze the weighted inference of decoding, we split it into two criteria: first assessing flat inferences, then evaluating weights. This reflects the classic trade-off in multi-criteria decision theory (Mardani et al. 2015) between quantity (number of inferences) and quality (inference weights).

4.1 Criterion of Inference

Axioms. These properties ensure that a measure considers a decoding as reasonable if the premises infer the claim (Ideal version) or the more the premises fully infer the claim, the better the decoding (Increasing version). In Figure 2, this is inspired by validity (argumentation), soundness (XAI), and quality (philosophy).

Definition 11. We denote by $|X|$ the cardinality of X.

Let $\mathbf{L} = \langle \mathtt{W}, \mid\!\sim, t\rangle$ be a weighted logic and σ a criterion measure on $\mathbf{L}$. We say that σ satisfies the axiom **Ideal Inference ($\mathbf{I}_1$)** iff $\forall\, E \in \mathtt{Enth}, \forall\, D = \langle \Delta, \beta\rangle \in \mathtt{aArg}$:

if $\mathtt{flat}(\Delta) \vdash \mathtt{flat}(\beta)$, then $\sigma(E, D) = 1$.

σ satisfies the axioms **Lenient Increasing Inference ($\mathbf{I}_2$)** iff, $\forall\, E \in \mathtt{Enth}, \forall\, D = \langle \Delta, \beta\rangle, D' = \langle \Delta', \beta\rangle \in \mathtt{aArg}$:

if $|\{f : \mathtt{flat}(\Delta) \vdash f \text{ and } \mathtt{flat}(\beta) \vdash f\}| \geq$
$|\{f : \mathtt{flat}(\Delta') \vdash f \text{ and } \mathtt{flat}(\beta) \vdash f\}|$,
then $\sigma(E, D) \geq \sigma(E, D')$.

The axiom **Strict Increasing Inference ($\mathbf{I}_3$)** is defined as above, but $\geq$ is replaced by $>$.

Criterion measure. We illustrate a measure of inference based on the weighted propositional logic (Definition 5), by adapting the dependent finite Cn definition from (David 2021) to handle finite consequences. The definition for $\mathtt{fCn}(\Delta)$ below gives a single formula per equivalence class for all consequences of minimal subsets Γ of Δ that do not add new literals.

Definition 12. Let $\Delta \subseteq \mathtt{wLan}$, and $\mathtt{n}$ a normalization method on $\mathtt{Lan}$, the **Flat Finite Cn** of Δ is

$\mathtt{fCn}(\Delta) = \{\mathtt{n}(f) : \mathtt{flat}(\Delta) \vdash f \text{ s.t. } f \in \mathtt{Lan} \text{ and},$
$\mathtt{lit}(f) \subseteq \mathtt{lit}(\mathtt{flat}(\Gamma)) \text{ s.t. } \Gamma \subseteq \Delta, \mathtt{flat}(\Gamma) \vdash f$
and $\nexists \Gamma' \subset \Gamma$ s.t. $\mathtt{flat}(\Gamma') \vdash f\}$.

This definition captures minimal (without irrelevant literals) consequences derivable from Δ, providing a basis for evaluating whether a decoding's claim is logically entailed by its premises.

Example 3. Let

- $\Delta = \{\langle r, 0.7\rangle, \langle \neg r \vee h, 0.8\rangle\} \subseteq \mathtt{wLan}$;
- $\beta = \langle r \wedge h \wedge x, 0.7\rangle \in \mathtt{wLan}$. Hence, we have:
- $\mathtt{fCn}(\Delta) = \{r, \neg r \vee h, h, r \vee h\}$;
- $\mathtt{fCn}(\beta) = \{r, h, x, r \vee h, r \vee x, h \vee x, r \vee h \vee x\}$.

To finitely represent the semantics of formulae, a classic approach is to use prime implicates (Darwiche and Marquis 2002) for their compactness. However, they are too semantically compact to syntactically extract their overlap, e.g., the prime implicates $\{p, q\}$ and $\{p \vee q\}$ share no common formula despite their semantic link.

The next measure assigns a quality of 1 to decodings that fully infer their claim and reduces its score for each consequence that is not deductible from the premises.

Definition 13. Let $\mathbf{L} = \langle \mathtt{W}, \mid\!\sim, t\rangle$ be a weighted logic. Let $\sigma^{\mathtt{pi}}_{\mathbf{L}}$ the criterion measure on $\mathbf{L}$ called **Proportional Inference**, i.e., $\forall\, E \in \mathtt{Enth}, \forall\, D = \langle \Delta, \beta\rangle \in \mathtt{aArg}$:

$$\sigma^{\mathtt{pi}}_{\mathbf{L}}(E, D) = \frac{|\mathtt{fCn}(\beta)|}{|\mathtt{fCn}(\beta)| + |\mathtt{fCn}(\beta) \setminus \mathtt{fCn}(\Delta)|}$$

We extend the running example with a decoding D_4, to show the different behaviors of the measure.

Example 4. *(Cont. running ex.)* Let $\mathbf{L} = \mathbf{wLog}$.

- $\sigma^{\mathtt{pi}}(E, D_1) = 1 \quad \sigma^{\mathtt{pi}}(E, D_2) = 1 \quad \sigma^{\mathtt{pi}}(E, D_3) = 1$
- let $D_4 = \langle \{\langle r, 0.7\rangle, \langle \neg r \vee h, 0.8\rangle\}, \langle r \wedge h \wedge x, 0.7\rangle\rangle$: $\sigma^{\mathtt{pi}}(E, D_4) = \frac{7}{7+4} = \frac{7}{11} \approx 0.64$.

Proposition 1. $\sigma^{\mathtt{pi}}$ satisfies all Inference axioms.

Some criteria are specifically designed to evaluate a decoding D independently of the original enthymeme E. The Inference criterion is one such example: it focuses solely on the internal coherence of D, namely whether its premises entail its claim, without reference to the source argument. This independence supports a modular evaluation strategy, where each criterion isolates a particular aspect of quality. Such separation is essential for flexible analysis, as it allows us to assess decodings along multiple orthogonal dimensions, which can later be combined or prioritized depending on the needs of the application.

4.2 Criterion of Weighting

Axioms. Weighting ensures minimal difference between the premises and claim weights, with quality decreasing as the gap increases. In Figure 2, this is inspired by validity (argumentation), soundness (XAI), and quality (philosophy).

A weighted consequence operator, use a weight aggregator ($\mathtt{v}$) to infers the weight of a weighted formula from a set of formulae. In our weighted propositional logic, $\mathtt{v}$ is the function $\mathtt{min}$ on the weights of a set of formulae.

Definition 14. Let $\mathtt{W}$ be a weighted language. A **weight aggregator** is a function $\mathtt{v} : 2^{\mathtt{W}} \to [0, 1]$ that assigns a weight to any set of weighted formulae.

Definition 15. Let $\mathbf{L} = \langle \mathtt{W}, \mathtt{\sim}, t\rangle$ be a weighted logic, $\mathtt{v}$ the weight aggregator of $\mathbf{L}$, and σ a criterion measure on $\mathbf{L}$. σ satisfies the axiom **Ideal Weighting ($\mathbf{W}_1$)** iff, $\forall\, E \in \mathtt{Enth}, \forall\, D = \langle \Delta, \beta\rangle, D' = \langle \Delta', \beta'\rangle \in \mathsf{aArg}$:

if $\mathtt{v}(\Delta) = \mathtt{v}(\beta)$, then $\sigma(E, D) = 1$.

Let $\mathtt{abs}(x)$ be the absolute value of x. Similarly, σ satisfies **Lenient Decreasing Weighting ($\mathbf{W}_2$)** iff:

if $\mathtt{abs}(\mathtt{v}(\Delta) - \mathtt{v}(\beta)) \geq \mathtt{abs}(\mathtt{v}(\Delta') - \mathtt{v}(\beta'))$,

then $\sigma(E, D) \leq \sigma(E, D')$.

The axiom **Strict Decreasing Weighting ($\mathbf{W}_3$)** is defined as above but $\geq$ is replaced by $>$ and $\leq$ by $<$.

Criterion measure. We propose a strict version discriminating all variations from the difference.

Definition 16. Let $\mathbf{L} = \langle \mathtt{W}, \mathtt{\sim}, t\rangle$ be a weighted logic. Let $\sigma^{\mathtt{dw}}_{\mathbf{L}}$ the criterion measure on $\mathbf{L}$ called the **Difference Weighting**, i.e., $\forall\, E \in \mathtt{Enth}, \forall D = \langle \Delta, \beta\rangle \in \mathsf{aArg}$:

if $\Delta = \emptyset$, then $\sigma^{\mathtt{dw}}_{\mathbf{L}}(E, D) = 1$; otherwise,

$\sigma^{\mathtt{dw}}_{\mathbf{L}}(E, D) = 1 - \mathtt{abs}(\mathtt{min}[\mathtt{weight}(\Delta)] - \mathtt{weight}(\beta))$.

Example 5. *(Cont. running ex.)* Let $\mathbf{L} = \mathbf{wLog}$.

- $\sigma^{\mathtt{dw}}(E, D_1) = 1, \sigma^{\mathtt{dw}}(E, D_2) = \frac{3}{4}, \sigma^{\mathtt{dw}}(E, D_3) = 1$.

Proposition 2. $\sigma^{\mathtt{dw}}$ satisfies all Weighting axioms.

The Weighting criterion captures how well the premises aligns with the claim. For instance, if the argument's premises infer "Tweety flies" with weight 0.9, but the claim states it at 0.1, there is a meaningful mismatch.

4.3 Criterion of Minimality

Axioms. Decoding should be selective to avoid overwhelming the user (Ideal version); the more information in the premises that is unecessary to infer the claim, the worse the decoding (Decreasing version). Note that if the premises do not imply the claim, then any information is potentially required to infer the claim, thus minimality is not weakened. In Figure 2, this is inspired by minimality (argumentation), parsimony (XAI), and quality/manner (philosophy).

Definition 17. Let $\mathbf{L} = \langle \mathtt{W}, \mathtt{\sim}, t\rangle$ be a weighted logic, and σ a criterion measure on $\mathbf{L}$. We say that σ satisfies the axiom **Ideal Minimality ($\mathbf{M}_1$)** iff $\forall\, E = \langle \Gamma, \alpha\rangle \in \mathtt{Enth}, \forall\, D = \langle \Delta, \beta\rangle \in \mathsf{aArg}$, the following holds:

if $\forall\, \Delta' \subset \Delta, \Delta' \not\mathtt{\sim} \beta$, then $\sigma(E, D) = 1$.

We say that σ satisfies the axiom **Lenient Decreasing Minimality ($\mathbf{M}_2$)** iff, $\forall\, E \in \mathtt{Enth}$, $\forall\, D = \langle \Delta, \beta\rangle, D' = \langle \Delta', \beta\rangle \in \mathsf{aArg}$, the following holds:

if $|\{\Gamma : \Gamma \subset \Delta \text{ s.t. } \Gamma \mathtt{\sim} \beta\}| \geq |\{\Gamma' : \Gamma' \subset \Delta' \text{ s.t. } \Gamma' \mathtt{\sim} \beta\}|$,

then $\sigma(E, D) \leq \sigma(E, D')$.

The axiom **Strict Decreasing Minimality ($\mathbf{M}_3$)** is defined as the point above, but $\geq$ is replaced by $>$, $\leq$ is replaced by $<$, and both sets are non-empty.

Criterion measure. We propose a strategy based on the number of minimal subsets. Recall that we use a normalized language, which allows knowledge to be counted.

Definition 18. Let $\mathbf{L} = \langle \mathtt{W}, \mathtt{\sim}, t\rangle$ be a weighted logic. We denote by $\mathtt{inf}_{\mathbf{L}}$ the function on $2^{\mathtt{W}} \times \mathtt{W}$ such that, $\forall\, \Delta \subseteq \mathtt{W}$, $\forall\, \beta \in \mathtt{W}$, the following holds:

$$\mathtt{inf}_{\mathbf{L}}(\Delta, \beta) = \{\Gamma : \Gamma \subseteq \Delta \text{ and } \Gamma \mathtt{\sim} \beta\}.$$

Let $\sigma^{\mathtt{dm}}_{\mathbf{L}}$ be the criterion measure called the **Divided Minimality**, i.e., $\forall\, E \in \mathtt{Enth}, \forall\, D = \langle \Delta, \beta\rangle \in \mathsf{aArg}$,

if $\mathtt{inf}_{\mathbf{L}}(\Delta, \beta) = \emptyset$, then $\sigma^{\mathtt{dm}}_{\mathbf{L}}(E, D) = 1$;

otherwise, $\sigma^{\mathtt{dm}}_{\mathbf{L}}(E, D) = \dfrac{1}{|\mathtt{inf}_{\mathbf{L}}(\Delta, \beta)|}$.

Example 6. *(Cont. running ex.)* Let $\mathbf{L} = \mathbf{wLog}$.

- $\sigma^{\mathtt{dm}}(E, D_1) = 1, \sigma^{\mathtt{dm}}(E, D_2) = 1, \sigma^{\mathtt{dm}}(E, D_3) = \frac{1}{2}$.

Proposition 3. $\sigma^{\mathtt{dm}}$ satisfies all Minimality axioms.

Importantly, our framework does not impose any specific strategy for generating decodings; it focuses only on evaluating the resulting decodings against a set of quality criteria. While logic-based decodings may naturally satisfy certain criteria such as Minimality, this is not always desirable. In some contexts, such as persuasion, redundancy can serve rhetorical purposes.

4.4 Criterion of Coherence

Axioms. Any explainable system (i.e., decoding) must be consistent with itself (Strong version) or, to go further; they must be consistent with the user's prior knowledge (Weak version). Also, the more subsets of inconsistent formulae a decoding contains, the worse the decoding (Decreasing version). In Figure 2, this is inspired by consistency (argumentation), coherence (XAI), and quality (philosophy).

Definition 19. Let $\mathbf{L} = \langle \mathtt{W}, \mathtt{\sim}, t\rangle$ be a weighted logic and σ a criterion measure on $\mathbf{L}$. We say that σ satisfies the axioms **Ideal Strong Coherence ($\mathbf{C}_1$)**, and **Ideal Weak Coherence ($\mathbf{C}_2$)** iff, $\forall\, E = \langle \Gamma, \alpha\rangle \in \mathtt{Enth}$, $\forall\, D = \langle \Delta, \beta\rangle \in \mathsf{aArg}$, the following first, and second point holds, respectively:

- if Δ is consistent, then $\sigma(E, D) = 1$;
- if $\Delta \cup \Gamma$ is consistent, then $\sigma(E, D) = 1$.

We say that σ satisfies the axiom **Lenient Decreasing Strong Coherence ($\mathbf{C}_3$)**, iff $\forall\, E = \langle \Gamma, \alpha\rangle \in \mathtt{Enth}$, $\forall\, D = \langle \Delta, \beta\rangle, D' = \langle \Delta', \beta\rangle \in \mathsf{aArg}$, the following holds:

if $|\ \{\Phi \subseteq \Delta : \Phi \in \mathtt{inc} \text{ and } \nexists \Psi \subset \Phi \text{ s.t. } \Psi \in \mathtt{inc}\}\ | \geq$
$|\ \{\Phi' \subseteq \Delta' : \Phi' \in \mathtt{inc} \text{ and } \nexists \Psi' \subset \Phi' \text{ s.t. } \Psi' \in \mathtt{inc}\}\ |$

then $\sigma(E, D) \leq \sigma(E, D')$.

The axiom **Strict Decreasing Strong Coherence ($\mathbf{C}_4$)** is defined as above, but $\geq$ is replaced by $>$, and $\leq$ by $<$.
Lenient Decreasing Weak Coherence ($\mathbf{C}_5$) is defined by replacing Δ with $\Delta \cup \Gamma$, and Δ' with $\Delta' \cup \Gamma$.
Strict Decreasing Weak Coherence ($\mathbf{C}_6$) is defined by replacing Δ by $\Delta \cup \Gamma$, Δ' by $\Delta' \cup \Gamma$, $\geq$ by $>$, and $\leq$ by $<$.

The condition of the weak coherence is more restrictive because even if information in the premises of the enthymeme is not used in the decoding, it can prevent a decoding if the latter is inconsistent with it. Hence, consistent decodings may be disallowed. However, from a user point of view, this constraint can be valuable.

Criterion measures. We propose a strategy based on the number of minimal inconsistent subsets.

Definition 20. Let $\mathbf{L} = \langle \mathtt{W}, \mid\!\sim, t\rangle$ be a weighted logic, $\forall\, E = \langle \Gamma, \alpha\rangle \in \mathtt{Enth}, \forall\, D = \langle \Delta, \beta\rangle \in \mathtt{aArg}$, we define
$\mathtt{nb\_sInc}(E, D) = |\{\Phi \subseteq \Delta : \Phi \in \mathtt{inc}, \nexists \Psi \subset \Phi : \Psi \in \mathtt{inc}\}|$
$\mathtt{nb\_wInc}(E, D) = |\{\Phi \subseteq \Delta \cup \Gamma : \Phi \in \mathtt{inc}, \nexists \Psi \subset \Phi : \Psi \in \mathtt{inc}\}|$
We denote by $\sigma^{\mathtt{dsc}}_{\mathbf{L}}$ the criterion measure on $\mathbf{L}$ called **Divided Strong Coherence**, and by $\sigma^{\mathtt{dwc}}_{\mathbf{L}}$ the criterion measure on $\mathbf{L}$ called **Divided Weak Coherence**:

$$\sigma^{\mathtt{dsc}}_{\mathbf{L}}(E, D) = \frac{1}{1 + \mathtt{nb\_sInc}(E, D)}$$

$$\sigma^{\mathtt{dwc}}_{\mathbf{L}}(E, D) = \frac{1}{1 + \mathtt{nb\_wInc}(E, D)}$$

Example 7. *(Cont. running ex.)* Let $\mathbf{L} = \mathbf{wLog}$.

- $\sigma^{\mathtt{dsc}}(E, D_1) = \sigma^{\mathtt{dwc}}(E, D_1) = 1$;
- $\sigma^{\mathtt{dsc}}(E, D_2) = \sigma^{\mathtt{dwc}}(E, D_2) = 1$;
- $\sigma^{\mathtt{dsc}}(E, D_3) = 1$, and $\sigma^{\mathtt{dwc}}(E, D_3) = \frac{1}{2}$.

Proposition 4. $\sigma^{\mathtt{dsc}}$ satisfies all Coherence axioms. $\sigma^{\mathtt{dwc}}$ satisfies $\mathbf{C}_2$, $\mathbf{C}_5$ and $\mathbf{C}_6$.

4.5 Criterion of Similarity

Axioms. Adjusting an explanation to users requires modeling their background knowledge as much as possible, i.e., a decoding is preferable when it uses as much information as possible from the enthymeme (increasing similarity) and a minimum of new information (decreasing similarity). Moreover, a decoding must be based on the elements present in the enthymeme, aligned with its premises (preservation) and claim (preservation). In Figure 2, this is inspired by fidelity (XAI), and relation (philosophy).

Definition 21. Let $\mathbf{L} = \langle \mathtt{W}, \mid\!\sim, t\rangle$ be a weighted logic, and σ a criterion measure on $\mathbf{L}$. We say that σ satisfies the axiom **Lenient Similarity ($\mathbf{S}_1$)** iff, $\forall\, E = \langle \Gamma, \beta\rangle \in \mathtt{Enth}, \forall\, D = \langle \Delta, \beta\rangle, D' = \langle \Delta', \beta\rangle \in \mathtt{aArg}$,

if $a \geq a', b \leq b',$ then $\sigma(E, D) \geq \sigma(E, D')$,

where $a = |\Delta \cap \Gamma|$, $a' = |\Delta' \cap \Gamma|$, (common information)
$b = |\Delta \setminus \Gamma|$, $b' = |\Delta' \setminus \Gamma|$. (distinct information)

σ satisfies the axioms **Strict Increasing Similarity ($\mathbf{S}_2$)**, and **Strict Decreasing Similarity ($\mathbf{S}_3$)** iff the following first, second, and third point holds, respectively:

- if $a > a', b \leq b',$ then $\sigma(E, D) > \sigma(E, D')$;
- if $a \geq a' > 0,\ b < b',$ then $\sigma(E, D) > \sigma(E, D')$.

Definition 22. Let $\mathbf{L} = \langle \mathtt{W}, \mid\!\sim, t\rangle$ be a weighted logic, and σ a criterion measure on $\mathbf{L}$. σ satisfies the axioms **Premises Preservation ($\mathbf{S}_4$)**, and **Claim Preservation ($\mathbf{S}_5$)** iff, $\forall\, E = \langle \Gamma, \alpha\rangle \in \mathtt{Enth}$, $\forall\, D = \langle \Delta, \beta\rangle \in \mathtt{aArg}$, the following first, and second point holds, respectively:

- if $\Delta \cap \Gamma = \emptyset,$ then $\sigma(E, D) = 0$;
- if $\alpha \neq \beta,$ then $\sigma(E, D) = 0$.

Criterion measures. We propose syntactic similarity measure from the literature to deal with similarity.

Tversky's ratio model (Tversky 1977) is a general similarity measure which encompasses different well known similarity measure such as (Jaccard 1901), (Dice 1945), (Sørensen 1948), (Anderberg 1973) and (Sneath, Sokal, and others 1973). These measures have been studied in the literature to evaluate arguments in propositional logic (Amgoud and David 2018; Amgoud, David, and Doder 2019) and first-order logic (David, Delobelle, and Mailly 2023).

Definition 23. Let $\mathtt{W}$ be a weighted language, $\Gamma, \Delta \subseteq \mathtt{W}$, and $x, y \in (0, +\infty)$. Let $\mathtt{Tve}_{x,y}(\Gamma, \Delta)$ the xy**-Tversky Measure**, defined by:

$$\mathtt{Tve}_{x,y}(\Gamma, \Delta) = \begin{cases} 1 & \text{if } \Gamma = \Delta = \emptyset; \\ \dfrac{a}{a + (x \times b) + (y \times c)} & \text{otherwise,} \end{cases}$$

where $a = |\Gamma \cap \Delta|$, $b = |\Gamma \setminus \Delta|$, and $c = |\Delta \setminus \Gamma|$.

The above classic measures can be obtained with $\alpha = \beta = 2^{-n}$. The Jaccard measure is obtained with $n = 0$ (i.e., $\mathtt{Tve}_{1,1} = \mathtt{jac}$), Dice with $n = 1$ (i.e., $\mathtt{Tve}_{0.5,0.5} = \mathtt{dic}$), Sorensen with $n = 2$ (i.e., $\mathtt{Tve}_{0.25,0.25} = \mathtt{sor}$), Anderberg with $n = 3$ (i.e., $\mathtt{Tve}_{0.125,0.125} = \mathtt{and}$), and Sokal and Sneah 2 with $n = -1$ (i.e., $\mathtt{Tve}_{2,2} = \mathtt{ss2}$). Similarity measures taking into account the structure of weighted formulae would be significant in improving accuracy.

Definition 24. Let $\mathbf{L} = \langle \mathtt{W}, \mid\!\sim, t\rangle$ be a weighted logic, and $x, y \in (0, +\infty)$. We denote by $\sigma^{\mathtt{ts}}_{\mathbf{L}xy}$ the criterion measure on $\mathbf{L}$ called the xy**-Tversky Similarity** on x and y, i.e., $\forall\, E = \langle \Gamma, \alpha\rangle \in \mathtt{Enth}, \forall\, D = \langle \Delta, \beta\rangle \in \mathtt{aArg}$,

$$\sigma^{\mathtt{ts}}_{\mathbf{L}xy}(E, D) = \mathtt{Tve}_{x,y}(\Gamma, \Delta) \times \mathtt{Tve}_{x,y}(\alpha, \beta).$$

A similarity score of 1 means the decoding matches the enthymeme exactly, but since enthymemes are incomplete, a good decoding should never achieve this.

Example 8. *(Cont. running ex.)* Let $\mathbf{L} = \mathbf{wLog}$.

- $\sigma^{\mathtt{ts}}_{\mathtt{and}}(E, D_1) = \frac{1}{1.5}$, and $\sigma^{\mathtt{ts}}_{\mathtt{jac}}(E, D_1) = \frac{1}{5}$;
- $\sigma^{\mathtt{ts}}_{\mathtt{and}}(E, D_2) = \frac{2}{2.375}$, and $\sigma^{\mathtt{ts}}_{\mathtt{jac}}(E, D_2) = \frac{2}{5}$;
- $\sigma^{\mathtt{ts}}_{\mathtt{and}}(E, D_3) = \frac{1}{1.625}$, and $\sigma^{\mathtt{ts}}_{\mathtt{jac}}(E, D_3) = \frac{1}{6}$.

Proposition 5. For any $x, y \in (0, +\infty)$, $\sigma^{\mathtt{ts}}_{x,y}$ satisfy all Similarity axioms. So do the extended classical measures $\sigma^{\mathtt{ts}}_{\mathtt{jac}}$, $\sigma^{\mathtt{ts}}_{\mathtt{dic}}$, $\sigma^{\mathtt{ts}}_{\mathtt{sor}}$, $\sigma^{\mathtt{ts}}_{\mathtt{and}}$, and $\sigma^{\mathtt{ts}}_{\mathtt{ss2}}$.

4.6 Criterion of Granularity

Axioms. Due to the diversity of users' experience and knowledge, a single explanation cannot meet all expectations. Users should be able to personalize the explanation to their

needs, such as adjusting the granularity of the decoding. We propose two strategies: concise and detailed. In Figure 2, this is inspired by complexity/personalisation (XAI), and quality/manner (philosophy).

Definition 25. Let $\mathbf{L} = \langle \mathtt{W}, \mid\!\sim, t\rangle$ be a weighted logic, and σ a criterion measure on $\mathbf{L}$. We say that σ satisfies the axiom **Lenient Concise Granularity ($\mathbf{G}_1$)** iff, $\forall\, E \in \mathtt{Enth}, \forall\, D = \langle \Delta, \beta\rangle, D' = \langle \Delta', \beta'\rangle \in \mathsf{aArg}$,

$$\text{if } |\Delta| \leq |\Delta'|, \text{ then } \sigma(E, D) \geq \sigma(E, D').$$

σ satisfies the axioms **Strict Concise Granularity ($\mathbf{G}_2$)**, **Lenient Detailed Granularity ($\mathbf{G}_3$)**, and **Strict Detailed Granularity ($\mathbf{G}_4$)** iff the following first, second, and third point holds, respectively:

- if $|\Delta| < |\Delta'|$, then $\sigma(E, D) > \sigma(E, D')$;
- if $|\Delta| \leq |\Delta'|$, then $\sigma(E, D) \leq \sigma(E, D')$;
- if $|\Delta| < |\Delta'|$, then $\sigma(E, D) < \sigma(E, D')$.

Criterion measures. Let us examine the granularity measure, which favors concise decodings.

Definition 26. Let $\mathbf{L} = \langle \mathtt{W}, \mid\!\sim, t\rangle$ be a weighted logic. We denote by $\sigma^{\mathtt{cg}}_{\mathbf{L}}$ the criterion measure on $\mathbf{L}$ called the **Concise Granularity**, i.e., $\forall\, E \in \mathtt{Enth}, \forall\, D = \langle \Delta, \beta\rangle \in \mathsf{aArg}$, the following holds:

$$\sigma^{\mathtt{cg}}_{\mathbf{L}}(E, D) = \frac{1}{|\Delta| + 1}.$$

Example 9. *(Cont. running ex.)* Let $\mathbf{L} = \mathbf{wLog}$.

- $\sigma^{\mathtt{cg}}(E, D_1) = \frac{1}{3}, \sigma^{\mathtt{cg}}(E, D_2) = \frac{1}{4}, \sigma^{\mathtt{cg}}(E, D_3) = \frac{1}{4}$.

Next, let us see the dual version.

Definition 27. Let $\mathbf{L} = \langle \mathtt{W}, \mid\!\sim, t\rangle$ be a weighted logic, $\sigma^{\mathtt{dg}}_{\mathbf{L}}$ be the criterion measure on $\mathbf{L}$ called the **Detailed Granularity**, $\forall\, E \in \mathtt{Enth}, D = \langle \Delta, \beta\rangle \in \mathsf{aArg}$,

$$\sigma^{\mathtt{dg}}_{\mathbf{L}}(E, D) = 1 - \frac{1}{|\Delta| + 1}.$$

Example 10. *(Cont. running ex.)* Let $\mathbf{L} = \mathbf{wLog}$.

- $\sigma^{\mathtt{dg}}(E, D_1) = \frac{2}{3}, \sigma^{\mathtt{dg}}(E, D_2) = \frac{3}{4}, \sigma^{\mathtt{dg}}(E, D_3) = \frac{3}{4}$.

Proposition 6. $\sigma^{\mathtt{cg}}$ satisfies the Granularity axioms $\mathbf{G}_1$, $\mathbf{G}_2$ while $\sigma^{\mathtt{dg}}$ satisfies $\mathbf{G}_3$, $\mathbf{G}_4$.

4.7 Insights on Axioms and Measures

The satisfaction of some axioms ensures that a decoding constitutes a valid argument (as defined in Definition 7).

Proposition 7. Let $\mathbf{L} = (\mathtt{W}, \mid\!\sim, t)$ be a weighted logic, $\sigma_1, \sigma_2, \sigma_3, \sigma_4$ be criterion measures on $\mathbf{L}$, each satisfying one of the following axioms respectively: Ideal Inference, Ideal Weighting, Ideal Strong Coherence, and Ideal Minimality. Let $E \in \mathtt{Enth}$ and $D \in \mathsf{aArg}$. If $D \in \mathsf{Arg}$, then $\sigma_1(E, D) = \sigma_2(E, D) = \sigma_3(E, D) = \sigma_4(E, D) = 1$.

A set of **axioms** is **compatible** if and only if there exists a criterion measure that satisfies all of them. Otherwise, the axioms are **incompatible**.

Theorem 1. For the proposed axioms, the set of all Lenient axioms is compatible with either, the set of all Ideal axioms, or with the set of all Preservation ones; i.e., $\{\mathbf{I}_2, \mathbf{W}_2, \mathbf{M}_2, \mathbf{C}_3, \mathbf{C}_5, \mathbf{S}_1, \mathbf{G}_1, \mathbf{G}_3\}$, with either $\{\mathbf{I}_1, \mathbf{W}_1, \mathbf{M}_1, \mathbf{C}_1, \mathbf{C}_2\}$, or $\{\mathbf{S}_4, \mathbf{S}_5\}$ is compatible. For each criterion $X \in \{\mathbf{I}, \mathbf{W}, \mathbf{M}, \mathbf{C}, \mathbf{S}\}$, the set of axioms for X is compatible. For criterion $\mathbf{G}$, each of $\{\mathbf{G}_1, \mathbf{G}_2\}$, $\{\mathbf{G}_3, \mathbf{G}_4\}$ and $\{\mathbf{G}_1, \mathbf{G}_3\}$ is compatible; all other non-singleton subsets of the axioms of $\mathbf{G}$ is incompatible.

Theorem 2. For the proposed axioms, for each Strict axiom X of any criterion, and for each axiom Y of any other criterion, $\{X, Y\}$ is incompatible. For each Ideal axiom I of any criterion, and for each Preservation axiom P for the Similarity criterion, $\{I, P\}$ is incompatible.

Next we consider implication between axioms. An **axiom ax_1 implies** (i.e., $\rightarrow$) an axiom ax_2 iff for all measures σ, if σ satisfies ax_1, then σ satisfies ax_2.

Theorem 3. For Coherence criterion, the Strong axioms imply Weak ones; for any criterion, any Strict axiom implies its Lenient version; and any Ideal axiom implies its Lenient version i.e., applying also the transitivity we obtain $\mathbf{I}_3 \rightarrow \mathbf{I}_2$, $\mathbf{I}_1 \rightarrow \mathbf{I}_2$; $\mathbf{W}_3 \rightarrow \mathbf{W}_2$, $\mathbf{W}_1 \rightarrow \mathbf{W}_2$; $\mathbf{M}_3 \rightarrow \mathbf{M}_2$, $\mathbf{M}_1 \rightarrow \mathbf{M}_2$; $\mathbf{C}_1 \rightarrow \mathbf{C}_2$, $\mathbf{C}_1 \rightarrow \mathbf{C}_3$, $\mathbf{C}_1 \rightarrow \mathbf{C}_5$, $\mathbf{C}_2 \rightarrow \mathbf{C}_5$, $\mathbf{C}_4 \rightarrow \mathbf{C}_3$, $\mathbf{C}_6 \rightarrow \mathbf{C}_5$, $\mathbf{C}_3 \rightarrow \mathbf{C}_5$, $\mathbf{C}_4 \rightarrow \mathbf{C}_6$, $\mathbf{C}_4 \rightarrow \mathbf{C}_5$; $\mathbf{S}_2 \rightarrow \mathbf{S}_1$, $\mathbf{S}_3 \rightarrow \mathbf{S}_1$; $\mathbf{G}_2 \rightarrow \mathbf{G}_1$, $\mathbf{G}_4 \rightarrow \mathbf{G}_3$.

We show next that our criterion measures do not overlap. A **criterion measure** is **orthogonal** to a set of criteria if the measure fails to satisfy any axiom for this set of criteria. Let Ω be the set of criteria of this paper.

Corollary 1. If a criterion measure σ for $X \in \Omega$ satisfies a Strict axiom for X, then σ is orthogonal to $\Omega \setminus \{X\}$.

Corollary 2. For any criterion $X \in \Omega$, any criterion measure σ for X, σ is orthogonal to $\Omega \setminus \{X\}$.

These results are fundamental, as they stress that a modular approach is necessary for a deep analysis of each criterion, as no measure can satisfy all the axioms. Moreover, strict axioms are key to prevent redundancy among measures, thus avoiding over-valuing certain aspects.

5 Quality Measure

Criterion measures are designed to evaluate different aspects of the quality of an enthymeme decoding. To assess overall quality, we combine these values into a single one. To do that we use an **aggregation function** $\oplus$, defined as following: $\oplus : [0, 1]^n \rightarrow [0, 1]$, where $n \in \mathbb{N}$.

Definition 28. Let $\mathbf{L} = \langle \mathtt{W}, \mid\!\sim, t\rangle$ be a weighted logic, $\mathtt{C} = \langle \sigma_1, \ldots, \sigma_k\rangle$ a sequence of criterion measures on $\mathbf{L}$, and $\oplus$ an aggregation function. We denote by $\mathtt{Q}^{\mathtt{C}}_{\oplus}$ the **quality measure** based on $\mathtt{C}$ and $\oplus$, i.e., the function on $\mathtt{Enth} \times \mathsf{aArg}$ such that, $\forall\, E \in \mathtt{Enth}$, $\forall\, D \in \mathsf{aArg}$, the following holds:

$$\mathtt{Q}^{\mathtt{C}}_{\oplus}(E, D) = \oplus\langle \sigma_1(E, D), \ldots, \sigma_k(E, D)\rangle$$

We propose a sequence of criterion measures because we believe that the order can be significant, especially for future work involving the assignment of specific properties or coefficient of importance depending on the different criteria. In

certain contexts, a subset of criteria may have the property of being "mandatory", meaning that if their quality drops below 1 (or a threshold), the overall quality becomes zero. For example, in decoding a scientific argument, generally the information is assumed to be verified, so decodings that are inconsistent with the enthymeme are deemed invalid. In contrast, when decoding a political argument, where fallacies are more likely, a decoding that is incoherent with the original enthymeme may be permissible.

Let see some examples of aggregation function. For a sequence of values $T = \langle v_1, \ldots, v_k \rangle$, the aggregation function $\texttt{av}(T)$ computes the average of the values in T, and $\texttt{pr}(T)$ returns the product of all the values in T.

Let us see now an example of a sequence of criterion measures, the Jaccard Coherently Weak Concise, defined as:

$$\texttt{jcwc} = \langle \sigma^{\texttt{pi}},\ \sigma^{\texttt{dw}},\ \sigma^{\texttt{dm}},\ \sigma^{\texttt{dwc}},\ \sigma^{\texttt{ts}}_{\texttt{jac}},\ \sigma^{\texttt{cg}} \rangle$$

Continuing our example, we analyze the best decoding for the enthymeme E explaining why Bob is happy.

$\mathtt{Q}^{\texttt{jcwc}}_{\texttt{av}}(E, D_1) =$	$\texttt{av}(1, 1, 1, 1, \frac{1}{5}, \frac{1}{3})$	$\approx$	**0.756**;
$\mathtt{Q}^{\texttt{jcwc}}_{\texttt{av}}(E, D_2) =$	$\texttt{av}(1, \frac{3}{4}, 1, 1, \frac{2}{5}, \frac{1}{4})$	$\approx$	0.733;
$\mathtt{Q}^{\texttt{jcwc}}_{\texttt{av}}(E, D_3) =$	$\texttt{av}(1, 1, \frac{1}{2}, \frac{1}{2}, \frac{1}{6}, \frac{1}{4})$	$\approx$	0.569.
$\mathtt{Q}^{\texttt{jcwc}}_{\texttt{pr}}(E, D_1) =$	$\texttt{pr}(1, 1, 1, 1, \frac{1}{5}, \frac{1}{3})$	$\approx$	0.067;
$\mathtt{Q}^{\texttt{jcwc}}_{\texttt{pr}}(E, D_2) =$	$\texttt{pr}(1, \frac{3}{4}, 1, 1, \frac{2}{5}, \frac{1}{4})$	$=$	**0.075**;
$\mathtt{Q}^{\texttt{jcwc}}_{\texttt{pr}}(E, D_3) =$	$\texttt{pr}(1, 1, \frac{1}{2}, \frac{1}{2}, \frac{1}{6}, \frac{1}{4})$	$\approx$	0.010.

There are at least two possible goals for a quality measure's output: i) extracting the k-best decodings via ranking, or ii) identifying "acceptable" decodings using a threshold.

To identify the best decoding according to $\mathtt{Q}^{\texttt{jcwc}}_{\texttt{av}}$, D_1 ("a researcher is generally happy") ranks first due to its *better alignment of weights*: its support weight ($\texttt{min} = 0.7$) exactly matches the weight of its claim (0.7), whereas D_2 ("being loved makes people happy") shows a less coherent support weight ($\texttt{min} = 0.95$). In contrast, under $\mathtt{Q}^{\texttt{jcwc}}_{\texttt{pr}}$, D_2 ranks higher thanks to a *better similarity score*, and a higher product of similarity and support weight ($\frac{2}{5} \times \frac{3}{4} > \frac{1}{5} \times 1$). This example highlights the crucial role of the aggregation function: even with the same underlying criterion measures, different aggregation strategies can lead to different rankings of the best decoding. When focusing on "acceptable" decodings, we also observe that the acceptance threshold is not fixed (e.g., 0.5), but instead depends on the combination of criterion scores and the chosen aggregation method. Altogether, this illustrates the importance of analyzing the mathematical properties of aggregation functions, as they directly affect the selection of optimal decodings.

One of the key advantages of our quality measure approach lies in its explainability: by decomposing the evaluation of a decoding into distinct, interpretable criteria and an explicit aggregation strategy, we gain insight into *why* a given decoding receives a particular score. This modular structure not only improves transparency, but also allows for fine-grained control and adaptation to different application needs or user preferences.

6 Discussion and Conclusion

This paper presents the first formal framework for evaluating enthymeme decodings using weighted logic and quality criteria. We explore how content can be added or removed from an argument, and introduce a general mechanism for assessing decodings across multiple dimensions. At the core of our approach is a set of criterion measures that satisfy logical axioms, ensuring desirable behavior. The resulting quality measure is parametric and modular, making it adaptable to various goals, users, and contexts.

Understanding and evaluating implicit arguments is essential for building systems capable of robust reasoning in realistic discourse. While argumentative XAI approaches leverage argumentation frameworks to explain decisions or model dialectical interactions (Čyras et al. 2021; Vassiliades, Bassiliades, and Patkos 2021), they typically rely on pre-constructed argument graphs. Their focus is on clarifying how a conclusion was reached and which arguments support or challenge it. In contrast, our work targets an earlier and complementary step: decoding implicit arguments (enthymemes). By enhancing the qualities of these arguments, our approach improves both the interpretability and accuracy of the underlying argumentative structure, providing a more reliable foundation for downstream reasoning.

Unlike approaches such as (Al Khatib et al. 2021), relying on human-annotated criteria (relevance, argumentativeness, content richness, plausibility, bias) and trained classifiers to evaluate completions, our framework operates in a formal logic setting. It supports clear evaluation through logic-based measures with formal guarantees. While some criteria, such as bias or argumentativeness, depend on contextual or pragmatic factors and remain hard to formalize, they point to valuable directions for future integration. Notably, dimensions like rhetorical quality and fallacy degree offer promising extensions. The former captures how effectively a decoding serves communicative goals (e.g., emphasis, structure, audience alignment), while the latter reflects the presence of reasoning flaws such as false dilemmas or emotional appeals. In future work, we will investigate how such dimensions could be captured in our framework in order to identify not only logically valid, but also pragmatically sound and argumentatively robust decodings, better aligned with human judgment. In addition to proposing new criteria, existing ones may also be extended. Similarity could, for example, move beyond formula-to-formula to set-to-set comparisons (e.g., $\{\texttt{minor(x)}\}$ vs. $\{\texttt{human(x)}, \texttt{under18(x)}\}$), or be refined with formula weights to assess local coherence across matched elements, unlike Weighting, which evaluates global weight coherence between the premises and the claim of a decoding.

Finally, we will further analyze quality measures and aggregation functions. Most of the proposed measures are non-parametric, except for Similarity, which can be tuned using thresholds or scaling factors for better practical alignment. Criteria can be defined by users based on their semantic relevance to the task. While configuring measures and aggregation functions may be non-trivial, a promising solution is to learn these parameters from annotated examples.

Acknowledgements

The work by Victor David was supported by the French government, managed by the Agence Nationale de la Recherche under the Plan d'Investissement France 2030, as part of the Initiative d'Excellence d'Université Côte d'Azur under the reference ANR-15-IDEX-01.

References

Al Khatib, K.; Trautner, L.; Wachsmuth, H.; Hou, Y.; and Stein, B. 2021. Employing argumentation knowledge graphs for neural argument generation. In *In Proc. of ACL-IJCNLP*, 4744–4754.

Alsinet, T.; Chesñevar, C. I.; Godo, L.; and Simari, G. R. 2008. A logic programming framework for possibilistic argumentation: Formalization and logical properties. *Fuzzy Sets and Systems* 159(10):1208–1228.

Amgoud, L., and David, V. 2018. Measuring similarity between logical arguments. In *Proc. of KR*, 98–107.

Amgoud, L., and David, V. 2021. Similarity measures based on compiled arguments. In *Proc. of ECSQARU*, 32–44.

Amgoud, L., and Doder, D. 2019. Compilation of logical arguments. In *Proc. of IJCAI*, 1502–1508.

Amgoud, L.; David, V.; and Doder, D. 2019. Similarity measures between arguments revisited. In *Proc. of ECSQARU*, 3–13.

Anderberg, M. R. 1973. *Cluster analysis for applications. Probability and Mathematical Statistics: A Series of Monographs and Textbooks*. Academic Press, Inc., New York.

Arieli, O.; Borg, A.; Hesse, M.; and Straßer, C. 2022. Explainable logic-based argumentation. In *Computational Models of Argument*. IOS Press. 32–43.

Ben-Naim, J.; David, V.; and Hunter, A. 2024. Understanding enthymemes in argument maps: Bridging argument mining and logic-based argumentation. *arXiv preprint arXiv:2408.08648*.

Besnard, P., and Hunter, A. 2001. A logic-based theory of deductive arguments. *Artificial Intelligence* 128:203–235.

Black, E., and Hunter, A. 2012. A relevance-theoretic framework for constructing and deconstructing enthymemes. *Journal of Logic and Computation* 22:55–78.

Boltužić, F., and Šnajder, J. 2016. Fill the gap! analyzing implicit premises between claims from online debates. In *Proc. of the Workshop on Argument Mining (ArgMining)*, 124–133.

Čyras, K.; Rago, A.; Albini, E.; Baroni, P.; and Toni, F. 2021. Argumentative xai: a survey. In *Proc. of IJCAI*, 4392–4399.

Darwiche, A., and Marquis, P. 2002. A knowledge compilation map. *Journal of Artificial Intelligence Research* 17:229–264.

David, V., and Hunter, A. 2025. A logic-based framework for decoding enthymemes in argument maps involving implicitness in premises and claims. In *Proc. of IJCAI*.

David, V.; Delobelle, J.; and Mailly, J.-G. 2023. Similarity measures between order-sorted logical arguments. In *Proc. of JIAF*.

David, V. 2021. *Dealing with Similarity in Argumentation*. Ph.D. Dissertation, Université Paul Sabatier-Toulouse III.

Dice, L. R. 1945. Measures of the amount of ecologic association between species. *Ecology* 26(3):297–302.

Dong, X.; Gabrilovich, E.; Heitz, G.; Horn, W.; Lao, N.; Murphy, K.; Strohmann, T.; Sun, S.; and Zhang, W. 2014. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In *Proceedings of SIGKDD*, 601–610.

Dupin de Saint-Cyr, F. 2011. Handling enthymemes in time-limited persuasion dialogs. In *Proc. of SUM*, volume 6929, 149–162. Springer.

Faure, M. 2010. Rhetoric and persuasion: Understanding enthymemes in the public sphere. *Acta Academica* 42:61–96.

Grice, H. P. 1975. Logic and conversation. In *Speech Acts*. Brill. 41–58.

Habernal, I.; Wachsmuth, H.; Gurevych, I.; and Stein, B. 2017. The argument reasoning comprehension task: Identification and reconstruction of implicit warrants. *arXiv preprint arXiv:1708.01425*.

Hahn, U., and Tešić, M. 2023. Argument and explanation. *Philosophical Transactions of the Royal Society A*.

Han, S.; Schoelkopf, H.; Zhao, Y.; Qi, Z.; Riddell, M.; Zhou, W.; Coady, J.; Peng, D.; Qiao, Y.; Benson, L.; et al. 2022. Folio: Natural language reasoning with first-order logic. *arXiv preprint arXiv:2209.00840*.

Hosseini, S. A.; Modgil, S.; and Rodrigues, O. 2014. Enthymeme construction in dialogues using shared knowledge. In *Proc. of COMMA*, volume 266 of *FAIA*, 325–332. IOS Press.

Hunter, A. 2007. Real arguments are approximate arguments. In *Proc. of AAAI*, volume 7, 66–71.

Hunter, A. 2022. Understanding enthymemes in deductive argumentation using semantic distance measures. In *Proc. of AAAI*, volume 36, 5729–5736.

Jaccard, P. 1901. Nouvelles recherches sur la distributions florale. *Bulletin de la Société Vaudoise des Sciences Naturelles* 37:223–270.

Lajus, J.; Galárraga, L.; and Suchanek, F. 2020. Fast and exact rule mining with amie 3. In *Proc. of ESWC*, 36–52. Springer.

Lalwani, A.; Chopra, L.; Hahn, C.; Trippel, C.; Jin, Z.; and Sachan, M. 2024. NL2FOL: translating natural language to first-order logic for logical fallacy detection. *arXiv preprint arXiv:2405.02318*.

Lang, J.; Liberatore, P.; and Marquis, P. 2003. Propositional independence-formula-variable independence and forgetting. *Journal of Artificial Intelligence Research* 18:391–443.

Lawrence, J., and Reed, C. 2020. Argument mining: A survey. *Computational Linguistics* 45(4):765–818.

Leiva, D. S. O.; Gottifredi, S.; and García, A. J. 2023. Automatic knowledge generation for a persuasion dialogue sys-

tem with enthymemes. *International Journal of Approximate Reasoning* 160:108963.

Lippi, M., and Torroni, P. 2016. Argumentation mining: State of the art and emerging trends. *ACM Transactions on Internet Technology (TOIT)* 16(2):1–25.

Lu, X.; Liu, J.; Gu, Z.; Tong, H.; Xie, C.; Huang, J.; Xiao, Y.; and Wang, W. 2022. Parsing natural language into propositional and first-order logic with dual reinforcement learning. In *Proc. of COLING*, 5419–5431.

Mardani, A.; Jusoh, A.; Nor, K.; Khalifah, Z.; Zakwan, N.; and Valipour, A. 2015. Multiple criteria decision-making techniques and their applications–a review of the literature from 2000 to 2014. *Economic Research-Ekonomska Istraživanja* 28(1):516–571.

Modgil, S., and Prakken, H. 2014. The aspic+ framework for structured argumentation: a tutorial. *Argument & Computation* 5(1):31–62.

Nakov, P.; Corney, D.; Hasanain, M.; Alam, F.; Elsayed, T.; Barrón-Cedeño, A.; Papotti, P.; Shaar, S.; and Martino, G. D. S. 2021. Automated fact-checking for assisting human fact-checkers. *arXiv preprint arXiv:2103.07769*.

Panisson, A. R.; McBurney, P.; and Bordini, R. H. 2022. Towards an enthymeme-based communication framework in multi-agent systems. In Kern-Isberner, G.; Lakemeyer, G.; and Meyer, T., eds., *Proc. of KR*, 267–277.

Rashkin, H.; Choi, E.; Jang, J. Y.; Volkova, S.; and Choi, Y. 2017. Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proc. of EMNLP*, 2931–2937.

Ryu, H.; Kim, G.; Lee, H. S.; and Yang, E. 2024. Divide and translate: Compositional first-order logic translation and verification for complex logical reasoning. *arXiv preprint arXiv:2410.08047*.

Simari, G., and Loui, R. 1992. A mathematical treatment of defeasible reasoning and its implementation. *Artificial Intelligence* 53(2-3):125–157.

Singh, K.; Inoue, N.; Mim, F. S.; Naito, S.; and Inui, K. 2022. Irac: A domain-specific annotated corpus of implicit reasoning in arguments. In *Proc. of LREC*, 4674–4683.

Sneath, P. H.; Sokal, R. R.; et al. 1973. *Numerical taxonomy. The Principles and Practice of Numerical Classification*.

Sokol, K., and Flach, P. 2020. Explainability fact sheets: A framework for systematic assessment of explainable approaches. In *Proc. of Conference on Fairness, Accountability, and Transparency*, 56–67.

Sørensen, T. 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on danish commons. *Biologiske Skrifter* 5:1–34.

Toni, F. 2014. A tutorial on assumption-based argumentation. *Argument & Computation* 5(1):89–117.

Tversky, A. 1977. Features of similarity. *Psychological Review* 84(4):327–352.

Vassiliades, A.; Bassiliades, N.; and Patkos, T. 2021. Argumentation and explainable artificial intelligence: a survey. *The Knowledge Engineering Review* 36:e5.

Wei, K.; Sun, X.; Zhang, Z.; Jin, L.; Zhang, J.; Lv, J.; and Guo, Z. 2022. Implicit event argument extraction with argument-argument relational knowledge. *IEEE Transactions on Knowledge and Data Engineering* 8865–8879.

Xydis, A.; Hampson, C.; Modgil, S.; and Black, E. 2020. Enthymemes in dialogues. In *Proc. of COMMA*, 395–402.

Yang, Y.; Xiong, S.; Payani, A.; Shareghi, E.; and Fekri, F. 2023. Harnessing the power of large language models for natural language to first-order logic translation. *arXiv preprint arXiv:2305.15541*.

Zimanyi, E. 1992. *Incomplete and uncertain information in relational databases*. Ph.D. Dissertation, Université Libre de Bruxelles, Brussels, Belgium.