Dialectical Argumentative Characterisations for Real-world Resource-bounded Agents

Castagna, Federico

Awarding institution:
King's College London

Dialectical argumentative characterisations for real-world resource-bounded agents

Federico Castagna

Department of Informatics
Faculty of Natural, Mathematical & Engineering Sciences
King's College London

A thesis submitted in fulfilment for the
degree of Doctor of Philosophy

2022

To Mina and Mario,
Nini and Remo

Abstract

Real-world interactions involve a constant exchange of information between agents (be they humans or AIs) that are characterized by a limited availability of resources. These dialectical interactions, and the entailed properties, constitute the primary focus of this dissertation. They can be formalized by argumentative models of non-monotonic reasoning that provide real-world dialectical characterisations of arguments by resource-bounded agents. This thesis covers a wide range of implementations for such dialectical methods that span from proof theories to labelling algorithms, from argument schemes to dialogue protocols, and from explainable AI to decision support systems.

The main contributions consist of: (1) the design of (sound and complete) dialectical argument game proof theories and (2) algorithmic procedures for the enumeration of dialectical labellings. (3) The formalisation of Explanation-Question-Response (*EQR*) protocols and specific *EQR argument schemes* yields dialogue specialized in conveying explanations. Upon these results, (4) the presentation of *D-schemes*, i.e., dialectical versions of EQR schemes, allows for EQR dialogue implementations capable of delivering explanations more suited to real-world resource-bounded agents. Finally, (5) a practical code-based implementation shows how a software chatbot can perform (a partial) EQR dialogue to assist users seeking information.

From this thesis stem different possible research directions, theoretical and practical. Focusing on the practical application, the herein automatization of the developed dialogue protocols allows the implementation of software capable of seamlessly enhancing clinical decision support systems by answering patients' clarification needs. Furthermore, by implementing a fully-fledged *EQR* dialogue on the presented bot, it should be possible in the future to improve the software scope and enable a multi-purpose chatbot tailored to different explanation contexts and related functions.

Acknowledgements

This thesis is the product of considerable efforts, everchanging ideas, and shifts in focus. It is the result of the curiosities and the motivations that have pushed me in the past years. It is the outcome of the experiences and the product of all the (bad and good) memories that have accompanied each step of this steep journey. If I were to describe this adventure, I would probably resort to the classical metaphor of the roller-coaster, with its ups and downs, although in the end, you simply learn how to enjoy the ride. Reaching the end of this road has not been easy, but I have been very fortunate in finding supportive and friendly people along the way. I would like to thank them in chronological order.

None of this would have been possible without the steady support from my family, Anna, Giancarlo and Eleonora, to whom I owe a lot (special mention deserve my, always encouraging, uncle Marco).

Besides my first, there is also a second family to whom I am deeply grateful, my friends Andrea, Carlo, Didi, Foppa, Fossa, Italo, Jack, Marco, Matteo, Mirko, and Oggio. Even if we may argue about board games or mangas, they have always been cheering for me.

My gratitude goes also to Marcello D'Agostino, my Master's supervisor, as well as one of the best logicians I know.

Obviously, I cannot forget to mention the 'Lucky Biscuits bunch', especially Andreas and Caroline, precious friends and PhD colleagues.

I am also extremely thankful to my 'saviours' for allowing me to change the topic of my research and for having always helped me throughout this journey: Luca Viganò and Peter McBurney.

Luca is the kind of person upon which you can always rely on. He gave me valuable advice and guided me towards a solution when I needed the most.

Peter is one of the best researchers I have had the pleasure to work with. His keen insights fostered the development of my ideas, later transformed into thesis chapters. He told me to research what picks my interest and follow my intuitions (in the worst-case scenario, there is always the possibility of doing a second PhD “This is something you may think about for your next PhD” - he used to say). I will forever be in his debt for the immense help he gave me.

Simon Wells and Katie Atkinson also deserve a special mention: their sharp comments and detailed feedback contributed to improving and shaping the final version of this thesis.

Last but not least, I want to thank Annalisa. Her bright smile is what lead me to achieve this goal.

Contents

Chapter 1

Introduction

> *"The ability to engage in arguments is essential for humans to understand new problems, to perform scientific reasoning, to express, clarify and defend their opinions in their daily lives."*

Quoted from P.M.Dung's seminal paper [53], this sentence seems to extend a famous passage of Aristotle that compares every human being to an 'animal capable of reasoning [through speaking]' (i.e., '$\zeta\omega o\nu\ \lambda o\gamma o\nu\ \varepsilon\chi\omega\mu$' [12]). That is to say, arguing and providing the rationales underpinning one's own beliefs via exchange of *arguments* fulfils a fundamental role in human reasoning. Indeed, since Aristotle's Organon [4, 118] and its considerable influence on the history of Western thought, rich scholarly literature has been investigating the intertwined notions of arguments, reasoning, and logic. For example, Walton claimed that *"logic is the evaluation of reasoning in arguments"* [135], whereas Mercier and Sperber emphasised the idea that reasoning evolved to produce and evaluate arguments [93]. Trying to consolidate possessed information by formulating reasons (arguments) that challenge or defend them is an ordinary procedure in which humans engage. This process is not only common but even necessary: how would it be possible, otherwise, to decide what to believe or trust without being misled by a non-reliable source of information? This 'scaffolding' (as defined in [95]) role of dialogues and arguments can be seen in both social and lone thinking practices where the reasoner(s) evaluates the possessed information by constructing counter-arguments that assess their acceptability. That is to say, every reasoning process entails a dialogue (even if it is just an imaginary dialogue that a person makes 'within himself/herself'), and every dialogue entails arguments irrespective of the type of interacting agents: be they humans and AIs, among themselves and with humans. Thanks to its important role, argumentation has thrived as a rich, interdisciplinary area of research spanning philosophy, linguistics, psychology, and artificial intelligence [57, 126, 15]. Able to characterize a promising paradigm for modelling reasoning in the presence of conflict and uncertainty, formal-logical accounts of argumentation theory have come to be increasingly central as a core study within Artificial Intelligence. According to such theory, in order to determine if a piece of information is acceptable, it will suffice to prove that the argument (in which the considered information is embedded) is justified under specific semantics. Dung's abstract argumentation framework (AF) [53] has been considered the formalism from which stemmed most of the subsequent studies in this fruitful research field. Nevertheless, although a plethora of works has successfully shown various additions and instantiations of Dung's abstract AFs and achieved different goals [54, 106, 98], none of these approaches managed to provide a full rational account for real-world resource-bounded agents. Undoubtedly, the introduction of rationality postulates [27, 28], as well as desiderata for practical applications [56], have enabled preclusion of counter-intuitive results in AFs instantiations. However, such requirements may demand a consumption of resources that far exceed those available to real-world agents.

Dialectical Classical Logic Argumentation (Dialectical Cl-Arg) is an approach that provides real-world dialectical characterisations of Classical Logic arguments by resource-bounded agents. This method satisfies the rationality postulates and practical desiderata (under minimal requirements) and revolves around the core notion of dialectical defeats. Such defeats enable argumentative interactions more aligned with the dialectical reasoning of real-world agents. An interesting feature of Dialectical Cl-Arg is that it makes an epistemic distinction between committed premises and suppositions assumed true for the sake of the argument. This distinction, which efficiently solves the "foreign commitment problem" [i], allows an argument to be challenged by supposing its premises, outlining in this way a dialectical move typical of real-world dialogues.

D-ASPIC$^+$ is a general framework for argumentation considered halfway between Dung's abstract approach and its concrete instantiations. Similarly to Dialectical Cl-Arg, although from a higher-level perspective, D-ASPIC$^+$ provides a full rational account of arguments for real-world resource-bounded agents.

These two dialectical approaches represent the formal tools used to tackle the research questions around which the thesis has been structured.

[i] As extensively explained in [29], the foreign commitment problem is the issue that arises in dialogical applications when agents are forced to commit to the premises of their interlocutors in order to challenge their arguments.

1.1 Research Objectives

The current paradigm discerns intelligent agents as those entities that receive percepts from the environment and perform actions [109] guided by their goals. This broad description comprises simple and complex agents, each capable of learning and exchanging information. Human beings belong to a subset of this category and share membership in this class with any other entity that interacts with them or their surroundings: together they are denoted as real-world agents. A common feature associated with this group is the scarce availability of resources, here intended in the general sense of 'anything that has utility' (thus, also including abstract elements such as knowledge, memory, and time). Given this limitation, real-world agents regularly exchange information among themselves and their environment in order to conduct their functions. Although a large corpus of studies in argumentation and dialogue formalisms should prove their adequacy in modelling such communication by logically encoding the involved data, this is not always the case. The main problem is the lack of consideration of the resource constraints that characterise real-world entities: the majority of the research in the field of argumentation disregards such a feature, especially those approaches that logically instantiate arguments.

The gap that the present thesis will therefore try to target can be summarized by the following overarching question:

- *Can we envisage and design logically structured argumentative proof-theoretical models of dialectical interactions that approximate real-world resource-bounded agent reasoning?*

In particular, the thesis aims to address the following research questions:

1. (a) Does a proof theoretical account of real-world exchanges of arguments differ from the standard argument games?
   (b) If so, can we still devise a procedure that enjoys soundness and completeness properties?
2. (a) Can the dialectical inconsistencies of arguments be efficiently represented by the standard 3-value labelling method and algorithms?
   (b) Otherwise, can we adapt such a labelling approach to provide sound and complete algorithms more suited to dialectical characterisations of arguments?
3. (a) Can we envisage an argument scheme capable of enhancing explanations specifically addressed to real-world resource-bounded agents?

(b) If so, can we design a dialogue protocol informed by this template and the enhanced explanation it conveys?

(c) Finally, can we outline a practical implementation for such an argument scheme?

1.2 Methodology

Throughout the thesis, two methodologies have been employed: formal (for the majority of the dissertation sections) and practical. The discussed topics involve many theoretical subjects, especially the research conducted from Chapter 3 to Chapter 6, whose contents maintain a high level of abstractness. The required logical manipulations for achieving the desired results revolve around a formalism that includes: definitions, theorems, corollaries, lemmas, and related proofs. Such theoretical tools proved to be enough for the introduction of each abstract element and the demonstration of any related property.

On the other hand, Chapter 7 concerns a concrete application of the findings that results from the previous Chapters and demands a more practical methodology. For this reason, a detailed example and a code-based implementation are provided. The external repository that contains the designed software can be accessed and tested by the reader via the supplied (GitHub) weblink[ii]. The developed program is meant to be interpolated among the architecture of an already existing piece of software which, along with the provided virtual example, testify to the validity of the implementation. In addition, a formal assessment of the introduced algorithm is also given through (an outlined) proof of its soundness and completeness properties.

1.3 Thesis Overview

The structure of the thesis is the following. Chapter 2 presents a brief review of all the relevant literature referenced throughout the dissertation (including any essential formalism and proofs). For the reader's convenience, a table has been positioned at the end of the first paragraph which highlights the connection existing between each reviewed notion and the respective section of the thesis.

Chapter 3 tackles the first research objective and studies a procedure for developing dialectical argument game proof theories tailored for real-world resources-bounded agents. To achieve this, it expressly adapts the standard argument games described in [97] to Dialectical Cl-Arg and analyses the attained outcome. This chapter is arranged

[ii] `https://github.com/FCast07/EQRbot`

into four main parts. The first handles the admissible and preferred Dung's semantics by designing an appropriate protocol for the dialectical games. Similarly, the second section devises a dialectical game protocol for Dung's grounded semantics. The third, instead, provides a list of the properties enjoyed by these new proof theories, whereas the fourth part concludes the chapter with a discussion about possible efficiency improvements for the games.

Chapter 4 focuses on addressing the second research question by working on a suitable modification of the standard 3-value labelling approach to accommodate Dialectical Cl-Arg arguments. After having introduced the required formalism in the first section, the following sections devise algorithms (and their respective optimized version) that procedurally output admissible, preferred and grounded labelling for the Dialectical Cl-Arg labelling approach (denoted as dialectical labelling). Notice that the optimized version of the grounded algorithm is identical to the standard 3-value labelling. This is demonstrated by resorting to a particular definition that, with some ingenuity, takes advantage of a particular issue experienced by each Dialectical Cl-Arg admissible extension. Finally, as an additional assessment of their respective soundness and completeness results, a formal equivalence between Dialectical Cl-Arg related argument games and algorithms is shown.

Chapters 5 to 7 deal with the third research objective. In particular, Chapter 5 studies the EQR scheme, a new type of Argument Scheme (AS) expressly designed to be employed as the core element of an Explanation-Question-Response dialogue. Particularly suited for conveying explanations to real-world agents, the purpose of such AS is to formalise the consequences arising (and the presumptive reasoning leading to them) by acting upon a specific expert opinion. The EQR scheme presents three variants depending on the particular focus provided: assertion (*EQR claim* scheme), endorsement (*EQR endorsement* scheme) or precise endorsement (*EQR endorsed-by-whom* scheme). The chapter is divided into three main sections, each of which illustrates a specific EQR version listing also all the critical questions (CQs) associated. In addition, a different protocol for the Explanation-Question-Response dialogue (shortened as *EQR dialogue*) is informed by every one of the three variants, along with the corresponding attacks (derived from the different sets of CQs). The related protocol syntax and (axiomatic) semantics are also shown in detail.

Chapter 6 builds upon the investigation commenced in the previous chapter and studies a combination of the EQR scheme and D-ASPIC+ arguments. The result (denoted as *D-scheme*) provides variants of the EQR template that, when instantiated, yield fully-fledged D-ASPIC+ arguments, thus better accounting for resource-bounded agents. The remaining sections devise EQR formal dialogue protocols that involve D-scheme instantiations, present an implementation of such dialogues and demonstrate the existing equivalence with the dialectical argument games for the admissible/preferred semantics.

Chapter 7 proposes a practical application of the EQR claim scheme as a template for conveying explanations to patients to assist them to better manage their health conditions. The content of this chapter is organized into three main sections. The first presents a background framework about clinical decision support systems (cDSS) and a brief summary of the research field of Explainable AI (XAI). The second part reviews the already existing argument schemes employed in medical settings. The last section shows instead the benefits provided by the EQR claim scheme when used as a tool for cDSS. Indeed, the instantiations of such a scheme will feed a (purposely engineered) chatbot that will convey the requested explanation following the EQR dialogue protocol.

Lastly, Chapter 8 provides a comprehensive overview of the main results of this thesis, along with a discussion concerning the envisaged future theoretical work and other potential practical applications.

1.4 Summary of Contributions

Here is a brief summary of the main contributions resulting from the extensive investigation of each research question:

- Design of specific argument game proof theories for Dialectical Cl-Arg (i) admissible/preferred and (ii) grounded semantics. The presented protocols (and the optionally provided efficiency enhancements) allow for the unfolding of games that enjoy the soundness and completeness properties.

- Development of (iii) 3-value labelling for Dialectical Cl-Arg and entailed properties. Taking advantage of the new dialectical labelling approach, algorithms for the generation of (iv) dialectical preferred and (v) grounded labellings have been devised.

- Introduction of the (vi) EQR argument scheme (including all of its three variants *claim*, *endorsed* and *endorsed-by-whom*) along with (vii) the Explanation-Question-Response dialogue protocol, i.e., a protocol that revolves around the EQR schemes.

- Integration of the EQR scheme and D-ASPIC$^+$ framework to generate (viii) *D-Scheme* instantiations and the (ix) formal dialogue protocol that employs them as its core element.

- Practical implementation of the EQR *claim* scheme as an (x) explanation template for a clinical decision support system. This implementation is integrated within a (xi) chatbot software capable of delivering explanations (partially) following the previously developed dialogue protocol.

These contributions have currently produced the following peer-reviewed publications:

[32] Federico Castagna. Argument games for dialectical classical logic argumentation. *Online Handbook of Argumentation for AI*, pages 2–6, 2020.

[33] Federico Castagna. A dialectical characterisation of argument game proof theories for classical logic argumentation. *International Workshop on Advances in Argumentation in Artificial Intelligence (AI^3)*, 2021.

[34] Federico Castagna. Labelling procedures for dialectical classical logic argumentation. *Online Handbook of Argumentation for AI*, pages 7-11, 2021.

[38] Federico Castagna, Simon Parsons, Isabel Sassoon, and Elizabeth I. Sklar. Providing explanations via the EQR argument scheme. *In Computational Models of Argument: Proceedings of COMMA 2022*, pages 351–352, 2022.

[35] Federico Castagna. Towards a fully-fledged formal protocol for the explanation-question-response dialogue. *Online Handbook of Argumentation for AI*, pages 17–21, 2022.

[37] Federico Castagna, Alexandra Garton, Peter Mcburney, Simon Parsons, Isabel Sassoon, and Elizabeth Sklar. EQRbot: a chatbot delivering EQR argument-based explanations. *Frontiers in Artificial Intelligence*, 6:39, 2023.

[36] Federico Castagna. Dialectical argument game proof theories for classical logic. *Journal of Applied Logic (IfCoLog)*, 2023.

Chapter 2

Literature Review

This chapter provides the relevant literature and background notions upon which the contents of the thesis are based. Each of the presented topics is indeed particularly relevant to specific chapters, as shown in the following table.

Sections\Chapters	**3**	**4**	**5**	**6**	**7**
Dung's AF	√	√		√	
Argument Games	√				
Dialogues			√	√	√
Labelling		√			
Dialectical Cl-Arg	√	√			
D-ASPIC$^+$				√	
Argument Schemes			√	√	√

2.1 Dung's Argumentation Frameworks

Informal studies on argumentation are underpinned by a rich literary heritage [59], but it is only in the past decades that logic-based models of argumentation have been intensively investigated as core components of AI-driven and Multi-Agents systems [15, 42]. Dung's seminal paper [53] has been the starting point for most of the recent interest and research in the field of abstract argumentation and its argumentative characterisations of non-monotonic inference. Indeed, the main strength of his approach is the simple and intuitive use of arguments as a means to formalise non-monotonic reasoning, also showing how humans handle conflicting information in a dialectical way. In a nutshell, the idea is that correct reasoning is related to the admissibility of a statement: the argument is acceptable (i.e. justified) only if it is defended against any attacks from counter arguments.

In this section, I am going to outline the formal definitions that mainly portray Dung's argumentation approach:

Definition 1 (AFs, Conflict freeness and Acceptability)**.** *An argumentation framework* (AF) *is a pair:*

$$AF = \langle AR, attacks \rangle$$

where AR is a set of arguments, and 'attacks' is a binary relation on AR, i.e. attacks $\subseteq AR \times AR$*. Let* $\mathscr{S} \subseteq AR$*, then:*

- $\mathscr{S}$ *is* conflict free *iff* $\forall X, Y \in \mathscr{S}$*:* $\neg attacks(X,Y)$*;*
- $X \in AR$ *is acceptable wrt* $\mathscr{S}$ *iff* $\forall Y \in AR$ *such that* $attacks(Y,X)$*:* $\exists Z \in \mathscr{S}$ *such that* $attacks(Z,Y)$*.*

In the previous definition, we could describe Z as the argument that *defends* X, thus granting its acceptability. In some AFs, *indirect defences* might also occur. That is to say, giving a sequence of acceptable arguments ending with X and starting with A, it can be the case that X (indirectly) defends A.

Definition 2 (Indirect defence)**.** *Let* $\langle AR, attacks \rangle$ *be an AF, and* $X, A \in AR$*. According to the following recursive definition, an argument X* indirectly defends *an argument A if:*

$a)$ X defends A*;*

$b)$ X defends Z*, and* Z indirectly defends A.

Notice that an unattacked argument is, trivially, defended and indirectly defended by itself.

Definition 3 (Dung's semantics)**.** *Let* $\langle AR, attacks \rangle$ *be an AF, and* $\mathscr{S} \subseteq AR$ *be conflict free. Then:*

- $\mathscr{S}$ *is an* admissible *extension iff* $X \in \mathscr{S}$ *implies X is acceptable w.r.t.* $\mathscr{S}$*;*
- *An admissible extension* $\mathscr{S}$ *is a* complete *extension iff* $\forall X \in AR$*: X is acceptable w.r.t.* $\mathscr{S}$ *implies* $X \in \mathscr{S}$*;*
- *The least complete extension* (*with respect to set inclusion*) *is called the* grounded extension*;*
- *A maximal complete extension* (*with respect to set inclusion*) *is called a* preferred extension.

The notion of complete extension provides the link between preferred extensions (*credulous semantics*), and grounded extension (*sceptical semantics*)[i]. In general for E $\in$ {admissible, complete, preferred, grounded}, X is sceptically or credulously justified under the E semantics if X belongs to *all*, respectively *at least one*, E extensions. In [53]

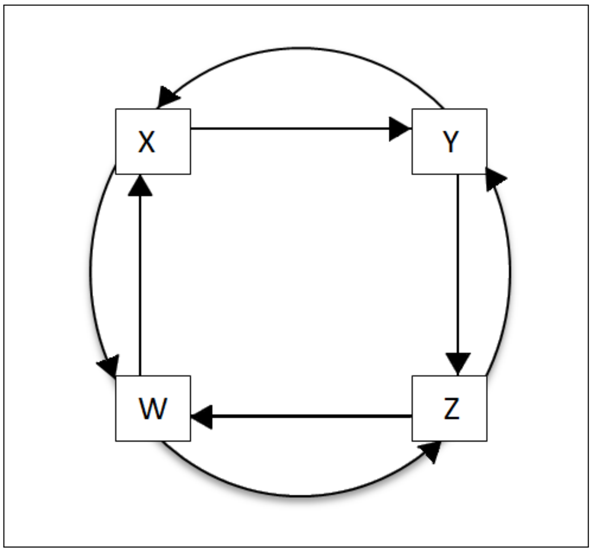

Figure 2.1: An abstract argumentation framework. The arrows represent attack relations between the arguments of the AF.

Dung showed also an important result in relation to the newly introduced argumentation approach: the *fundamental lemma* and its resulting theorem.

Lemma 1 (Fundamental Lemma)**.** *Let $\langle AR, attacks \rangle$ be an AF, and $\mathscr{S} \subseteq AR$ be admissible. Let also $X, X' \in AR$ be arguments which are acceptable w.r.t. $\mathscr{S}$. Then:*

(1) *$\mathscr{S}' = \mathscr{S} \cup \{X\}$ is admissible, and*

(2) *X' is acceptable w.r.t. $\mathscr{S}'$.*

Intuitively, the fundamental lemma guarantees the generation of new admissible extensions by adding acceptable arguments to existing admissible sets. A very interesting consequence of the fundamental lemma is that every admissible extension is a subset of a preferred extension. Formally:

Theorem 1. *Let $\langle AR, attacks \rangle$ be an AF. Then, for each admissible set $\mathscr{S} \subseteq AR$, there exists a preferred extension $\mathscr{S}' \subseteq AR$ such that $\mathscr{S} \subseteq \mathscr{S}'$.*

[i]The grounded extension is sceptical in the sense that it always induces a unique extension of admissible arguments: in case of an unsolvable conflict between two arguments, it leaves both arguments out of the extension. On the other hand, preferred extensions, in the case of an unsolvable conflict between two arguments, build multiple extensions and separate the conflicting arguments.

Another interesting result presented in [53] is the alternative definition of Dung's extensions (i.e. the 'fixpoint semantics') via the AF's *characteristic function* $\mathscr{F}_{AF}$:

$$\mathscr{F}_{AF}(\mathscr{S}) = \{X \mid X \text{ is acceptable w.r.t. } \mathscr{S}\}, \text{where } \mathscr{S} \subseteq \text{AR}$$

For any conflict free $\mathscr{S} \subseteq$ AR, we then have that $\mathscr{S}$ is: admissible iff $\mathscr{S} \subseteq \mathscr{F}_{AF}(\mathscr{S})$; complete iff $\mathscr{S}$ is a fixed point of $\mathscr{F}_{AF}$, i.e., $\mathscr{S} = \mathscr{F}_{AF}(\mathscr{S})$; grounded iff $\mathscr{S}$ is the least fixed point of $\mathscr{F}_{AF}(\mathscr{S})$; preferred iff $\mathscr{S}$ is a maximal fixed point of $\mathscr{F}_{AF}(\mathscr{S})$.

Dung's AF can be extended such that 'preferences' can be taken into account as well. It can be useful, indeed, to have a way of deciding, among two or more conflicting arguments, which ones are preferred, hence, which attacks will succeed as *defeats*. This leads to the formal definition of defeats:

Definition 4 (Defeats)**.** *Let $\langle AR, attacks \rangle$ be an AF. Then 'defeats' $\subseteq$ 'attacks' is the defeat relation defined by the strict partial ordering $\prec$ over AR, such that:*

$$defeats(X,Y) \text{ iff } attacks(X,Y) \text{ and } X \not\prec Y$$

That is to say, an argument X defeats an argument Y if and only if X attacks Y and Y is not strictly preferred over X. In the remainder of the dissertation, $X \Rightarrow Y$ will stand for "$defeats(X,Y)$", and $X \not\Rightarrow Y$ will stand for "$\neg defeats(X,Y)$".

As anticipated, abstract AFs represent general frameworks capable of providing argumentative characterisations of non-monotonic logics[ii]. That is to say, given a set of formulae Δ of some logical language *L*, AFs can be instantiated by such formulae. These instantiations paved the way for a plethora of different investigations concerning the so-called 'structured' argumentation (as opposed to the abstract approach). Studies showed how the conclusions of justified arguments defined by the instantiating Δ are equivalent to those obtained from Δ by the inference relation of the logic *L*.

2.2 Argument Games

Argument game-based proof theories provide procedural structures capable of determining the status of an argument. Given an argumentation framework, these games identify the membership of an argument to a specific extension simulating a conversation between two opposing contenders. The semantics meant to be captured dictates the rules of

[ii]In [53], Dung employs Reiter's Default logic [108] and Pollock's Inductive Defeasible logic [103] as an example of non-monotonic reasoning rendered via abstract argumentation.

the played game, which serves to describe how the players can achieve victory. In other words, these proof theories formalise reasoning processes allowing to (sceptically or credulously) defend the data encoded in an argument via a rule-guided interplay of arguments. Intuitively, they can be seen as fictitious dialogues that an agent performs 'within itself' to evaluate the reliability of a specific piece of information it possesses. Let us now review the fundamental notions of argument games as described in [97]: we will consider these as identifying the *'standard argument games'*. Notice, however, that these definitions have been modified to accommodate defeats rather than attacks among the arguments of an AF.

To begin with, an argument game is played by two players, PRO (for *proponent*) and OPP (for *opponent*), each of which is referred to as the other's 'counterpart'. PRO starts the game by moving an initial argument X that it wants to test. After that, both players take turns in moving arguments against their counterpart's moves. This generates disputes:

Definition 5 (Dispute)**.** *A sequence of moves in which each player moves one argument at a time against its counterpart's move is referred to as a* dispute*. Formally, d = X—Y—Z— $\cdots$ is a dispute, and X—Y denotes a player moving argument Y against an argument X moved by its counterpart (similarly, Y—Z). A* sub-dispute *d' of a dispute d is any sub-sequence of d that starts with the same initial argument. For example, if d = X—Y—Z, then d' = X—Y would be a sub-dispute of d.*

Notice that, to avoid any ambiguity, each argument of a dispute will be labelled with either P or O (that stands for either one of the two players, PRO or OPP). Hence, d = (P)X—(O)Y—(P)Z is a dispute where PRO moves the argument X, followed by Y moved by OPP and countered by another move from PRO, Z.

We can now introduce the notion of the (unique) dispute tree, which represents the 'playing field' of the argument games. In other words, the dispute tree is the data structure that contains all the potential moves (and sequence of moves) available to the players.

Definition 6 (Dispute Tree)**.** *Let $AF = \langle AR, defeats\rangle$ be a finite dialectical argumentation framework, and let* A $\in AR$*. The* dispute tree *induced by* A *in the AF is the (upside-down) tree $\mathscr{T}$ of arguments, such that $\mathscr{T}$'s root node is* A*, every branch of the tree (from root to leaf) is a different dispute, and $\forall X, Y \in AR$: X is a child of Y in $\mathscr{T}$ iff* defeats(X,Y).

For the remainder of this section, we are going to write PRO$(\mathscr{T})$ and OPP$(\mathscr{T})$ to denote the sets of all PRO and OPP arguments in the dispute tree $\mathscr{T}$. Also, LAST(d) will identify the last argument played in a dispute d.

An argument game is said to be won by the proponent only if it has a winning strategy. That is to say, only if it can successfully defend the argument it wants to test (i.e., the root of $\mathscr{T}$) against any possible arguments moved by the opponent. PRO loses otherwise.

Definition 7 (Winning Strategy)**.** *Let $\mathscr{T}$ be the dispute tree induced by* A *in a finite dialectical $AF = \langle AR, defeats \rangle$. Let also d be a dispute in $\mathscr{T}$. Then, a* winning strategy $\mathscr{T}'$ *for* A *is the dispute tree $\mathscr{T}$ pruned in a way such that:*

(7.1) *The set $\mathscr{T}'_D$ of disputes in $\mathscr{T}'$ is a non-empty finite set such that each dispute $d \in \mathscr{T}'_D$ is finite and is won by PRO (i.e., $LAST(d) \in PRO(\mathscr{T})$);*

(7.2) *$\forall d \in \mathscr{T}'_D$, $\forall d'$ such that d' is some sub-dispute of d, $LAST(d') = X$ and $X \in PRO(\mathscr{T})$, then $\forall Y \in OPP(\mathscr{T})$ such that $Y \Rightarrow X$, there is a $d'' \in \mathscr{T}'_D$ such that d'— Y is a sub-dispute of d''.*

Informally, the previous definition states that a winning strategy is the dispute tree $\mathscr{T}$ pruned in a way such that (7.1) $\mathscr{T}'_D$ is a non-empty finite set, its disputes are finite, end with a PRO argument and (7.2) are such that OPP has moved exhaustively (i.e., all the moves that OPP could have moved, had been moved) and also PRO has countered every defeating argument moved by OPP.

Depending on the semantics the game is meant to capture, there are different kinds of protocols that need to be observed. For example, each protocol requires that the players move only one argument at a time during their respective turns. However, the procedure for the admissible/preferred game forbids the repetition of the arguments already played by the opponent in the same dispute, whereas the protocol for the grounded game does the exact contrary. The authors of [97] proved the correspondence existing between the membership of an argument A to an extension E and the presence of a winning strategy for the same argument A. Let Φ_E represents the protocol of an argument game meant to capture an extension E (where E $\in$ {admissible/preferred, grounded}), then the following theorem holds[iii]:

Theorem 2. *Let $AF = \langle AR, defeats \rangle$ be a finite argumentation framework. Then, there exists a Φ_E-winning strategy $\mathscr{T}'$ for A such that the set $PRO(\mathscr{T}')$ of arguments moved by PRO in $\mathscr{T}'$ is conflict free, iff A is in the E extension of the AF.*

[iii]Although the theorem refers to an AF based on defeats rather than attacks, the proofs are the same as the ones presented in [97]

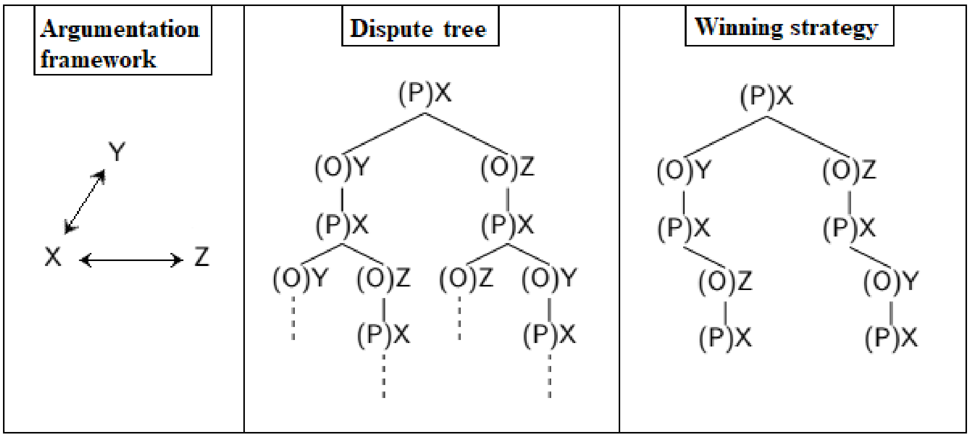

Figure 2.2: The picture depicts the dispute tree $\mathscr{T}$, induced by argument X, and the corresponding winning strategy $\mathscr{T}'$ (assuming that OPP cannot repeat its moves in the same disputes, as per the protocol of the admissible/preferred game).

2.3 Dialogues

The view of computation as distributed cognition and interaction [83] led to the rise of multi-agent systems, where agents are software entities with control over their own execution. This new paradigm required the design of means of communication between such intelligent agents [89]. The choice fell upon formal dialogues, due to their potential expressivity despite being still subject to specific restrictions. Dialogue games are rule-governed interactions among players (i.e., agents) that take turns in making utterances (i.e., moves) following the rules of the game. Unlike argument games, dialogues games represent a higher level of agents interactions where each participant has its own beliefs, goals, desires and knows only a small (possibly, none) amount of information regarding the other players. That is to say, if an argument game corresponds to a dialogue that an agent plays 'within itself', the latter represents a dialogue in a public venue that can simultaneously engage multiple agents.

2.3.1 Types of Dialogue

Dialogue games are commonly categorized according to elements such as: what the participants know, what the participants seek to get from the dialogue, and what the dialogue rules are intended to bring about [21]. The following is an extended list (with no ambition of being exhaustive) of the standard dialogue types presented in [141]:

- *Information-Seeking*: one participant seeks the answer to some question(s) from another participant, who is believed by the first to know the answer(s) (e.g., [72]).

- *Inquiry*: the participants collaborate to answer some question(s) whose answers are not known to any one participant (e.g., [20]).

- *Persuasion*: one participant seeks to persuade another to accept a proposition she does not currently endorse. This can mean that the persuadee holds the opposite or is agnostic about the position put forward by the persuader. (e.g., [105]).

- *Negotiation*: the participants bargain over the division of some scarce resources. If a negotiation dialogue terminates with an agreement, then the resource has been divided in a manner acceptable to all participants. (e.g., [92]).

- *Deliberation*: the participants collaborate (hence, share the responsibility) to decide what action or course of action should be adopted in some situation. Appeals to value assumptions, such as goals and preferences, may influence the agents' deliberation (e.g., [85]).

- *Eristic*: the participants quarrel verbally as a substitute for physical fighting, aiming to vent perceived grievances[iv].

- *Verification*: one participant seeks the answer to some question from another agent. The former wants to verify if the second believes that p (i.e., the proposition with which the dialogue is concerned) is true (e.g., [43]).

- *Query*: one participant always challenges the answer about p from another participant. The former's interest lies more on the second's argument for p rather than if she believes p or not (e.g., [43]).

- *Command*: One participant tells another what to do. If challenged, instructions may be justified, possibly by referencing further actions which the commanded action is intended to enable (e.g., [65]).

- *Education*: One participant wants to teach another something. Unlike information-seeking dialogues, in education dialogues the asking agent does know the answer to the question she is posing (i.e., she is *quizzing* the learner). (e.g., [117]).

- *Discovery*: A new idea arises out of exchanges between participants. Unlike inquiry dialogues, here the focus is on the discovery of something not previously known. The question whose truth is to be ascertained may (or not) emerge in the course of the dialogue (e.g., [86]).

[iv] As stated in [21]: '[...] *eristic dialogues have not been widely studied in computer science'*.

2.3.2 Dialogues Combinations and Control layers

Overall, a dialogue game can be composed of multiple mixtures of dialogues, each of which might be of a different type. Drawing from the classification detailed in [87], we can identify the combination patterns listed below:

> Iteration Let D be a dialogue. The *iteration* of D to its n-fold repetition is also a dialogue, where each occurrence is undertaken until closure, and then is followed immediately by the next occurrence.
>
> Sequencing If D_1 and D_2 are both dialogues, then their *sequence* is also a dialogue, which consists of undertaking D_1 until its closure and then immediately undertaking D_2.
>
> Parallelization If D_1 and D_2 are both dialogues, then conducting them in *parallel* can be also considered as a dialogue, which consists of undertaking D_1 and D_2 simultaneously, until each is closed.
>
> Embedding If D_1 and D_2 are both dialogues, then their *embedding* is also a dialogue, which consists of undertaking D_1, and then switching to dialogue D_2 which is undertaken until its closure, whereupon dialogue D_1 resumes immediately after the point where it was interrupted and continues until closure.

The selection and transitions between different dialogue types can be rendered via a *Control Layer* [87], defined in terms of *atomic dialogue-types* and *control dialogues*. The first element is based upon a finite set of dialogue-types. Control dialogues, instead, are dialogues that have as their discussion subjects not topics, but other dialogues. They include the so-called Commencement and Termination Dialogues in charge of opening (respectively, closing) the subject dialogue, thus contributing to the management of dialogue combinations and their transitions.

Similarly, Walton and Krabbe [141] studied the interaction between multiple dialogues. Their analysis resulted in an informal and general classification of possible dialogue shifts: (a) 'from one type to another', a sequence composed of multiple kinds of dialogues similar to the sequencing combination patterns; (b) 'internal shifts', which occur within the same dialogue type without normative changes; (c) 'from one flavour to another', where the transitions concern only flavours[v]. Shifts can be licit or illicit, the

[v] [...] *we sometimes say that there is one type of dialogue along with a flavour of another type of dialogue, meaning that the one type of dialogue is more explicit and dominant, while the secondary type of dialogue is present in a more subdued or less explicit form.* [141]

second kind is usually associated with fallacies (i.e., a sort of faulty reasoning leading to an invalid argument).

2.3.3 Dialogue Components

Following the study outlined in [89], we can now summarize the three main features of formal dialogues: *syntax*, *semantics* and *pragmatics*.

Syntax The syntax of a language prescribes the instructions on how to form words, phrases and their combinations. Similarly, determining the syntax of a dialogue game involves the specification of the utterances available to the agents and the rules that govern the interactions among such utterances. In addition, it is standard to consider utterances as composed of (1) an inner layer comprising the topics of discussion and (2) an outer (or wrapper) layer comprising the locutions.

Semantics Research concerning dialogue games is at a crossroads between multiple fields of study. Indeed, the interplay among participants in the dialogue is a form of communication that draws from human linguistics knowledge. However, the language must also be necessarily formal while, at the same time, being interpretable by computers. It might then be helpful to consider different types of semantics according to the specific focus, and final deployment, of the dialogue.

1. *Axiomatic*: It defines each locution in terms of its pre and (possibly) post-conditions. Pre-conditions identify what must exist before the locution can be uttered, and post-conditions determine the consequences of such utterance. Public axiomatic approaches enable access to all conditions from each agent in the dialogue, whereas private axiomatic approaches restrict such access to a smaller subset.

2. *Operational*: It considers each locution as a computational instruction that operates successively on the states of some abstract machine. That is to say, it interprets these locutions as commands in some computer programme language.

3. *Denotational*: It assigns, for each element of the language syntax, a relationship to an abstract mathematical entity, its denotation[vi].

[vi]The possible worlds (i.e., Kripkean) semantics is an example of denotational semantics for a logical language [79].

While the dialogue unfolds, agents usually incur *commitments*. That is to say, a speaker asserting the truth of a statement, may be committed to justifying such statement (even if it does not correspond to their real beliefs) against opponents' challenges or retract its assertion. The commitments of all the agents are then tracked and stored in a public database, called a *commitment store*. This position adopts Hamblin's understanding of commitments as purely dialectical obligations [70]. Walton and Krabbe consider instead commitments as obligations connected to a course of action that subsumes under this paradigm also dialectical commitments: "[...] *whose partial strategies assign dialogical actions that center on one proposition"* [141]. On the other hand, Singh [115] and Colombetti [44] regard commitments as *social*, i.e., expressions of wider inter-personal, social, business or legal relationships between the participants, and utterances in a dialogue are a means by which these relationships may be manipulated or modified.

Pragmatics Pragmatics deals with those aspects of the language that do not involve considerations about truth and falsity. Such aspects usually include the *illocutionary force* of the utterances along with *speech acts*, i.e., non-propositional utterances intended to or perceived to change the state of the world[vii]. More precisely, drawing from the analysis of [67] based on relevant literature on the topic, such as Austin's and Searle's works [9, 112, 113], we can define speech acts as 'verbal actions' that accomplish something. Locution would correspond to the simple performance of an utterance, whereas illocution would be the actual intention of the speaker behind the locution meaning. For example, the sentence "You're standing on my foot" uttered in a crowded place is a statement (locution) with the illocutionary force of a command (that is to say, the real meaning is the imperative "move away").

2.3.4 Burden of Proof

One last important aspect considered by the dialogues literature regards the so-called *'burden of proof'*. Multiple authors have investigated the matter and proposed different definitions. For instance, according to Walton [134], the burden of proof is *"an allocation made in reasoned dialogue which sets a strength (weight) of argument required by one side to reasonably persuade the other side."*, whereas van Eemeren and Grootendorst [58] described it as occurring when *"a party that advances the* [dialogue] *standpoint is obliged to defend it if the other party asks him to do so"*. In general, we could say

[vii]An example of analysis of the different pragmatical meanings existing between, say, 'commands' and 'promises' can be found in [88]. Furthermore, the work presented in [91] introduces a specific syntax that accounts for the pragmatical uptake and revocation of utterances over actions.

that participants in a dialogue incur a burden of proof when declaring a proposition as their thesis, thereby compelling them to offer evidence or backing when such a thesis is challenged. In an evenly matched dispute, where the plausibility of the participants' thesis is balanced, any new argument moved may tilt the burden of proof. Nevertheless, in some specific circumstances, the burden of proof can be much heavier on one particular side. As an example, consider any criminal trial: the prosecutor must prove guilt "beyond reasonable doubt" to win her case, which means that she bears a greater encumbrance than her counterpart. Walton further examined how such obligations affect the diverse dialogue types concluding *"If there is no thesis to be proved or cast into doubt* [...]*, there is no burden of proof in that dialogue"* [133]. This quote contributes also to shedding light on the difference between commitments and the burden of proof. Indeed, while the former depends on single utterances and will always occur, the latter concerns the dialogue's thesis and may or may not be present.

2.4 Labellings within Abstract AFs

Besides the most common approach that involves argument extensions, the labelling method is regarded as an alternative way for identifying justified arguments. This method consists of assigning exactly one label, which can either be *IN*, *OUT* or *UNDEC*, to each argument of an AF. The label *IN* indicates that the argument is justified, *OUT* indicates that the argument is overruled, and *UNDEC* indicates that the status of the argument is undecided (since there is not enough information to justify or overrule it). The following Definitions and Theorem, taken from [97], will outline the background of the labelling approach. The general idea is to establish labelling semantics that can easily be mapped to the ones introduced by Dung, showing in this way that the two methods (labelling and extensions) are indeed equivalent.

Definition 8 (Labelling function)**.** *Let $AF = \langle AR, attacks \rangle$ be an argumentation framework.*

- *A labelling is a total function $\mathscr{L}: AR \mapsto \{\text{IN, OUT, UNDEC}\}$;*
- *We define:* $\mathsf{in}(\mathscr{L}) = \{X | \mathscr{L}(X) = \text{IN}\}$; $\mathsf{out}(\mathscr{L}) = \{X | \mathscr{L}(X) = \text{OUT}\}$; $\mathsf{undec}(\mathscr{L}) = \{X | \mathscr{L}(X) = \text{UNDEC}\}$.

Definition 9 (Legal labelling)**.** *Let $\mathscr{L}$ be a labelling for $AF = \langle AR, attacks \rangle$ and $X \in AR$.*

(1) *X is legally* IN *iff X is labelled* IN *and every Y that attacks X is labelled* OUT*;*

(2) *X is legally* OUT *iff X is labelled* OUT *and there is at least one Y that attacks X and Y is labelled* IN*;*

(3) *X is legally* UNDEC *iff X is labelled* UNDEC *and not every Y that attacks X is labelled* OUT*, and there is no Y that attacks X such that Y is labelled* IN.

Definition 10 (Illegal labelling)**.** *Let* $AF = \langle AR, attacks \rangle$*,* $X \in AR$ *and* $\mathtt{l} \in$ {IN, OUT, UNDEC}*. X is said to be illegally* $\mathtt{l}$ *iff X is labelled* $\mathtt{l}$*, and it is not legally* $\mathtt{l}$*.*

Definition 11 (Labelling semantics)**.** *An admissible labelling* $\mathscr{L}$ *is a labelling without arguments that are illegally IN and without arguments that are illegally OUT;*

- *A complete labelling* $\mathscr{L}$ *is an admissible labelling without arguments that are illegally UNDEC* [viii]
- $\mathscr{L}$ *is a grounded labelling iff* $\mathscr{L}$ *is a complete labelling and there does not exist a complete labelling* $\mathscr{L}'$ *such that* $\mathsf{in}(\mathscr{L}') \subset \mathsf{in}(\mathscr{L})$*;*
- $\mathscr{L}$ *is a preferred labelling iff* $\mathscr{L}$ *is a complete labelling and there does not exist a complete labelling* $\mathscr{L}'$ *such that* $\mathsf{in}(\mathscr{L}') \supset \mathsf{in}(\mathscr{L})$*.*

The equivalence between labelling and Dung's semantics has been stated in the following theorem (proved in [25]):

Theorem 3. [25] *Let* $AF = \langle AR, attacks \rangle$ *be an argumentation framework and* $E \subseteq AR$*. For* s $\in$ *{admissible, complete, grounded, preferred}: E is an* s *extension of the AF iff there exists an* s *labelling* $\mathscr{L}$ *with* $\mathsf{in}(\mathscr{L}) = E$*.*

Example 1. *Let us consider the AF of Figure 2.1 and apply the labelling method to its argument in order to determine its extensions. The labelling will be rendered as triples of the form* $(\mathsf{in}(\mathscr{L}), \mathsf{out}(\mathscr{L}), \mathsf{undec}(\mathscr{L}))$*. There are three complete labellings:*

(1) $(\{X, Z\}, \{Y, W\}, \emptyset)$*;*

(2) $(\{Y, W\}, \{X, Z\}, \emptyset)$*;*

[viii] Another equivalent definition has been provided in [30]. Given an AF = $\langle AR, attacks \rangle$ and its labelling $\mathscr{L}$, it holds that $\mathscr{L}$ is a complete labelling iff for each $A \in AR$:

(1) *A* is labelled *IN* iff every attacker of *A* is labelled *OUT*, and

(2) *A* is labelled *OUT* iff *A* has a attacker that is labelled *IN*.

Interestingly, this definition prevents any argument to be illegally *UNDEC* even without explicitly referring to the *UNDEC* label.

(3) $(\emptyset, \emptyset, \{X,Y,Z,W\})$.

(1) *and* (2) *correspond to preferred labelling, whereas* (3) *identifies the grounded labelling* (*i.e., the empty set*).

The labelling method is particularly effective when combined with algorithms specifically designed for arguments' computation. Writing computer programs that involve arguments' status is much simpler when this can be encoded as an *IN*, *OUT*, and *UNDEC* classification problem since it eschews the need to check for each potential extension within the framework. In addition, the literature provides a multitude of available labelling-based algorithms that can be harnessed to serve the specific tasks at hand [41, 99, 97].

2.5 Dialectical Classical Logic Argumentation

2.5.1 Classical Logic Argumentation (Cl-Arg)

Classical Logic commonly refers to a logic that enjoys specific properties such as the law of the excluded middle, law of non-contradiction, duality of logical operators, monotonicity, etc[ix]. In order to clearly understand what is it meant by building AFs instantiated by Classical Logic, it is first necessary to become more familiar with its basic notions.

Definition 12 (Cl-Arg background)**.** [49] *Classical Logic is composed of the following elements:*

- ***Syntax*** *'L' is a first order language consisting of:* (*i*) *the logical operators* $\wedge$, $\vee$, $\supset$, $\neg$, $\forall$, $\exists$; (*ii*) *a countable set of individual variables;* (*iii*) *a* (*possibly empty*) *set of function symbols of various arities, where a function symbols of arity* 0 *is interpreted as a constant symbol;* (*iv*) *a non-empty set of predicate symbols of various arities that includes the* 0*-ary symbol* $\curlywedge$ *interpreted as a patently false atomic formula.*

- ***Consequence relation*** *The first order classical consequence relation* $\vdash_c$ *will be used as well, and writing* $Cn(\Delta)$ *will denote* $\{\alpha | \Delta \vdash_c \alpha\}$. *If* $Cn(\Delta) = L$ *we say that* Δ *is inconsistent; else* Δ *is consistent.*

- ***Complement function*** *Let* ϕ, ψ *be classical well-formed formulae. Then* $\overline{\phi} = \psi$ *if* ϕ *is of the form* $\neg\psi$; *else* $\psi = \neg\phi$.

[ix] An exhaustive introduction to Classical Logic can be found at `http://plato.stanford.edu/entries/logic-classical/`.

- ***Symbols*** *Let Δ be a set of classical formulae. Then* $symbols(\Delta) = \{P \mid P$ *is either a predicate or function symbol in* $\Delta\}$.

- ***Base*** *A* base *$\mathscr{B}$ is a finite set of classical wff such that* $\curlywedge \notin symbols(\mathscr{B})$[x]. *Also, two bases $\mathscr{B}_1$ and $\mathscr{B}_2$ are syntactically disjoint (denoted with $\mathscr{B}_1 \parallel \mathscr{B}_2$) iff* $symbols(\mathscr{B}_1) \cap symbols(\mathscr{B}_2) = \emptyset$.

It is now possible to formally introduce Classical Logic Arguments:

Definition 13 (Classical Logic Arguments)**.** [2, 16, 66]
(Δ, ϕ) is an argument defined by a base of formulae $\mathscr{B}$, if $\Delta \subseteq \mathscr{B}$, and:

1. $\Delta \vdash_c \phi$;
2. $Cn(\Delta) \neq L$ *(Δ is consistent);*
3. $\neg\exists\Delta' \subset \Delta$ *such that* $\Delta' \vdash_c \phi$ *(Δ is said to be 'subset minimal').*

Δ and ϕ are respectively referred to as the premises *($Prem$) and the* conclusion *(Con) of (Δ, ϕ).*

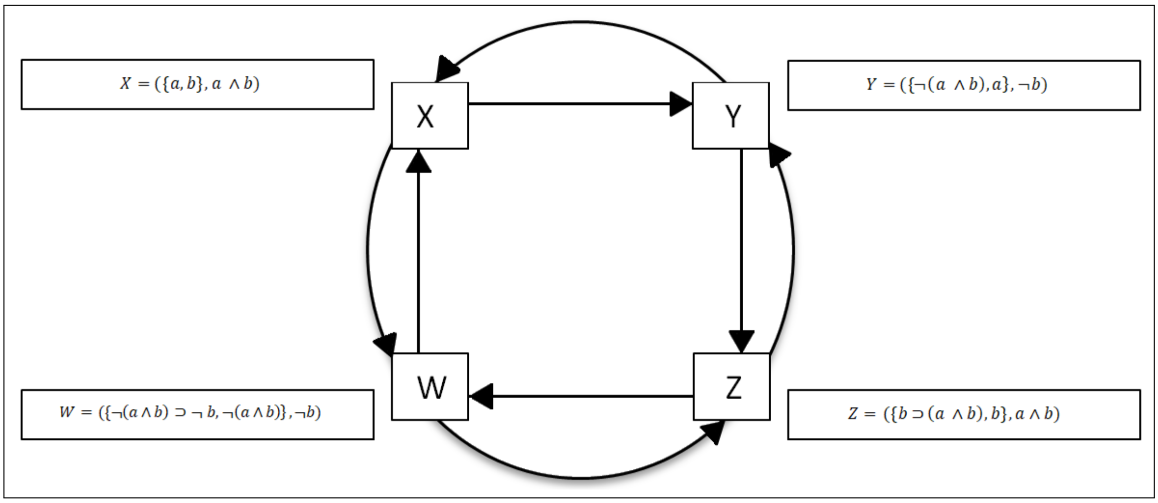

Figure 2.3: An example of Cl-Arg argumentation framework using the same structure of the AF of Figure 2.1.

Intuitively, Classical Logic instantiations of AFs interpret conflicting arguments as stressing mutual logical inconsistencies. Since Classical Logic does not include specific machinery for solving such conflicts, to accomplish this task we are going to define a defeat relation based on the preference $\prec$ (i.e., a strict partial order) over the arguments of the considered AFs.

[x]From now on $\curlywedge$ will express that an inconsistency has been reached when constructing an argument.

Definition 14 (CL-arg attacks and defeats)**.** [49]

- ***Undermine Attacks*** *$Y = (\Delta, \phi)$ undermine attacks $X = (\Delta', \phi')$ on ψ if $\phi = \overline{\psi}$ for some $\psi \in \Delta'$.*

- ***Defeats*** *Let $\langle AR, attacks \rangle$ be an AF where AR and 'attacks' are the arguments and the undermine attacks defined by $\mathscr{B}$. Let also $\prec \subseteq AR \times AR$ be a strict partial ordering. Finally, assume that $(Y, X) \in$ 'attacks', and Y undermine attacks X on ψ (i.e., on $X' = (\{\psi\}, \psi)$). Then $(Y, X) \in$ 'defeats' (Y defeats X) if $Y \not\prec X'$.*

Among the existing preference relations, we are going to use the 'Elitist' studied in [98]. That is to say, given a partial preordering $\leq$ over $\mathscr{B}$, then:

$$(\Gamma, \phi) \prec_E (\Delta, \theta) \text{ iff } \exists \alpha \in \Gamma \text{ such that } \forall \beta \in \Delta, \alpha < \beta \qquad (\textit{Elitist Preference})$$

Henceforth, the subscript '*E*' will be omitted in favour of the more simple notation $\prec$.

2.5.2 Rationality Postulates

The rationality postulates are specific properties whose satisfaction ensure that any concrete instantiations of an argumentation framework fulfil some rational criteria [27, 28]. Since Cl-Arg satisfies each of such postulates, we are going to introduce them by employing the formalisms of classical logic argumentation. Let $\langle \text{AR}, attacks \rangle$ be an AF, let also E refers to a complete extension and $C(E)$ denotes the set of conclusions of arguments in E. Then, using the intuitive presentation of [49]:

Definition 15 (Closure)**.** *This postulate refers to two kinds of 'closures':*

- (***Closure under sub-arguments***) *Sub-argument closure states that if an argument is in a complete extension then all its sub-arguments are in the same extension. Formally:*

  Let $X = (\Gamma, \delta) \in E$. If $\psi \in \Gamma$, then $(\{\psi\}, \psi) \in E$

- (***Closure under strict rules***) *Closure under strict rules states that if a strict rule with consequent φ can be applied to the conclusions of arguments in a complete extension, then there is an argument in that extension that concludes φ. Formally:*

  If $C(E) \vdash_c \varphi$, then $\varphi \in C(E)$

Definition 16 (Direct Consistency)**.** *Direct consistency is satisfied if the conclusions of arguments in a complete extension are consistent, i.e. don't contain contradictions. Formally:*

$$\forall \psi, \varphi \in C(E)\colon \psi \neq \overline{\varphi}$$

Definition 17 (Indirect Consistency)**.** *Indirect consistency means that the closure under strict rules of the arguments in a complete extension is consistent. Formally:*

$$\forall \psi, \varphi \in Cn(C(E))\colon \psi \neq \overline{\varphi}$$

The last two rationality postulates are less intuitive than the others. An example may outline a better presentation before their formal introduction.

Example 2 (Examples of Non-Interference and Crash Resistance)**.** *Having two syntactically disjoint bases $\mathscr{B}_1 = \{r\}$ and $\mathscr{B}_2 = \{q, \neg q\}$, from $\mathscr{B}_1$ is it possible to define the (Cl-arg) argument $X = (\{r\}, r)$, while from $\mathscr{B}_2$ is it possible to define the (Cl-arg) arguments $Y = (\{q, \neg q\}, \curlywedge)$, $Y' = (\{q\}, q)$ and $Y'' = (\{\neg q\}, \neg q)$. Considering then an AF composed only by X, it follows that its grounded extension will be equal to X itself. Whereas, considering an AF based only on Y, Y' and Y'', the grounded extension will be empty: it is trivial to see that if all the arguments of an AF attack each other and if these attacks succeed as defeats*[xi]*, then the grounded extension will be equal to $\emptyset$. $\mathscr{B}_2$ is a contaminating set that could influence the outcome of $\mathscr{B}_1$. The AF based on $\mathscr{B}_1 \cup \mathscr{B}_2$ will have $\emptyset$ as a grounded extension, rather than X: that is to say that the outcome of $\mathscr{B}_1$ will be the same as $\mathscr{B}_2$ (and, therefore, influenced by $\mathscr{B}_2$). This is due to the fact that from the inconsistent premises of Y it is possible to derive any formula and, therefore, attack and defeat any argument (for example, $Y^* = (\{q, \neg q\}, \neg r)$ will attack and defeat X). Hence, $\mathscr{B}_1 \cup \mathscr{B}_2$ does not satisfy Non-Interference or Crash Resistance.*

Definition 18 (Non-Interference)**.** *Non-interference means that for two syntactically disjoint knowledge bases $\mathscr{B}_1$ and $\mathscr{B}_2$, neither should 'influence' each other's argumentation defined inferences. Formally, given frameworks AF and AF' instantiated, respectively, by $\mathscr{B}_1$ and $\mathscr{B}_1 \cup \mathscr{B}_2$, where $\mathscr{B}_1 \parallel \mathscr{B}_2$:*

E is an extension of AF iff there is an extension E' of AF' where: $C(E)_{|symbols(\mathscr{B}_1)} = C(E')_{|symbols(\mathscr{B}_1)}$

[xi] Indeed, in this example, every attack succeeds as defeat.

Definition 19 (Crash Resistance). *A set of formulae is said to be contaminating iff it yields the same outcome when merged with a syntactically disjoint set of formulae. That is to say, a contaminating set of formulae makes all other unrelated sets of formulae irrelevant when being merged with it. Hence, a logical formalism satisfies Crash Resistance iff there does not exist a set of formulae that is contaminating. Formally, there is no* $\mathscr{B}_1$ *such that for any syntactically disjoint* $\mathscr{B}_2$*, and the frameworks AF and AF′ instantiated respectively by* $\mathscr{B}_1$ *and* $\mathscr{B}_1 \cup \mathscr{B}_2$*, occurs that:*

$$\{C(E)|E \text{ is an extension of } AF\} = \{C(E)|E \text{ is an extension of } AF'\}$$

2.5.3 Dialectical Cl-Arg

Unlike standard formalisation of Cl-Arg, real-world agents behave pragmatically and do not need to: (i) always construct every argument defined by a base, (ii) enforce consistency and subset minimality checks on their arguments (nor do they have enough computational power to do these checks, given their limited resources). Dialectical Cl-Arg provides a formalisation of real-world modes of dialectical reasoning from resource-bounded agents whilst satisfying both the rationality postulates [27, 28] and practical desiderata [56].

Definition 20 (Dialectical Arguments). [49] $X = (\Delta, \Gamma, \alpha)$ *is a* dialectical argument *defined by a base* $\mathscr{B}$ *of classical wff, if* $(\Delta \cup \Gamma) \subseteq \mathscr{B}$*,* $\Delta \cap \Gamma = \emptyset$*, and* $\Delta \cup \Gamma \vdash_c \alpha$*. If* $\alpha = \curlywedge$ *then X is said to be a* falsum *argument. If* $\Gamma = \emptyset$ *then X is said to be* unconditional*; else X is* conditional*. Finally, if* $\Delta = \emptyset$ *then X is said to be* unassailable.

Δ*,* Γ *and* α *are respectively referred to as the* premises $(Prem(X))$*,* suppositions $(Supp(X))$ *and* conclusion $(Con(X))$ *of* $X = (\Delta, \Gamma, \alpha)$*. Also, the union of premises and suppositions of X can be referred to as the* assumptions $(Assumptions(X))$ *of the argument.*

Attacks and defeats for Dialectical Cl-Arg work differently than their respective counterparts for Classical Logic Argumentation (Cl-Arg). The reason is the presence of suppositions embedded in the internal structure of the arguments. Intuitively, it is common practice for interlocutors in dialogues to differentiate between their own arguments' premises, regarded as true, and their opponents' premises that they want to challenge: "by considering what I deem to be valid and supposing what you have committed to, I can show your premises inconsistency". This motivates such an epistemic distinction between information considered true (i.e., $Prem(X)$, the *premises* of an argument X) and opponents' information supposed true (i.e., $Supp(X)$, the *supposition* of an argument X) which proves useful also in solving the so-called 'foreign commitment problem' [29].

Definition 21 (Attacks and Defeats)**.** [49] *Let AR be a set of dialectical arguments defined by a base* $\mathscr{B}$*. The attack relation 'attacks'* $\subseteq AR \times AR$ *is defined as follows. For any* $X = (\Delta, \Gamma, \alpha)$*,* $Y = (\Pi, \Sigma, \beta) \in AR$*: attacks*$(X,Y)$ *iff:*

- *if* $\alpha \neq \curlywedge$ *then* $\overline{\alpha} \in \Pi$ *(X attacks Y on* $\overline{\alpha}$*, equivalently on* $Y' = (\{\overline{\alpha}\}, \emptyset, \overline{\alpha})$*);*
- *if* $\alpha = \curlywedge$ *(X attacks Y on any* $\phi \in \Gamma \cap \Pi$*, equivalently on any* $Y' = (\{\phi\}, \emptyset, \phi)$*).*

Let $\prec$ *be a strict partial ordering over AR. Then, for every X, Y such that attacks*(X,Y)*, defeats*(X,Y) *iff exactly one of the following holds:*

- *either X is an argument of the form* $(\emptyset, \Gamma, \curlywedge)$*;*
- *else,* $\exists \psi \in Prem(Y)$ *such that attacks*(X,Y) *on* ψ*, and* $X \not\prec (\{\psi\}, \emptyset, \psi)$*.*

$X \Rightarrow Y$ *will stand for "defeats*(X,Y)*", and* $X \not\Rightarrow Y$ *will stand for "*$\neg$*defeats*(X,Y)*".*

The strict partial ordering of Definition 21 refers to the Elitist Preference Ordering, adapted for Dialectical Cl-Arg. In addition, the authors of [49] show that such preference is also 'redundance-coherent' in the sense that arguments are not strengthened when redundantly weakening with syntactically disjoint assumptions[xii]. This is an important property that ensures the satisfaction of the non-contamination (i.e., Non-Interference and Crash Resistance) rationality postulates for Dialectical Cl-Arg.

Definition 22 (Elitist Preference Ordering)**.** *Let* X,Y *be dialectical classical logic arguments defined by a base* $\mathscr{B}$*, and* $\leq$ *a partial preordering over* $\mathscr{B}$*. Then:*

(i) $X \prec Y$ *iff* $\exists \alpha \in Assumptions(X)$ *such that* $\forall \beta \in Assumptions(Y)$*,* $\alpha < \beta$*.*

(ii) $\prec$ *is redundance-coherent iff:* $\forall X, X', Y$ *such that* $X = (\Gamma, \emptyset, \alpha)$*,* $X' = (\Delta \cup \Gamma, \emptyset, \alpha)$*, and* $\Delta \parallel \Gamma \cup \{\alpha\}$*: if* $X \prec Y$ *then* $X' \prec Y$*.*

Cl-Arg assumes instantiation of an AF by all arguments defined by a base $\mathscr{B}$ of classical wff, a task that proves to be unfeasible for agents with limited resources. As such, dialectical arguments (Definition 20) along with the described defeat relation (Definition 21) allow us to introduce a *dialectical AF* as an argumentation framework $\langle \text{AR}, \textit{defeats} \rangle$ where AR is any subset of the dialectical arguments defined by a base $\mathscr{B}$.

Defeats and dialectical defeats for dialectical AFs present an important difference: the

[xii]Here 'weakening' denotes that a logical entailment from, say Δ, continues to be valid when adding some Γ to Δ.

$A_1 = (\{a\}, \emptyset, a)$	$B_1 = (\{b\}, \emptyset, b)$
$F_1 = (\{b, \neg a \vee \neg b\}, \emptyset, \neg a)$	$G_1 = (\{a, \neg a \vee \neg b\}, \emptyset, \neg b)$
$F_2 = (\{b\}, \{\neg a \vee \neg b\}, \neg a)$	$G_2 = (\{a\}, \{\neg a \vee \neg b\}, \neg b)$
$F_3 = (\{\neg a \vee \neg b\}, \{b\}, \neg a)$	$G_3 = (\{\neg a \vee \neg b\}, \{a\}, \neg b)$
$N_1 = (\{a \supset b\}, \{\neg b\}, \neg a)$	$N_2 = (\{a \supset b, \neg b\}, \emptyset, \neg a)$
$N_3 = (\{a \supset b, a\}, \emptyset, b)$	$O_1 = (\{\neg(a \supset b)\}, \emptyset, \neg(a \supset b))$
$L_1 = (\{\neg b\}, \emptyset, \neg b)$	$X_3 = (\{b\}, \{\neg b\}, \curlywedge)$
$C_1 = (\{\neg a \vee \neg b\}, \emptyset, \neg a \vee \neg b)$	$H_1 = (\{a, b\}, \emptyset, \neg(\neg a \vee \neg b))$
$X_1 = (\emptyset, \{a, b, \neg a \vee \neg b\}, \curlywedge)$	$X_2 = (\{a, b, \neg a \vee \neg b\}, \emptyset, \curlywedge)$

Figure 2.4: Example of dialectical arguments defined by a base $\mathscr{B} = \{a, b, \neg a \vee \neg b, \neg b, a \supset b, \neg(a \supset b)\}$.

reference to a set $\mathscr{S}$ of arguments. The general idea is that, when challenging the acceptability of an argument with respect to a set $\mathscr{S}$, the defeating argument can also suppose premises from all the arguments in $\mathscr{S}$. Whereas, the argument that defends $\mathscr{S}$ can only suppose the premises of the defeating argument. This new kind of defeat compelled the authors of [49] to adjust the standard semantics accordingly.

Definition 23 (Dialectical defeats and semantics for dialectical AFs)**.** *Let $\langle AR, defeats\rangle$ be a dialectical AF, $\mathscr{S} \subseteq AR$ and $X, Y \in AR$. Then:*

1) *X* dialectically defeats *Y* with respect to *$\mathscr{S}$, denoted $X \Rightarrow_{\mathscr{S}} Y$, if $defeats(X, Y)$ and $Supp(X) \subseteq Prem(\mathscr{S} \cup \{Y\})$.*

2) *$\mathscr{S}$ is* conflict free *if $\forall Z, Y \in \mathscr{S}$, $Z \not\Rightarrow_{\mathscr{S}} Y$.*

3) *Y is* acceptable *with respect to $\mathscr{S}$ if $\forall X$ such that $X \Rightarrow_{\mathscr{S}} Y$, $\exists Z \in \mathscr{S}$ such that $Z \Rightarrow_{\{X\}} X$.*

4) *Let $\mathscr{S}$ be conflict free. Then $\mathscr{S}$ is: an* admissible *extension iff $X \in \mathscr{S}$ implies X is acceptable with respect to $\mathscr{S}$; a* complete *extension iff $\mathscr{S}$ is admissible and if X is acceptable with respect to $\mathscr{S}$ then $X \in \mathscr{S}$; a* preferred *extension iff it is a set inclusion maximal complete extension; the* grounded *extension iff it is the set inclusion minimal complete extension.*

Example 3. *Consider Figure 2.5. Let A_1, $B_1 \in \mathscr{S}$ be the dialectical arguments introduced in Figure 2.4, and let $Z_1 = (\{a \supset \neg b\}, \{a\}, \neg b)$ be a dialectical argument that defeats B_1 with respect to $\mathscr{S}$, i.e., $Z_1 \Rightarrow_S B_1$. Notice that such defeat occurs only due to the presence of the formula $a \in Prem(A_1)$. The supposition of the formula a by the dialectical argument Z_1 (i.e., $Supp(Z_1) \subseteq Prem(\mathscr{S} \cup \{B_1\})$ allows concluding $\neg b$, hence defeating argument*

B_1. *However,* $Z_0 = (\{a \supset \neg b\}, a \supset \neg b)$*, the Cl-Arg argument that has the same premises as* Z_1*, is not capable of moving the same defeat to* B_1*. Indeed, the absence of the formula* a *among the premises prevents* Z_0 *from classically entailing the conclusion* $\neg b$*, hence precluding the defeat of argument* B_1*. This example shows, by supposing formulae (from single arguments or sets), how additional attacks and defeats may arise for Dialectical Cl-Arg arguments in comparison with Cl-Arg arguments.*

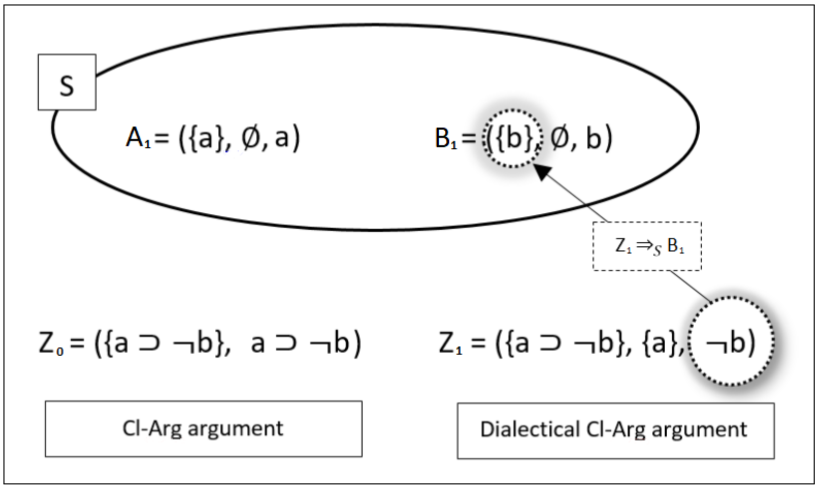

Figure 2.5: An example of differences between Cl-Arg and Dialectical Cl-Arg.

The conclusions of an extension in Dialectical Cl-Arg may derive from conditional arguments that only suppose the truth of the premises without any commitment. As such, we should revise the previously used notation. Once the extensions are defined, we detach only the conclusions of *unconditional* arguments all of whose assumptions are premises presumed true.

Definition 24 (Conclusions of an Extension in Dialectical Cl-Arg)**.** *Let* E *be an extension of a dialectical AF. Then* $C(E) = \{\phi \mid (\Delta, \emptyset, \phi) \in E\}$.

Dialectical AFs may enjoy some specific properties, as explained in [49]. Here we are going to outline five of them (P1, P2, P3, P4, P4′), which will be used later in the next chapters.

Proposition 1. *Given a dialectical AF =* $\langle AR, defeats \rangle$*:*

$(P1)$ $\forall X \in AR$*:* $\alpha \in Prem(X)$ *implies that* $(\{\alpha\}, \emptyset, \alpha) \in AR$ *(where* $(\{\alpha\}, \emptyset, \alpha)$ *is denoted as the* ‘elementary argument’ *of* X *defined by* α*);*

$(P2)$ $\forall X \in AR$: *if* $X' \in [X]$, *that is to say, if* X' *is the* logically equivalent argument *of* X (*i.e., the only difference between X and* X' *is the different distribution of premises and supposition*), *then* $X' \in AR$;

$(P3)$ *If* $(\Delta, \emptyset, \alpha) \in AR$ *and* $(\Gamma, \emptyset, \overline{\alpha}) \in AR$, *then either* $(\Delta, \emptyset, \curlywedge) \in AR$ *or* $(\Gamma, \emptyset, \curlywedge) \in AR$ *or* $(\Delta \cup \Gamma, \emptyset, \curlywedge) \in AR$;

$(P4)$ *If* $(\Gamma, \emptyset, \alpha) \in AR$, $\Delta \subseteq \Gamma$, $\Delta \neq \emptyset$ *and* $\Delta \parallel \Gamma \setminus \Delta \cup \{\alpha\}$, *then either* $(\Delta, \emptyset, \curlywedge) \in AR$ *or* $(\Gamma \setminus \Delta, \emptyset, \alpha) \in AR$;

$(P4')$ *If* $(\Gamma, \emptyset, \alpha) \in AR$, $\Delta \subseteq \Gamma$, $\Delta \neq \emptyset$ *and* $\Delta \parallel \Gamma \setminus \Delta \cup \{\alpha\}$, *then* $(\Delta, \emptyset, \curlywedge) \in AR$.

We can now refer to ⟨AR, *defeats*⟩ as a *partially instantiated dialectical AF* (*pdAF*) if AR corresponds to any subset of the dialectical arguments defined by a base $\mathscr{B}$ such that AR satisfies $P1$, $P2$, $P3$ and $P4$.

A *non-redundant pdAF* is, instead, a pdAF such that AR satisfies $P1$, $P2$, $P3$, $P4'$ and there are no redundantly contaminated arguments[xiii].

2.5.4 Rationality Postulates for Dialectical Cl-Arg

Dialectical Cl-arg satisfies the rationality postulates and does so by requiring that the AF enjoys properties $P1$-$P4$. This would impose minimally restrictive assumptions[xiv] as to the arguments that agents should be able to construct, thus providing a rational account of arguments more suited for the limited availability of resources that characterises real-world agents. The postulates are defined similarly as per Section 2.5.2. A detailed report of the proofs of their validity, along with each required lemma, theorem and any additional definition, is given in [49].

Theorem 4 (Sub-argument Closure)**.** *Let E be a complete extension of a dialectical AF* $=$ ⟨*AR, defeats*⟩ *such that AR satisfies P1. Let* $X \in E$. *Then if* $\alpha \in Prem(X)$ *then* $(\{\alpha\}, \emptyset, \alpha)$ $\in E$.

Theorem 5 (Direct Consistency)**.** *Let E be an admissible extension of a dialectical AF* $=$ ⟨*AR, defeats*⟩. *If AR satisfies P1, P2 and P3, then* $\forall \alpha, \beta \in C(E)$, $\alpha \neq \curlywedge$ *and* $\beta \neq \overline{\alpha}$.

[xiii]A redundantly contaminated argument is an argument that employs redundant assumptions, that is to say, a subset of the assumptions is unnecessary for drawing the argument conclusion. This may occur due to the fact that Dialectical Cl-Arg drops subset minimality checks. To avoid violation of the non-contamination postulates, the adopted preference relation has to be ‘redundance-coherent’. Indeed, this is the case of the Elitist preference of Definition 22.

[xiv]Especially the satisfaction of *P1-P3*.

Theorem 6 (Premise Consistency)**.** *Let* $\langle AR, \textit{defeats}\rangle$ *be a dialectical AF such that AR satisfies P2. If for some* $\Delta \subseteq Prem(E)$*:* $(\Delta, \emptyset, \curlywedge) \in AR$*, then E cannot be an admissible extension of* $\langle AR, \textit{defeats}\rangle$*.*

Closure under Strict Rules for Dialectical Cl-Arg slightly differs from its standard version. That is caused by the limited availability of resources that characterises real-world agents. Indeed, although it may be the case that $C(E) \vdash_c \alpha$, it may not be that there exists an $X \in E$ such that X concludes α, given that agents are not logically omniscient and do not construct all arguments from a base. Hence, the following version of the postulate:

Theorem 7 (Closure under Strict Rules)**.** *Let E be a complete extension of a dialectical* $AF = \langle AR, \textit{defeats}\rangle$*, where AR satisfies P1. Let* $E' \subseteq E$ *and* $C(E') \vdash_c \alpha$*. If there exists an* $X = (\Delta, \emptyset, \alpha) \in AR$ *such that* $\Delta = Prem(E')$*, then* $X \in E$*.*

Theorem 8 (Non-Interference)**.** *Non-interference is satisfied by (non-redundant) pdAFs.*

Theorem 9 (Crash Resistance)**.** *Crash Resistance is satisfied if there does not exist a contaminating base* $\mathscr{B}$ *for pdAFs and non-redundant pdAFs.*

2.5.5 Dung's Fundamental Lemma and Monotonicity of the Characteristic Function for Dialectical Cl-Arg

Among the most important key results of Dung's seminal paper [53] are the *fundamental lemma* and the *monotonicity of the AF's characteristic function* $\mathscr{F}_{AF}$ (that yields the constructive definition of the grounded extension via its iterations). However, unlike Dung's standard AFs, these properties cannot be straightforwardly shown, since when determining the acceptability of X w.r.t. E, the defeats on X are not independent from the set E under consideration. For dialectical AFs, the defeats on X w.r.t. E may be a subset of the defeats on X w.r.t. $E' \supset E$ (due to the additional premises committed to in E'). To avoid this issue, the authors of [49] have devised specific 'epistemically maximal' sets of arguments by means of whose it is possible to show the desired properties.

Definition 25 (Epistemically maximal sets)**.** *Let* $\langle AR, \textit{defeats}\rangle$ *be a dialectical AF. Then* $E \subseteq AR$ *is* epistemically maximal (em) *iff:*

$$\text{If } X = (\Delta, \Gamma, \alpha) \in E,\ \Gamma' \subseteq (\Gamma \cap Prem(E)),\ \text{then } X' = (\Delta \cup \Gamma', \Gamma \setminus \Gamma', \alpha) \in E \qquad (\bullet)$$

The function $Cl_{em} : 2^{AR} \to 2^{AR}$ *maps any E to its epistemically maximal set. As such,* $Cl_{em}(E)$ *denotes the smallest superset of E that is closed under condition* $(\bullet)$*.*

It is now possible to prove a variant of the fundamental lemma that involves *em* sets:

Lemma 2 (Fundamental Lemma for Dialectical Cl-Arg)**.** [49] *Let X, X' be acceptable w.r.t. an admissible extension E of a dialectical AF $= \langle AR, defeats \rangle$. Then:*

(1) *$Cl_{em}(E \cup \{X\})$ is admissible, and*

(2) *X' is acceptable w.r.t. $Cl_{em}(E \cup \{X\})$*

Lemma 2 entails:

Proposition 2. *Every admissible extension of a dialectical AF is a subset of a preferred extension.*

Proposition 2 guarantees that it suffices to show that an argument X is in an admissible extension, in order to prove that X is credulously justified under the preferred semantics (exactly as Dung's standard AFs).

Finally, by employing a variant of the framework characteristic function, i.e., $\mathscr{F}_p$ whose domain is composed of sets E that are *em* admissible, and that returns $Cl_{em}(\mathscr{F}(E))$, we can also show the constructive definition of the grounded extension. Indeed, starting with the empty set and iteratively applying $\mathscr{F}_p$, the monotonically increasing sequence approximates, and in the case of a *finitary* dialectical AF, it constructs, the least fixed point of $\mathscr{F}_p$, i.e., the grounded extension:

Proposition 3. [49] *Let $\langle AR, defeats \rangle$ be a dialectical AF, and $F^0 = \emptyset$, $F^{i+1} = \mathscr{F}_p(F^i)$. Let E be the grounded extension of $\langle AR, defeats \rangle$. Then:*

1. $E \subseteq \bigcup_{i=0}^{\infty}(F^i)$.
2. *If $\langle AR, defeats \rangle$ is* finitary*, i.e., $\forall X \in AR$, the set $\{Y \mid defeats(Y,X)\}$ is finite, then $E = \bigcup_{i=0}^{\infty}(F^i)$.*

We have thus outlined how Cl-Arg formalises Classical Logic instantiations of structured argumentation frameworks whilst satisfying the rationality postulates. Dialectical Cl-Arg moves a step further by approximating such a method to real-world resource-bounded agents. The dialectical version of the Fundamental Lemma and the monotonicity of the characteristic functions (Section 2.5.5) will be especially useful in the next two chapters when we are showing the soundness and completeness results of the newly introduced proof theories. Nevertheless, both approaches lack the consideration of inference rules that differs from the strict ones (classical entailment). Defeasible rules allow for the account of extra attacks/defeats, hence expanding the interacting options available to agents in specific circumstances. The following section will handle those additional formalisations.

2.6 Dialectical ASPIC$^+$

2.6.1 ASPIC$^+$

ASPIC$^+$ is a well established general framework for argumentation with preferences. Developed in papers such as [98, 106], ASPIC$^+$ is halfway between Dung's fully abstract approach [53] and its concrete instantiations. Indeed, ASPIC$^+$ specifies the internal (rule-based) structure of the arguments while maintaining a level of abstractness that allows the instantiations of its framework by various types of underlying logics [98]. The resulting frameworks will also satisfy the key properties and rationality postulates of [27, 28, 53].

Definition 26 (ASPIC $^+$Argumentation System)**.** *An argumentation system* (ASY) *is a tuple* $\langle L, -, \mathscr{R}, n\rangle$ *where:*

- *L is a logical language;*
- $-$ *is a function from L to* 2^L *such that:*
    1) φ *is a contrary of* ψ *(denoted as* $\varphi = \sim \psi$*) if* $\varphi \in \overline{\psi}$*,* $\psi \notin \overline{\varphi}$*;*
    2) φ *is a contradictory of* ψ *(denoted as* $\varphi = \neg\psi$*) if* $\varphi \in \overline{\psi}$*,* $\psi \in \overline{\varphi}$*;*
- $\mathscr{R} = \mathscr{R}_s \cup \mathscr{R}_d$ *is a set of strict* $(\mathscr{R}_s)$ *and defeasible* $(\mathscr{R}_d)$ *inference rules (such that* $\mathscr{R}_s \cap \mathscr{R}_d = \emptyset$*) of the form* $\varphi_1, \ldots, \varphi_n \rightarrow \psi$*, respectively,* $\varphi_1, \ldots, \varphi_n \rightsquigarrow \psi$ $(\psi \neq \curlywedge)$*, where* φ_i, ψ *are meta-variables ranging over wff of L;*
- $n : \mathscr{R}_d \rightarrow L$ *is a naming convention for defeasible rules.*

Notice that a strict rule is an expression indicating that if the antecedent holds, then the consequent is entailed without exception. On the other hand, the consequent of a defeasible rule only 'usually' follows when the antecedent holds.

According to the brief overview given in [51], an ASPIC$^+$ theory $\mathscr{T} = (\mathsf{ASY}, \mathscr{K})$ consists of an argumentation system and a knowledge base $\mathscr{K} \subseteq L \backslash \{\curlywedge\}$ comprising two disjoint subsets $\mathscr{K}_n$ (the 'infallible' axiom premises) and $\mathscr{K}_p$ (the 'fallible' ordinary premises). $\mathscr{K}_n$ represents infallible information and/or axioms in some deductive logic and $\mathscr{K}_p$ fallible information.

Definition 27 (ASPIC$^+$ argument)**.** [98] *An argument X on the basis of a knowledge base* $\mathscr{K}$ *in an argumentation system* $\langle L, -, \mathscr{R}, n\rangle$ *is one of the following:*

- φ *if* $\varphi \in \mathscr{K}$ *with:* $\mathsf{prem}(X) = \{\varphi\}$; $\mathsf{conc}(X) = \varphi$; $\mathsf{sub}(X) = \{\varphi\}$; $Rules(X) = \emptyset$.

- *1.* $X_1, \ldots, X_n \rightarrow \psi$ *if* $X_1, \ldots, X_n$ *are arguments such that there exists a strict rule* $\mathsf{conc}(X_1), \ldots, \mathsf{conc}(X_n) \rightarrow \psi$ *in* $\mathcal{R}_s$.

  2. $X_1, \ldots, X_n \rightsquigarrow \psi$ *if* $X_1, \ldots, X_n$ *are arguments such that there exists a defeasible rule* $\mathsf{conc}(X_1), \ldots, \mathsf{conc}(X_n) \rightsquigarrow \psi$ *in* $\mathcal{R}_d$.

  3. $\mathsf{prem}(X) = \mathsf{prem}(X_1) \cup \ldots \cup \mathsf{prem}(X_n)$; $\mathsf{conc}(X) = \psi$; $\mathsf{sub}(X) = \mathsf{sub}(X_1) \cup \ldots \cup \mathsf{sub}(X_n) \cup \{X\}$; $Rules(X) = Rules(X_1) \cup \ldots \cup Rules(X_n) \cup \{\mathsf{conc}(X_1), \ldots, \mathsf{conc}(X_n) \rightarrow / \rightsquigarrow \psi\}$.

  4. $TopRule(X)$ *refers to the last rule applied in* X.

Intuitively, $\mathsf{prem}(X)$ *stand for 'premises of X'*; $\mathsf{conc}(X)$ *for 'conclusion of X' and* $\mathsf{sub}(X)$ *for 'sub-argument of X'.*

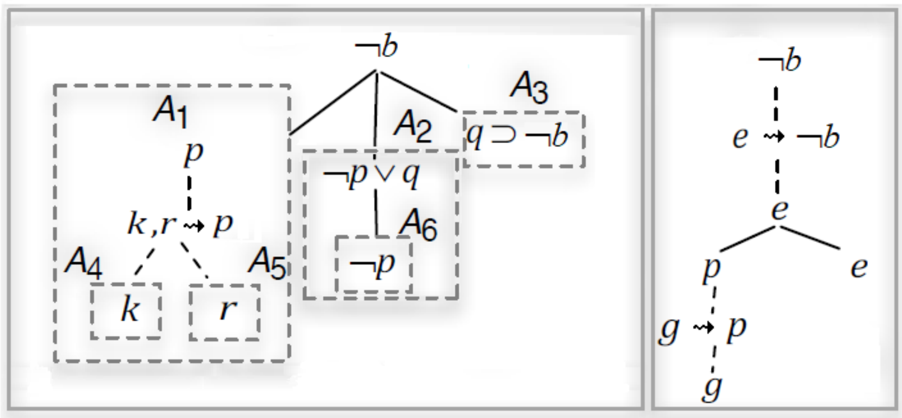

Figure 2.6: Example of ASPIC$^+$ arguments and sub-arguments (dashed squares), composed of strict (solid lines) and defeasible (dashed lines) rules, adapted from [51].

There are three kinds of attack licensed by ASPIC$^+$, each of which targets a different element of the attacked argument. Undercuts argue against the defeasible rule used to derive the attacked argument conclusion. Rebuttals counter the attacked argument conclusion while underminers target the attacked argument premises. Formally:

Definition 28 (ASPIC$^+$ attacks)**.** [98]

1) [***Undercut***] *X undercuts argument Y (on* Y'*) iff* $\mathsf{conc}(X) \in \overline{n(d)}$ *for some* $(Y') \in \mathsf{sub}(Y)$ *such that* (Y')*'s TopRule d is defeasible.*

2) [***Rebut***] *X rebuts argument Y (on* Y'*) iff* $\mathsf{conc}(X) \in \overline{\varphi}$ *for some* $Y' \in \mathsf{sub}(Y)$ *of the form* $Y''_1, \ldots, Y''_n \rightsquigarrow \varphi$ *(where* $\mathsf{conc}(X)$ *is either a contrary or contradictory of* φ*).*

3) [***Undermine***] *X undermines argument Y (on Y') iff* $\mathsf{conc}(X) \in \overline{\varphi}$ *for some* $Y' = \varphi$, $\varphi \in \mathsf{prem}_p(Y)$[xv] (*where* $\mathsf{conc}(X)$ *is either a contrary or contradictory of* φ).

Notice also that undercuts and attacks targeting contraries are preference independent, which means that they always succeed as defeats. Rebuttals and underminers may instead be preference dependent, i.e., their success as defeats might rely on some preferences ordering over the involved arguments.

Unfortunately, ASPIC$^+$ presents some limitations: the rationality postulates are satisfied only under specific (sometimes onerous) assumptions. For example, consistency holds only under the tacit assumption that agents have access to unlimited resources to generate the required arguments. In addition, ASPIC$^+$ is not guaranteed to satisfy non-contamination [28].

2.6.2 D-ASPIC$^+$

Dialectical ASPIC$^+$ (D-ASPIC$^+$) is a general framework that provides a full rational account of ASPIC$^+$ for real-world resource-bounded agents. Indeed, [51] shows how the consistency, closure [27] and non-contamination [28] rationality postulates are satisfied under minimal assumption as to the resource available to construct arguments. Similarly to Dialectical Cl-Arg [49], the framework presented by D-ASPIC$^+$ solves the *foreign commitment problem* [29] by employing a common dialectical move: the agents suppose their interlocutor's arguments in order to argue against them, whilst separating such suppositions from their own committed formulae. As such, a D-ASPIC$^+$ argument X differentially labels its *maximal fallible sub-arguments*[xvi] according to whether they are committed or supposed for the sake of the argument. Notice that, by supposing the arguments and claims of an interlocutor, we will get a conclusion that is strictly (as opposed to defeasibly) entailed.

Definition 29 (D-ASPIC$^+$ argument)**.** *An ASPIC$^+$ argument X is a D-ASPIC$^+$ argument $X = (\Delta, \Gamma, \phi)$ iff exactly one of $J1, \ldots, J4$ hold:*

(***J1***) *$X = \phi$, $\phi \in \mathscr{K}_p$ and X is exclusively labelled with either* Ⓒ *(for 'committed') or* Ⓢ *(for 'supposed');*

(***J2***) *$X = \phi$, $\phi \in \mathscr{K}_n$;*

[xv] $\mathsf{prem}_p(\mathrm{X}) = \mathsf{prem}(\mathrm{X}) \cap \mathscr{K}_p$, where $\mathsf{prem}(\mathrm{X})$ consists of all the premises of X and $\mathscr{K}_p$ refers to the 'fallible' ordinary premises of the knowledge base.

[xvi] The maximal fallible sub-arguments of an argument X are those sub-arguments with the last defeasible rule in X or else X's ordinary premises. That is, they are the maximal sub-arguments of X on which X can be attacked.

($\boldsymbol{J3}$) $X = X_1, \ldots, X_n \rightsquigarrow \varphi$ *and X is exclusively labelled with either* Ⓒ *or* Ⓢ*;*

($\boldsymbol{J4}$) $X = X_1, \ldots, X_n \rightarrow \varphi$ *and each* X_i *is a D-ASPIC*$^+$ *argument*[xvii]*, and* $\forall X' \in \mathsf{sub}(X)$*:*

- **C1** X' *is not labelled with* Ⓒ *and* Ⓢ*.*
- **C2** $X' \in \Delta$ *iff* X' *is labelled* Ⓒ*,* $X' \in \Gamma$ *iff* X' *is labelled* Ⓢ *in* X*. Then, no supposed* $X' \in \Gamma$ *is a proper sub-argument of some committed* $Z \in \Delta$*, since commitment to* Z *implies a commitment to all sub-arguments of* Z*.*

Consider that the structure of a D-ASPIC$^+$ argument $X = (\Delta, \Gamma, \alpha)$[xviii] envisages *commitments* ($\text{Comm}(X) = \Delta$) and *suppositions* ($Supp(X) = \Gamma$).

Observe also that when representing a D-ASPIC$^+$ argument X in the format $X = (\Delta, \Gamma, \alpha)$, I am going to explicitly specify (either in Δ or Γ, depending on the argument commitments) the defeasible rule used as *TopRule* (if any). This serves to unambiguously distinguish whether a conclusion is the result of a defeasible or a strict entailment.

Attacks and defeats are the same as for ASPIC$^+$ arguments, except that the attacker must target a fallible *committed* sub-argument of the attackee. In addition, notice that the attacks moved by falsum arguments (i.e., arguments concluding $\curlywedge$) having empty premises solve the *foreign commitment problem*. Indeed, the attack would amount to a dialectical demonstration that the supposed premises of the attackee are inconsistent. Finally, recall also that undercuts, along with rebuts and undermines targeting contraries, are preference independent attacks that always succeed as defeats.

Definition 30 (D-ASPIC$^+$ attacks and defeats)**.** [51] *Let* $\prec$ *be a strict partial ordering over a set AR of D-ASPIC*$^+$ *arguments. Let also* $X = (\Delta, \Gamma, \alpha) \in AR$ *and* $Y = (\Pi, \Sigma, \beta) \in AR$*.*

- *if* $\alpha \neq \curlywedge$*, then X attacks Y if X undercuts, rebuts or undermines some* $Y' \in \Pi$ *on* $Y'' \in \mathsf{sub}(Y')$*. X defeats Y iff X attacks Y and if the attack rebuts/undermines a contradictory then* $X \not\prec Y''$*.*
- *if* $\alpha = \curlywedge$ *then X attacks Y on any* $Y' \in \Gamma \cap \mathsf{sub}(\Pi)$*. If* $\Delta = \emptyset$*, then X defeats Y. If* $\Delta \neq \emptyset$ *then X defeats Y only if* $\exists Y' \in \Gamma \cap \mathsf{sub}(\Pi)$*,* $X \not\prec Y'$*.*

[xvii] In other words, traversing each path from the root (conclusion) to a leaf of an ASPIC$^+$ argument X, assign label Ⓒ or Ⓢ when first encountering either an ordinary premise or conclusion of a defeasible inference rule. Terminate the traversal once a label is assigned.

[xviii] Notice that, only from the structure point of view, D-ASPIC$^+$ and Dialectical Cl-Arg arguments are identical.

Given the analogy with Dialectical Cl-Arg, even the dialectical defeats will work in a similar way for D-ASPIC$^+$. That is to say, when establishing whether Y is acceptable w.r.t. a set of argument $\mathscr{S}$, one commits to fallible arguments in $\mathscr{S}$ and Y. Hence $X = (\Delta, \Gamma, \alpha)$ can challenge (i.e., defeat) Y if X supposes a subset Γ of these fallible arguments, arguing that together with Δ, Γ necessarily entails a claim that conflicts with some fallible element in Y. On the other hand, X is not required to define an admissible set, as such, it will be countered on an individual basis.

Definition 31 (D-ASPIC$^+$ dialectical defeats and semantics)**.** [51] *Let $\langle AR, defeats\rangle$ be a dialectical AF defined by a theory $\mathscr{T} = \langle \mathsf{ASY}, \mathscr{K}\rangle$[xix], where AR is a set of D-ASPIC$^+$ arguments, and 'defeats' $\subseteq AR \times AR$ is the set of defeats. Then:*

1) *X dialectically defeats Y with respect to $\mathscr{S} \subseteq AR$, denoted $X \Rightarrow_{\mathscr{S}} Y$, if $defeats(X,Y)$ and $Supp(X) \subseteq \mathsf{sub}\mathrm{Comm}(\mathscr{S} \cup \{Y\})$*[xx]*.*

2) *$\mathscr{S}$ is conflict free if $\forall Z, Y \in \mathscr{S}$, $Z \not\Rightarrow_{\mathscr{S}} Y$.*

3) *Y is acceptable with respect to $\mathscr{S}$ if $\forall X$ such that $X \Rightarrow_{\mathscr{S}} Y$, $\exists Z \in \mathscr{S}$ such that $Z \Rightarrow_{\{X\}} X$.*

4) *Let $\mathscr{S}$ be conflict free. Then $\mathscr{S}$ is: an* admissible *extension iff $X \in \mathscr{S}$ implies X is acceptable with respect to $\mathscr{S}$; a* complete *extension iff $\mathscr{S}$ is admissible and if X is acceptable with respect to $\mathscr{S}$ then $X \in \mathscr{S}$; a* preferred *extension iff it is a set inclusion maximal complete extension; the* grounded *extension iff it is the set inclusion minimal complete extension.*

The following example will clarify Definitions 29 and 30 while providing also an instance of the dialectical defeats formalised by Definition 31.

Example 4. *Let $Arg1 = (\{a\}, \emptyset, a)$, $Arg2 = (\{b\}, \emptyset, b)$, $Arg3 = (\{e\}, \emptyset, e)$, $Arg4 = (\{Arg3, d_1\}$, $Arg5 = (\{Arg2\}, \{Arg4\}, \neg a)$ and $Arg6 = (\{Arg1, Arg4\}, \emptyset, \neg b)$ be D-ASPIC$^+$ arguments, such that their structure specifies the committed and/or supposed sub-arguments (respectively denoted with Ⓒ or Ⓢ in Figure 2.7). Let also $\mathscr{R}_s = \vdash_{CL}$ (i.e., the consequence relation of classical logic), $\mathscr{R}_d = \{e \leadsto \neg a \vee \neg b\}$, $n(e \leadsto \neg a \vee \neg b) = d_1$, $\mathscr{K}_n = \emptyset$ and $\mathscr{K}_p = \{a, b, e\}$. Notice that each D-ASPIC$^+$ argument is depicted as an upside-down tree, whose leaves are premises, yielding the arguments' claim (the root node) via application of defeasible*

[xix] The argumentation theory $\mathscr{T}$ is characterized in the same way as ASPIC$^+$ (Section 2.6.1).

[xx] Intuitively, $\mathsf{sub}\mathrm{Comm}(\mathscr{S} \cup \{Y\})$ identifies all the sub-arguments of the arguments in $\mathscr{S}$ and Y that are labelled with Ⓒ.

rules (dashed lines) or strict rules (solid lines). In addition, the straight arrows identify dialectical defeats[xxi] *between arguments and highlight their specific targets.*

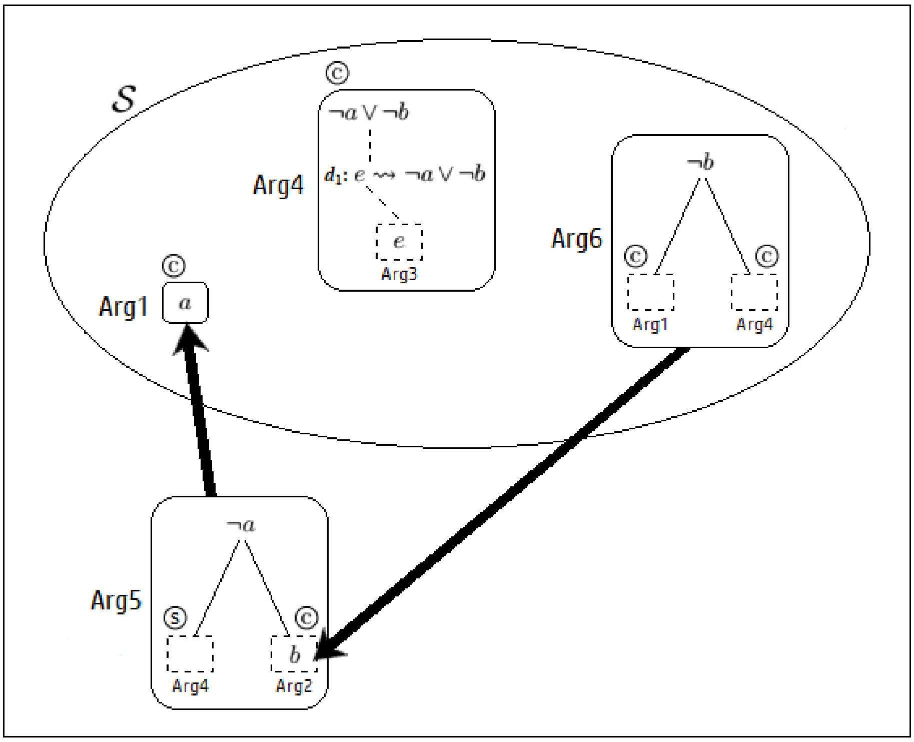

Figure 2.7: Example of D-ASPIC$^+$ arguments (solid squares), sub-arguments (dashed squares) and dialectical defeats (arrows).

*Figure 2.7 illustrates the dialectical defeat that occurs between Arg*5 *and Arg*1 *with respect to the set of arguments $\mathscr{S}$. That is to say, when challenging the acceptability of Arg*1*, Arg*5 *supposes Arg*4 *that belongs to the committed sub-arguments of $\mathscr{S}$. Then, Arg*4 *together with the sub-argument Arg*2 *concludes the formula $\neg a$. This conclusion targets the formula a and allows Arg*5 *to dialectically (undermine) defeat Arg*1*. On the other hand, argument Arg*6 *dialectically (undermine) defeats Arg*5*, defending in this way Arg*1 *and its acceptability with respect to $\mathscr{S}$.*

Finally, once again as in Dialectical Cl-Arg, D-ASPIC$^+$ enjoys properties P1-P4 [51] which suffice to satisfy the rationality postulate under minimal resource consumption[xxii].

[xxi] In this example, we are assuming that every attack succeeds as dialectical defeat.

[xxii] Although enjoying P4 under minimal assumptions might depend upon the employed proof-theoretical mean [51].

2.7 Argument Schemes

In everyday conversations, arguments are typically used to advocate or claim a conclusion based on the premises put forward as evidence to support the conclusion. However, the study of logic (and its subsequent development) in the past centuries has mainly stressed deductive logic and related arguments, making little to no space for the analysis of plausible reasoning. Using Douglas Walton's words:

> *"Recent concerns with the evaluation of argumentation in informal logic and speech communication have more and more begun to center around nondemonstrative arguments that lead to tentative (defeasible) conclusions, based on a balance of considerations. Such arguments do not appear to have structures of the kind traditionally identified with deductive and inductive reasoning. However, they are extremely common, and are often called "plausible" or "presumptive", meaning that they are only tentatively or provisionally acceptable, even when they are correct."*[136]

The same author proposes then the Argument Schemes (AS) model, i.e., forms of argument (structures of inference) that enable one to identify and evaluate common types of presumptive argumentation in everyday discourse [137]. Argument schemes are always paired with a set of specific critical questions that probe and assess the given argument in a particular case, in relation to the context in which it occurred[xxiii].

The evaluation of AS via critical questions may take place in two different manners, according to the most accredited theories presented in [140]: (a) *initiative shifting* and (b) *backup evidence*.

(a) After having moved a critical question, the initiative immediately shifts to the proponent that has to provide an answer or else the argument is considered defeated.

[xxiii] A recent study proposed by Lumer [84] thoroughly reviews Walton's AS approach and concludes by advocating its inadequacy in formalising argumentative reasoning.To clarify his statement, the author begins by defining four main principles that every argumentation theory should satisfy: (1) Approximate maximum epistemic effectiveness, i.e., lead to the best possible outcome (or, at least, an approximation of it); (2) Completeness; (3) Efficiency, i.e., aiming at the output with as little effort as possible; (4) Practical justifications of its criteria. Lumer then proceeds by showing how none of these conditions is fully met by Walton's approach. Indeed, the exhaustive list of schemes presented in [142] prove to be overabundant yet still incomplete (whilst also presenting superfluous or invalid schemes, the latter ensued from an absence of normative assessment). In addition, there is a lack of structural form and general arbitrary composition of the AS. Furthermore, their enthymematic patterns hinder the evaluation of their instantiations validity, and the critical questions do not seem to solve such a verification issue.

At the end of his research, however, the author suggests a solution to address the highlighted problems: "[...] *basing the construction of criteria for valid arguments on epistemological theories and principles like deductive logic, probability theory and rational decision theory."* [84]

That is to say, asking the question is enough to temporarily defeat the argument. Nevertheless, the proponent has the capability of retaining the argument validity by providing an appropriate answer to the critical question.

(*b*) On the other hand, the second theory states that asking a critical question does not suffice to defeat the argument. The question, if questioned, needs to be backed up with some evidence before it can shift any burden that would defeat the argument.

2.7.1 Argument Schemes Over Proposal for Action

As an example of an AS, we can illustrate the Argument Scheme Over Proposal for Action (ASOPA) which represents a rational inference pattern for proposing an action. Intelligent agents should be able to engage in practical reasoning in order to correctly interact with their environment and understand the best course of action to undertake in a given situation. As such, an AS that formalises the rationale underpinning this process may be particularly useful in AI research fields. According to the analysis conducted in [7], it results that the element of choice assumes an important role in practical reasoning:

> *"Given complete information, the world restricts us to a single rational choice of beliefs, but different people may rationally make different choices of goals and actions.* [...] *The "best"* [*course of actions*] *addresses the selection from the available options."*

Starting with the two AS for practical reasoning that Walton proposed in [138] (i.e., the necessary and the sufficient condition schemes), the authors of [7] 'split' Walton's notion of goal into three distinct elements: *states, goals and values*. *Goals* represent the effects that an agent wishes to attain, whereas *states* are the consequences of the undertaken action, whether the agents desire them or not. *Values*, in turn, provide the actual reasons for which an agent wishes to achieve a goal. According to Searle [114], the importance of values stems from the fact that they account for rational disagreement among people. On the other hand, a ranking of values may instead provide the preferred solution to a difference of opinions (although Searle specifies that such an ordering is the product, and not the presupposition, of practical reasoning).
These considerations lead to the introduction of the AS Over Proposal for Action:

AS Over Proposal for Action
Premise : In the current circumstances R *Premise* : we should perform action A *Premise* : to achieve new circumstances S *Premise* : which will realise some goal G *Premise* : which will promote some value v
Conclusion : A should be performed

The proposed scheme assumes the existence of:

- A finite set of distinct actions, called *Acts*, denoted with elements A, B, C, etc.
- A finite set of propositional formulae, called *Prop*, denoted with elements *p, q, w,* etc.
- A finite set of states, called *States*, denoted with elements *R, S, T,* etc. Each element describes a specific state of the world and corresponds to an assignment of truth values $\{T,F\}$ to each element of *Prop*.
- A finite set of propositional formulae, called *Goals*, denoted with elements *G, H* etc.
- A finite set of *Values* denoted with elements *v, w,* etc.
- A function *value* mapping each element of *Prop* to a pair $(v, sign)$, where $v \in Values$ and $sign \in \{+,-,=\}$.
- A ternary relation `apply` on $Acts \times States \times States$, with $\texttt{apply}(A,R,S)$ to be read as: *"Performing action A in state R, results in state S"*.

In addition, there are four statements that need to be satisfied if the argument represented by the formalisation is to be valid:

Statement 1: *R* is the case.

Statement 2: $\texttt{apply}(A,R,S) \in \texttt{apply}$.

Statement 3: $S \models G$ (i.e., "*G* is true in the state *S*").

Statement 4: $value(G) = (v,+)$.

We can represent the AS Over Proposal for Action following the diagrammatic form of [6]:

$$R \xrightarrow{A} S \models G \uparrow v$$

The intuitive meaning is: *"Performing action A in the current state R, results in a new state S that realises goal G and promotes value v"*.

The following list describes instead the set of critical questions paired with the scheme [7]:

- (**CQ1a**) Are there alternative ways of realising the same consequences?
- (**CQ1b**) Are there alternative ways of realising the same goal?
- (**CQ1c**) Are there alternative ways of promoting the same value?
- (**CQ2**) Is it possible to do action A?
- (**CQ3a**) Would doing action A promote some other value?
- (**CQ3b**) Does doing action A preclude some other action that would promote some other value?
- (**CQ4a\CQ4b**) Does doing action A have a side effect that demotes the value v\some other value?
- (**CQ5a\CQ5b**) Are the believed circumstances R possible and\or true?
- (**CQ5c**) Assuming CQ5a and CQ5b, has action A the stated consequences S?
- (**CQ5d**) Assuming CQ5a, CQ5b and CQ5c, will action A bring about the desired goal G?
- (**CQ6a**) Does goal G realise the intended value v?
- (**CQ6b**) Is the proposed value v a legitimate value?
- (**CQ7a**) Is situation S (believed by agent *a* to result from doing action A) a possible state of affairs?
- (**CQ7b**) Are the particular aspects of situation S (represented by goal G) possible?

The authors then use these sixteen questions as the basis for the development of a general theory of persuasion over action that results in a dialogue game protocol named the *PARMA Action Persuasion Protocol*.

In the forthcoming chapters, Argument Scheme theory will constitute the core dialectical block around which we will design the so-called Explanation-Question-Response (EQR) dialogue. The resulting protocol will harness the critical questions strategy and the general mechanism that informs the evaluation of AS instantiation in order to assess the explanation conveyed by the dialogue. The kernel of this procedure revolves around what we will denote as EQR scheme, the data unit that will store all the relevant information and conveniently deliver them upon request. ASOPA will be one of the two fundamental components that will structure such a new scheme.

The notions examined thus far represent the theoretical instruments and methods that will be deployed in the remainder of the thesis. In particular, the reviewed literature underlies the contents of the following chapters:

- Dung's AFs, Argument Games and Dialectical Cl-Arg lay the foundation for Chapter 3, where their combination yields Dialectical Argument Game Proof Theories.

- A similar background also underpins Chapter 4, where the additional contribution of the standard 3-value Labelling approach allows for the generation of Dialectical Labellings.

- Drawing from Dialogues and Argument Schemes, Chapter 5 and Chapter 7 will produce Explanation-Question-Response (EQR) patterns (Dailogue and Scheme) while examining their implementation in a clinical recommendation context.

- Leveraging also from Dung's AFs and D-ASPIC+, a resource-bounded real-world agents version of the EQR dialogue will be developed in Chapter 6 and then connected with the semantics formalism of Dialectical Argument Game Proof Theories and Labellings.

Chapter 3

Dialectical Argument Game Proof Theories

Argument game-based proof theories provide procedural structures capable of determining the status of an argument. Given an argumentation framework, argument games identify the membership of an argument to a specific extension simulating a dispute between two opposing contenders. The semantics meant to be captured dictates the rules of the played game, which serve to describe how the players can achieve victory. Dialectical Classical logic Argumentation (Dialectical Cl-Arg) is a recent approach that provides real-world dialectical characterisations of Cl-Arg arguments by resource-bounded agents while preserving the rational criteria established by the rationality postulates and practical desiderata. This chapter combines both subjects and introduces argument games for Dialectical Cl-Arg. The latter revolves around the core notion of *dialectical defeats*. Such defeats enable argumentative interactions more aligned with the dialectical reasoning of real-world resource-bounded agents. Thus, their presence requires the implementation of *dialectical argument game* proof theories capable of conveying the same idea as single-agent reasoning processes.

3.1 Developing Dialectical Argument Games

In the following sections, we are going to develop argument games for Dialectical Cl-Arg that accommodate the dialectical defeats and semantics introduced in Definition 23. The resulting proof theory will present some specific features that will distinguish it from the standard argument games, although the general structure remains similar. Intuitively, winning a dialectical game for an argument A means having a 'dialectical procedure' (depending on the semantics that the proof theory is meant to capture) for defending the information contained in A, hence showing the admissibility of the encoded data.

The main difference between a dispute tree $\mathscr{T}$ and a *dialectical dispute tree* $\mathscr{D}$ can be identified with the additional reference to a subset $\mathscr{S} \subseteq \text{PRO}(\mathscr{T})$. That is to say, $\mathscr{S}$ represents a candidate admissible set of PRO arguments such that PRO commits to their premises. Recall once again that, when challenging the acceptability of an argument with respect to a set $\mathscr{S}$, the defeating argument can suppose premises from all the arguments in $\mathscr{S}$. Whereas, the argument that defends $\mathscr{S}$ can only suppose the premises of the defeating argument. Another important difference between standard and dialectical games is that the latter handles *partially instantiated dialectical AFs* (*pdAFs*)[i]. As a consequence, each dialectical game enjoys specific properties that encapsulate the dialectical uses of arguments by real-world resource-bounded agents, thus succeeding in better approximating a process capable of bridging formal (proof-theoretical) and informal (real-world exchange of arguments) single-agent reasoning.

We can now formally introduce the (unique) dialectical dispute tree induced by A wrt a set $\mathscr{S}$:

Definition 32. [Dialectical Dispute Tree] *Let $\mathscr{T}$ be the dispute tree induced by* A *in a finite pdAF = $\langle AR, defeats\rangle$. Let also $\mathscr{S} \subseteq PRO(\mathscr{T})$. Then, the* dialectical dispute tree $\mathscr{D}$ *induced by* A *with respect to $\mathscr{S}$ is the dispute tree $\mathscr{T}$ pruned in a way such that $\forall X, Y \in AR$: X is a child of Y in $\mathscr{D}$ iff $defeats(X,Y)$ and:*

1. *If $X \in PRO(\mathscr{D})$ and $Y \in OPP(\mathscr{D})$, then $X \Rightarrow_{\{Y\}} Y$, i.e. X defeats Y and $Supp(X) \subseteq Prem(Y)$;*
2. *If $X \in OPP(\mathscr{D})$ and $Y \in PRO(\mathscr{D})$, then $X \Rightarrow_{\mathscr{S}} Y$, i.e. X defeats Y with respect to $\mathscr{S}$ and $Supp(X) \subseteq Prem(\mathscr{S} \cup \{Y\})$.*

The 'playing field' of the dialectical argument games (i.e., the data structure on the basis of which the games are played) is still depicted by the dispute tree $\mathscr{T}$. Indeed, the relationship existing between the dispute tree $\mathscr{T}$ induced by A in a finite pdAF and the dialectical dispute tree $\mathscr{D}$ induced by A wrt $\mathscr{S}$ is such that $\mathscr{D}$ is 'contained' in $\mathscr{T}$ (since $\mathscr{D}$ is a pruned version of $\mathscr{T}$), as shown in the following example.

[i]Refer to Proposition 1.

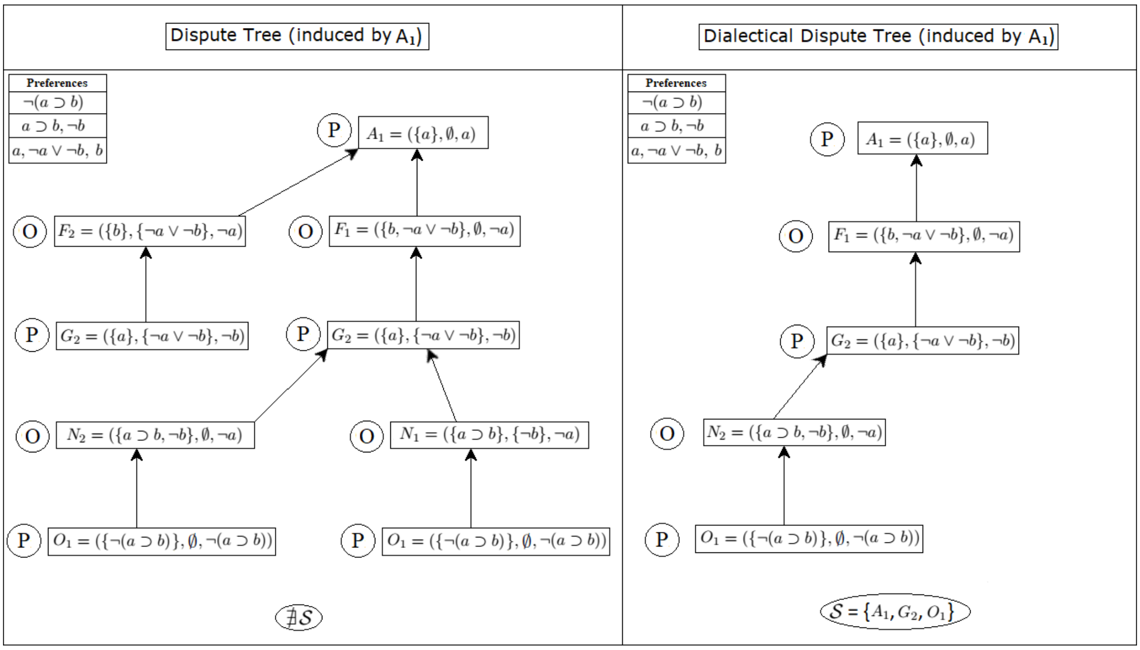

Figure 3.1: The (incomplete) dispute tree $\mathscr{T}$ (on the left) induced by A_1 in a finite pdAF = $\langle \text{AR}, \textit{defeats} \rangle$ and the corresponding (incomplete) dialectical dispute tree $\mathscr{D}$ (on the right) induced by A_1 wrt $\mathscr{S} = \{A_1, G_2, O_1\}$ in the same pdAF = $\langle \text{AR}, \textit{defeats} \rangle$.

Example 5. *Figure 3.1 presents the (incomplete) dispute tree $\mathscr{T}$ induced by A_1 in a finite pdAF = $\langle AR, defeats \rangle$ and the corresponding (incomplete) dialectical dispute tree $\mathscr{D}$ induced by A_1 wrt $\mathscr{S} = \{A_1, G_2, O_1\}$ in the same pdAF. Both trees are incomplete since the purpose of the example is just to show the relationship existing between them. For the same reason, we also avoid listing all the arguments of the pdAF.*

Observe that, unlike $\mathscr{T}$, where no set is taken into consideration, the defeats in $\mathscr{D}$ are parametrized to the set $\mathscr{S}$. This implies that, when defeating PRO's arguments, OPP can only suppose the premises of the arguments in the set $\mathscr{S}$ (besides the premises of the targeted argument). No such restrictions exist for $\mathscr{T}$. Notice that, even if we keep extending both trees, dispute $d = (P)A_1$—$(O)F_2$—$(P)G_2$ will never be part of $\mathscr{D}$. This is because, according to Definition 32 (which also emphasizes how dialectical defeats work), PRO can move G_2 only if $Supp(G_2) \subseteq Prem(F_2)$. However, this is never going to be the case since the formula $\neg a \vee \neg b \notin Prem(F_2)$. Therefore, even if the two trees were identical in every other branch, the absence of dispute d will still make $\mathscr{D}$ 'contained' in $\mathscr{T}$.

Dialectical argument games share with the standard argument games the notion of winning strategy: in order to win the game for an argument A, PRO must have a winning strategy for it. It will lose otherwise. However, the two definitions slightly differ since a dialectical winning strategy has to take into account the set $\mathscr{S}$ targeted by the dialectical defeats:

Definition 33. [Dialectical Winning strategy] *Let $\mathscr{D}$ be the dialectical dispute tree induced by* A *wrt $\mathscr{S}$ in a finite pdAF = $\langle AR, defeats \rangle$ and let d be a dispute in $\mathscr{D}$. Then, a* dialectical winning strategy *$\mathscr{W}$ for* A *corresponds to the dialectical dispute tree $\mathscr{D}$ pruned in a way such that:*

(33.1) *The set $\mathscr{W}_D$ of disputes in $\mathscr{D}$ is a non-empty finite set such that each dispute $d \in \mathscr{W}_D$ is finite and is won by PRO (i.e., $LAST(d) \in PRO(\mathscr{D})$);*

(33.2) *$\forall d \in \mathscr{W}_D$, $\forall d'$ such that d' is some sub-dispute of d, $LAST(d') = X$ and $X \in PRO(\mathscr{D})$, then $\forall Y \in OPP(\mathscr{D})$ such that $Y \Rightarrow_{\mathscr{S}} X$, there is a $d'' \in \mathscr{W}_D$ such that d'— Y is a sub-dispute of d''.*

Similarly to Definition 7, the previous definition states that a dialectical winning strategy corresponds to the dialectical dispute tree $\mathscr{D}$ pruned in a way such that (33.1) $\mathscr{W}_D$ is a non-empty finite set, its disputes are finite, end with a PRO argument and are such that (33.2) OPP has moved exhaustively and also PRO has countered each defeating argument moved by OPP. The difference is in the dialectical defeats: the nodes are no more connected by means of the defeats relations among arguments, but through dialectical defeats among arguments that target the set $\mathscr{S}$.

We now have all the elements needed to formally introduce the protocol of the dialectical admissible/preferred game. Similar to a list of instructions, this protocol determines the legal moves that can be performed by the players. The game unfolds as a result of the legal arguments played and terminates when there are no more valid moves available. When this happens, the status of the root of the tree is evaluated. The presence of a winning strategy for such an argument assigns the victory to PRO. Strictly speaking, OPP never wins: its purpose is to counter each argument moved by the proponent in order to assist it in testing the admissibility of the root argument (indeed, argument games are formalisations of single-agent reasoning processes). Nevertheless, OPP can still prevent PRO's victory by invalidating its winning strategy[ii].

[ii]In other words, OPP and PRO are two 'virtual entities' both played by the same real-world agent. The overall idea of argument game proof theories is to formalise a single-agent reasoning process (that can be interpreted as a sort of introspective reflection), allowing to establish the validity of the information embedded in the initial argument posited by PRO. OPP serves only in the role of a 'tester'.

3.1.1 Progressively Constructing Dialectical Dispute Trees

When we play a Φ-dialectical game we are increasingly building, starting from the root A and following the legal moves licensed by the protocol Φ, a dialectical dispute tree denoted as Φ-$\mathscr{D}^n$. Each node of such tree corresponds to an argument progressively played by either PRO or OPP that is labelled with a positive integer i (with $1 \leq i \leq n$). These additional labels allow identifying the order in which the arguments have been played, hence determining the current stage (i.e., the *n*th-stage) of the Φ-dialectical game. Recall that the dispute tree $\mathscr{T}$ induced by A represents the playing field of the games, and every Φ-dialectical game for A is contained within its data structure (i.e., Φ-$\mathscr{D}^n$ is a 'pruned-version' of $\mathscr{T}$). Moreover, being a dialectical dispute tree, even Φ-$\mathscr{D}^n$ is constructed wrt a set $\mathscr{S} \subseteq$ PRO($\mathscr{T}$), however, such $\mathscr{S}$ can gradually increase with each new move made by PRO during the game. Indeed, $\mathscr{S}$ is composed of the same arguments moved by PRO in Φ-$\mathscr{W}^n$ (i.e., a dialectical winning strategy for A of Φ-$\mathscr{D}^n$), which can be extended while the game proceeds[iii]. As it will be shown, observe also that $\mathscr{S}$ is still a different set than PRO(Φ-$\mathscr{W}^n$), meaning that it will modify its members according to the changes in PRO(Φ-$\mathscr{W}^n$), but it will never be empty even if there is no winning strategy Φ-$\mathscr{W}^n$.

In order to formally describe a Φ-dialectical game, we first need to define a *partial dialectical dispute tree* $\mathscr{D}^n$ which will stand as a potential 'game template' deprived of a protocol:

Definition 34. [Partial dialectical dispute tree] *A partial dialectical dispute tree* $\mathscr{D}^n$ *induced by* A *wrt* $\mathscr{S} \subseteq PRO(\mathscr{T})$ (*with* $\mathscr{S} \neq \emptyset$) *in a finite pdAF* = $\langle AR, defeats\rangle$ *is the* (*upside-down*) *tree that starts from the argument* A*, and it is progressively built up to the nth-move by one of the two players, such that each node of the tree is labelled with a positive integer* i (*for* $1 \leq i \leq n$). *Moreover, every branch of the tree* (*from root to leaf*) *constitutes a different dispute. Also* $\forall X, Y \in AR$: *X is a child of Y in* $\mathscr{D}^n$ *iff* $defeats(X,Y)$ *and:*

1. *If* $X \in PRO(\mathscr{D}^n)$ *and* $Y \in OPP(\mathscr{D}^n)$, *then* $X \Rightarrow_{\{Y\}} Y$, *i.e. X defeats Y and* $Supp(X) \subseteq Prem(Y)$;
2. *If* $X \in OPP(\mathscr{D}^n)$ *and* $Y \in PRO(\mathscr{D}^n)$, *then* $X \Rightarrow_{\mathscr{S}} Y$, *i.e. X defeats Y with respect to a set* $\mathscr{S}$ *and* $Supp(X) \subseteq Prem(\mathscr{S} \cup \{Y\})$.

[iii] Although the set $\mathscr{S}$ can increase the number of its members while the game goes on, it can never exceed the size of PRO($\mathscr{T}$). Indeed, keep in mind that every Φ-dialectical game for A is contained in the dispute tree $\mathscr{T}$ induced by A (since $\mathscr{T}$ corresponds to the playing field of the game).

Finally, $\mathscr{W}^n$ will denote a dialectical winning strategy for A *of $\mathscr{D}^n$ as per Definition 33 (substituting $\mathscr{D}$ with $\mathscr{D}^n$).*

Every stage of a Φ-dialectical game can then be identified with a specific dialectical dispute tree Φ-$\mathscr{D}^n$, i.e., a *partial dialectical dispute tree* of Definition 34 where each of its nodes also fulfils the legal move requirements according to the protocol Φ. Consider that every such stage of the game is not unique: playing the same game multiple times does not necessarily hold the same Φ-$\mathscr{D}^n$ at identical stages *n*. They can indeed differ depending on the way in which the legal arguments have been deployed by the players. As we are going to see, this notion is essential for a proper account of the dialectical defeats in the game protocol[iv].

3.1.2 Disqualified Defeats

It is interesting to notice that, during a Φ-dialectical game, a dialectical defeat that occurred in an early stage of the game might not take place in a more advanced phase of the same game. This can be caused by an update of the current $\mathscr{S}$, the set parametrized by OPP for performing dialectical defeats. We denote this anomaly as 'disqualified defeats'.

Definition 35. [Disqualified dialectical defeats] *Let Φ-$\mathscr{D}^n$ be the dialectical dispute tree of a Φ-dialectical game built up to the nth-move where X and Y denote arguments played respectively by OPP and PRO in Φ-$\mathscr{D}^n$. Let also $X \Rightarrow_{\mathscr{S}} Y$ by supposing $\alpha \in Prem(\mathscr{S})$. If, after the game goes on, we will reach a stage Φ-$\mathscr{D}^{n+k}$ (for $k > 0$) where $\alpha \notin Prem(\mathscr{S})$, then the defeat moved by X against Y will be invalidated and will be denoted as 'disqualified'. As such, X and all the arguments following it in the same dispute will be (temporarily) pruned from the tree.*

Consider indeed that the status of disqualified defeats might be temporary and be updated again in a further stage of the game (when these defeats will become valid once more). Definition 35 entails the following proposition:

Proposition 4. *Let Φ-$\mathscr{D}^n$ be the dialectical dispute tree of a Φ-dialectical game built up to the nth-move:*

[iv]Observe that it is possible for one (or more, depending on the protocol) dialectical winning strategy Φ-$\mathscr{W}^n$ for *A* of Φ-$\mathscr{D}^n$ to exist, although there is no dialectical winning strategy $\mathscr{W}$ for *A* of $\mathscr{D}$. This can happen, for example, when $\mathscr{D}$ is composed only by infinite disputes (recall that we need finite disputes to have winning strategies, as stated by Definition 33.1), whilst Φ-$\mathscr{D}^n$ is composed by finite disputes, due to the restrictions imposed by the protocol Φ. In this situation, it is possible to identify in Φ-$\mathscr{D}^n$ a winning strategy Φ-$\mathscr{W}^n$. Such an example is illustrated in Figure 3.2(b).

(I) *If the* n*th-move is an argument X played by OPP, then moving X cannot disqualify the dialectical defeat that X performs against a PRO argument.*

(II) *The presence of OPP arguments whose defeats have been disqualified will not affect the dialectical winning strategy.*

Proof.

(I) Since X is the last argument (legally) played in Φ-$\mathscr{D}^n$, it trivially does not comply with Definition 35.

(II) Even if the dialectical defeats moved by OPP arguments have been disqualified (hence are no more a threat for PRO), the requirements of the dialectical winning strategy have not changed. That is to say, every dispute of Φ-$\mathscr{W}^n$ must terminate with a PRO argument (Definition (33.1)).

□

Notice that every dialectical game protocol Φ takes into account disqualified defeats, which are then also contemplated by the dialectical dispute tree Φ-$\mathscr{D}^n$ (and dialectical winning strategy Φ-$\mathscr{W}^n$).

3.2 Dialectical Admissible/Preferred Games

We can now formally introduce the protocol for the dialectical admissible/preferred game. As already stated, during each dialectical argument game, the players have to comply with a protocol Φ that identifies the legal moves allowed.

Definition 36. [Dialectical Admissible Game legal moves] *Let $\mathscr{D}^n$ and $\mathscr{W}^n$ be defined as in Definition 34, let d be a dispute of $\mathscr{D}^n$ and d' be a sub-dispute of d. Let also $(PL_n)X$ (for $n > 0$) denotes the argument X played by either one of the two players $(P$ or $O)$ as the $(last)$ nth-move. Then Φ_P identifies legal moves in the following way:*

(36.0) *PRO moves the first argument.*

(36.1) *If $(PL_n)X$ and $n = 2k$ (for $k > 0$), then the next move $n+1$, say Y, is by PRO and it is such that:*

(a) $Y \Rightarrow_{\{Z\}} Z$, where $Z \in OPP(\mathscr{D}^n)$;

(b) There exists a $\mathscr{W}^{n+1}$ for A of $\mathscr{D}^{n+1}$.

(36.2) *If* $(PL_n)X$ *and* $n = 2k+1$ *(for* $k \geq 0$*), then the next move* $n+1$*, say Y, is by OPP and it is such that:*

(a) *Y* $\Rightarrow_{\mathscr{S}}$ *Z, where* $Z \in \mathscr{S}$ *and* $\mathscr{S} := PRO(\mathscr{W}^n)$[v]*;*

(b) *If* $d = d'$*—Z, then* $Y \notin OPP(d')$*;*

(c) *For each* $d = d'$*—J—*$\cdots$*, where* $J \in OPP(\mathscr{D}^n)$ *and its defeat has been disqualified, then* $LAST(d) = LAST(d')$ *until next OPP's turn.*

A Φ_P-dialectical game is said to be terminated when, during its turn, the corresponding player runs out of the legal moves identified by (36.1) or (36.2) of the protocol Φ_P. PRO wins only if it has a winning strategy once the game terminates. It loses otherwise.

The previous protocol can be informally summarised as follows. PRO starts the game by playing the first argument [(36.0)] and, after that, OPP will make its move. Then, the two players alternate in playing only one argument at a time to reply to one of their counterpart's arguments. Observe that when $\mathscr{S}$ is initialized in the game and, subsequently, every time its arguments are updated by the changes in PRO($\mathscr{W}^n$) [(36.2(a))], it is always the beginning of OPP's turn. This means that the condition for which $\mathscr{S} \neq \emptyset$ (as per Definition 34) is continuously respected[vi].

Notice that the established protocol allows *backtracking* to other arguments. That is to say, when PRO moves it can either target the last argument played by OPP or another argument moved by OPP in the dialectical dispute tree generated thus far (i.e., an argument member of the set OPP($\mathscr{D}^n$)) [(36.1(a))]. Similarly, when OPP moves it can either target the last argument played by PRO or another argument moved by PRO in the current dialectical winning strategy (i.e., an argument member of the set PRO($\mathscr{W}^n$)) [(36.2(a))]. The *relevance conditions* [(36.1(b)) for PRO; (36.2(a)) for OPP] ensure that: after PRO has made its move, there will be a winning strategy $\mathscr{W}^{n+1}$, hence providing the victory to PRO; after OPP has moved, instead, the previous winning strategy will cease to exist, thus preventing PRO from winning. That is to say, PRO will be forced to generate a dialectical winning strategy during each of its turns, while OPP will have to invalidate such a winning strategy during every one of its turns.

The restriction (36.2(b)) on the moves played by OPP is necessary (as also shown in

[v]The symbol ':=' denotes a variable initialization rather than an equivalence relation. That is to say, at the beginning of each OPP's turn, the content of $\mathscr{S}$ is initialized to the current PRO($\mathscr{W}^n$), i.e, the arguments member of $\mathscr{S}$ are the same as PRO($\mathscr{W}^n$). This operation overwrites the previous contents of $\mathscr{S}$.

[vi]That is because a situation in which $\mathscr{S} = $ PRO$(\mathscr{W}^n) = \emptyset$ never occurs at the beginning of OPP's turn.

the standard games of [97, 132] and [31]). Indeed, allowing OPP to repeat its arguments, since OPP is required to move exhaustively, could imply the generation of infinite disputes. To see why let us suppose that $(\mathrm{PL}_n)X$ (for $n > 1$) identifies an argument X played by either one of the two players (denoted as P or O) as its nth move in a Φ-dialectical game. Then, there could be an infinite dispute d like the following:

$$d = (P_1)A\text{—}\cdots\text{—}(O_n)Y\text{—}(P_{n+1})Z\text{—}(O_{n+2})Y\text{—}(P_{n+3})Z\text{—}(O_{n+4})Y\text{—}\cdots$$

Intuitively, since Z is capable of defending itself by defeating Y, there is no need to further extend the dispute by repeating the same arguments: this is because Z has already shown its acceptability wrt PRO($\mathscr{W}^{n+1}$). Therefore, the only way for avoiding infinite disputes (and infinite dialectical admissible/preferred games) is to prevent OPP from repeating its arguments in the same disputes.

Finally, $(36.2(c))$ ensures that the disqualified defeats (Definition 35) are taken into account throughout the game. That is to say, whenever a dialectical defeat moved by an argument J is disqualified, the protocol guarantees the pruning of J and all the arguments that follow in the same dispute, until the next turn of OPP, when a new check for disqualified defeats will occur.

Remark 1. *Similarly to the standard argument games presented in* [*97*]*, the protocol of the dialectical admissible games is identical to the protocol of the dialectical credulous preferred games. Indeed, it suffices to show the membership of an argument* A *in an admissible extension to show that* A *is credulously justified under the preferred semantics as well. That is because every admissible extension of a dialectical AF is a subset of a preferred extension. This is a consequence of the Fundamental Lemma* (*Lemma 2*) *and its entailed property* (*Proposition 2*)*.*

3.2.1 Soundness and Completeness

As it has been defined, the admissible/preferred game satisfies the properties of soundness and completeness. This proves the equivalence existing between the victory of the Φ_P-dialectical game for an argument A and the membership of the same A to an admissible/preferred extension of the corresponding finite pdAF.

Theorem 10. *Let* Φ_P*-*$\mathscr{D}^n$ *identifies a terminated* Φ_P*-dialectical game for* A*. Then, there exists a dialectical winning strategy* Φ_P*-*$\mathscr{W}^n$ *for* A*, such that the set PRO*(Φ_P*-*$\mathscr{W}^n$) *of arguments moved by PRO in* Φ_P*-*$\mathscr{W}^n$ *is conflict-free, iff* A *is included in an admissible extension* Adm *of the pdAF.*

Proof. **[Soundness]** We have to prove that if A is a member of the conflict-free set PRO(Φ_P-$\mathscr{W}^n$), then $A \in \mathsf{Adm}$. To simplify the notation, let E = PRO(Φ_P-$\mathscr{W}^n$). Assume that A is a member of the conflict-free set E, then:

- By Definition 33.2, the existence of the winning strategy implies that: each argument played by OPP against arguments moved by PRO in the winning strategy has been successfully countered by PRO. That is to say, $\forall X \in$ E, if $\exists Y \in$ AR such that $Y \Rightarrow_E X$, then $\exists$Z $\in$ E, such that $Z \Rightarrow_{\{Y\}} Y$, ensuring in this way that X is acceptable wrt E.
- Recall that the set of disputes of Φ_P-$\mathscr{W}^n$ is finite and composed of finite disputes (by Definition 33.1). As such, E is composed of a finite number of arguments.

We have thus shown that E is a finite, conflict-free set and every argument in E is acceptable wrt it. Therefore, E corresponds to an admissible extension, hence, if A is a member of the conflict-free set PRO(Φ_P-$\mathscr{W}^n$), then $A \in \mathsf{Adm}$. □

Proof. **[Completeness]** We are going to prove Completeness by showing that if $A \in \mathsf{Adm}$, then A is a member of the conflict-free set PRO(Φ_P-$\mathscr{W}^n$). We are going to do this by constructing a dialectical winning strategy Φ_P-$\mathscr{W}^n$ for A.

- Assume that $A \in \mathsf{Adm}$. Since the pdAF is finite, then it is also finitary, meaning that every argument in Adm has a finite number of defeaters. Then we can build a dialectical winning strategy Φ_P-$\mathscr{W}^n$ for A if PRO starts the game with A and: for each argument Y dialectically defeating A and moved by OPP, PRO chooses one argument X from Adm (even A itself) such that X$\Rightarrow_{\{Y\}}$Y. Notice that the generation of infinite disputes is prevented by the admissible/preferred protocol (Definition 36.2(b)). This procedure can be repeated for every argument Z dialectically defeating X, and so on, until OPP runs out of legal moves according to the protocol Φ_P (which will happen for sure since A is a member of an admissible set).

The result will be a dialectical winning strategy Φ_P-$\mathscr{W}^n$ for A, hence, A is a member of the conflict-free set PRO(Φ_P-$\mathscr{W}^n$). We have thus shown that, if $A \in \mathsf{Adm}$, then A is a member of the conflict-free set PRO(Φ_P-$\mathscr{W}^n$). □

3.3 Dialectical Grounded Games

The dialectical grounded game protocol Φ_G enjoys the same notations and definitions introduced thus far, but presents also important differences compared to the dialectical admissible/preferred game. Indeed, the protocol should be designed such that, when the game terminates and PRO is the winner, the set PRO(Φ_G-$\mathscr{W}^n$) of arguments moved by PRO in a dialectical winning strategy Φ_G-$\mathscr{W}^n$ is a subset of the grounded extension Grd of the pdAF. In this way, by iterating the framework characteristic function $\mathscr{F}$ from PRO(Φ_G-$\mathscr{W}^n$), we are able to obtain the grounded extension Grd. However, recall that it is the monotonicity of the function, in the case of a finitary pdAF[vii], that ensures the construction of the least fixed point of $\mathscr{F}$ which corresponds to the grounded extension.

In Dialectical Cl-Arg [49] the monotonicity of $\mathscr{F}$ holds only under the domain of *epistemically maximal* (*em*) admissible sets of arguments (described in Definition 25). Then, to get the grounded extension via the iteration of $\mathscr{F}$ from the set PRO(Φ_G-$\mathscr{W}^n$), we will need PRO(Φ_G-$\mathscr{W}^n$) to be *em*. Otherwise, we might have to face a situation in which argument A, whose membership in Grd we wanted to test via the dialectical grounded game, is not acceptable wrt Grd, although $A \in$ PRO(Φ_G-$\mathscr{W}^n$). To address this issue, we are going to adapt the protocol Φ_G accordingly.

Definition 37. [Dialectical Grounded Game legal moves] *Let $\mathscr{D}^n$ and $\mathscr{W}^n$ be characterized as in Definition 34, let d be a dispute of $\mathscr{D}^n$ and d' be a sub-dispute of d. Let also $(PL_n)X$ (for $n > 0$) denote the argument X played by either one of the two players (P or O) as the (last) nth-move. Then Φ_G identifies legal moves in the following way:*

(37.0) *PRO moves the first argument.*

(37.1) *If $(PL_n)X$ and $n = 2k$ (for $k > 0$), then the next move $n+1$, say Y, is by PRO and it is such that:*

(a) $Y \Rightarrow_{\{Z\}} Z$, where $Z \in OPP(\mathscr{D}^n)$;

(b) There exists a $\mathscr{W}^{n+1}$ for A of $\mathscr{D}^{n+1}$;

(c) If $d = d'$—Z, then $Y \notin PRO(d')$.

(37.2) *If $(PL_n)X$ and $n = 2k+1$ (for $k \geq 0$), then the next move $n+1$, say Y, is by OPP and it is such that:*

(a) $Y \Rightarrow_{\mathscr{S}} Z$, where $Z \in \mathscr{S}$ and $\mathscr{S} := PRO(\mathscr{W}^n)$.

[vii] Being finitary, it can be shown that $\mathscr{F}$ is also $\omega-$continuous (as explained in [53] for standard AFs and in [49] for pdAFs).

(b) *For each* $d = d'$—J—$\cdots$, *where* $J \in OPP(\mathscr{D}^n)$ *and its defeat has been disqualified, then* $LAST(d) = LAST(d')$ *until next OPP's turn.*

(37.3) *If, at the beginning of its turn, OPP cannot perform the move described by* $(37.2(a)$), *then apply function* Cl_{em} (*Definition 25*) *on PRO*($\mathscr{W}^n$).

Notice that a Φ_G-dialectical game is said to be terminated when, during its turn, at least one player runs out of the legal moves identified by (37.1) or (37.2) of the protocol Φ_G. PRO wins only if it has a winning strategy once the game terminates. It loses otherwise.

As per Definition 36, the previous protocol can be informally summarised as follows. PRO starts the game by playing the first argument [(37.0)] and after that OPP will make its move. Then, the two players alternate in playing only one argument at a time to reply to one of their counterpart's arguments. Observe that when $\mathscr{S}$ is initialized in the game and, subsequently, every time its arguments are updated by the changes in PRO($\mathscr{W}^n$) [(37.2(a))], it is always the beginning of OPP's turn. This means that the condition for which $\mathscr{S} \neq \emptyset$ (as per Definition 34) is continuously respected.

Notice also that the established protocol allows *backtracking* to other arguments. That is to say, when PRO moves it can either target the last argument played by OPP or another argument moved by OPP in the dialectical dispute tree generated thus far (i.e., an argument member of the set OPP($\mathscr{D}^n$)) [(37.1(a))]. Similarly, when OPP moves it can either target the last argument played by PRO or another argument moved by PRO in the current dialectical winning strategy (i.e., an argument member of the set PRO($\mathscr{W}^n$)) [(37.2(a))]. The *relevance conditions* [(37.1(b)) for PRO; (37.2(a)) for OPP] ensure that: after PRO has made its move, there will be a winning strategy $\mathscr{W}^{n+1}$, hence providing the victory to PRO; after OPP has moved, instead, the previous winning strategy will cease to exist, thus preventing PRO from winning. That is to say, PRO will be forced to generate a dialectical winning strategy during each of its turns, while OPP will have to invalidate such a winning strategy during every one of its turns.

The restriction (37.1(c)) emphasises the additional burden of proof entailed by the membership to the grounded extension. This is intuitively captured by the idea that in defending an argument X's membership to the grounded extension Grd, PRO must 'appeal to' some argument other than X itself. This is reflected in the game by the fact that PRO cannot repeat the arguments it has already moved in the same disputes.

Moreover, $(37.2(b))$ ensures that the disqualified defeats (Definition 35) are taken into account throughout the game. That is to say, whenever a dialectical defeat moved by an argument J is disqualified, the protocol guarantees the pruning of J and all the arguments that follow in the same dispute, until the next turn of OPP, when a new check for disqualified defeats will occur.

Finally, in light of the previously underlined epistemically maximal requirement, an additional one-time move has been included. According to the notions described in [49], adding all arguments up to some i to a set E, and then *em* closing, yields the same result as adding each argument one by one and closing prior to each subsequent addition. As such, once the game is terminated in favour of PRO and immediately before PRO is declared the winner, it suffices to apply function Cl_{em} (Definition 25) over the resulting set PRO($\mathscr{W}^n$) rendering it *em*, therefore, a subset of the grounded extension of the pdAF.

3.3.1 Soundness and Completeness

In the following proofs, we are going to employ the framework characteristic function $\mathscr{F}_p$, which iterates over admissible epistemically maximal extensions:

Definition 38. *Let $\langle AR, defeats \rangle$ be a pdAF and AR_p the set of all the em admissible subset of AR. Then $\mathscr{F}_p : AR_p \mapsto AR_p$, where $\mathscr{F}_p(E) = Cl_{em}(\mathscr{F}(E))$.*

We can now show that the dialectical grounded game satisfies the properties of soundness and completeness.

Theorem 11. *Let Φ_G-$\mathscr{D}^n$ identifies a terminated Φ_G-dialectical game for* A.
Then, there exists a dialectical winning strategy Φ_G-$\mathscr{W}^n$ for A*, such that the em closure $Cl_{em}(PRO(\Phi_G\text{-}\mathscr{W}^n))$ of the set of arguments moved by PRO in Φ_G-$\mathscr{W}^n$ is conflict-free, iff* A *is included in the grounded extension* Grd *of the pdAF.*

To simplify the notation, let us abbreviate $Cl_{em}(\text{PRO}(\Phi_G\text{-}\mathscr{W}^n))$ in Cl_{em}.

Proof. **[Soundness]** We have now to prove that if A is a member of the conflict-free set Cl_{em}, then $A \in$ Grd. Hence, assuming that A is a member of the conflict-free set Cl_{em}:

- Clearly, all of Φ_G-$\mathscr{W}^n$ leaves, say X_i, are in $\mathscr{F}_p(E_0)$ since they have no defeaters and are then acceptable wrt $\emptyset$. Now, consider that in every branch of Φ_G-$\mathscr{W}^n$, the arguments defended[viii] by each X_i are acceptable with respect to $\mathscr{F}_p(E_0)$ and so are in $\mathscr{F}_p(E_1)$. This process can be repeated until, say, $\mathscr{F}_p(E_i)$ when the root A of Φ_G-$\mathscr{W}^n$ is reached. Since $Cl_{em} \subseteq \mathscr{F}_p(E_i)$, and further iterations of $\mathscr{F}_p(E_i)$ will yield the generation of the least fixed point Grd, then A will be a member of Grd.

This suffices to show that if A is a member of the conflict-free set Cl_{em}, then $A \in$ Grd. □

Proof. **[Completeness]** We have to prove that if $A \in$ Grd, then A is a member of the conflict-free set Cl_{em}. Employing the acceptable arguments in the characteristic function $\mathscr{F}_p$ we are going to show that we can build a Φ_G-winning strategy for A.

- Assume that $A \in$ Grd. Since the pdAF is finite, it is also finitary, hence we know that there is a least number i such that $A \in \mathscr{F}_p(E_i)$. Then we will have a dialectical winning strategy Φ_G-$\mathscr{W}^n$ for A if PRO starts the game with A and: for each argument Y dialectically defeating A and moved by OPP, PRO chooses one argument X from $\mathscr{F}_p(E_{i-1})$ such that X$\Rightarrow_{\{Y\}}$Y. This procedure can be iterated for every argument Z dialectically defeating X, and so on, until PRO can choose an argument from $\mathscr{F}_p(E_0)$. $\mathscr{F}_p(E_0)$ has no defeaters and, as such, OPP cannot play any legal move (licensed by the protocol Φ_G) against it. Finally, the grounded game protocol will also ensure the epistemically maximality of the set of arguments moved by PRO in Φ_G-$\mathscr{W}^n$ (37.3).

The result yields a dialectical winning strategy Φ_G-$\mathscr{W}^n$ for A, such that A is a member of the conflict-free set Cl_{em}. We have thus shown that, if $A \in$ Grd, then A is a member of the conflict-free set Cl_{em}. □

[viii] Recall that an argument X defends an argument Z iff: when $\exists Y \in AR$ such that Y defeats Z, then X defeats Y.

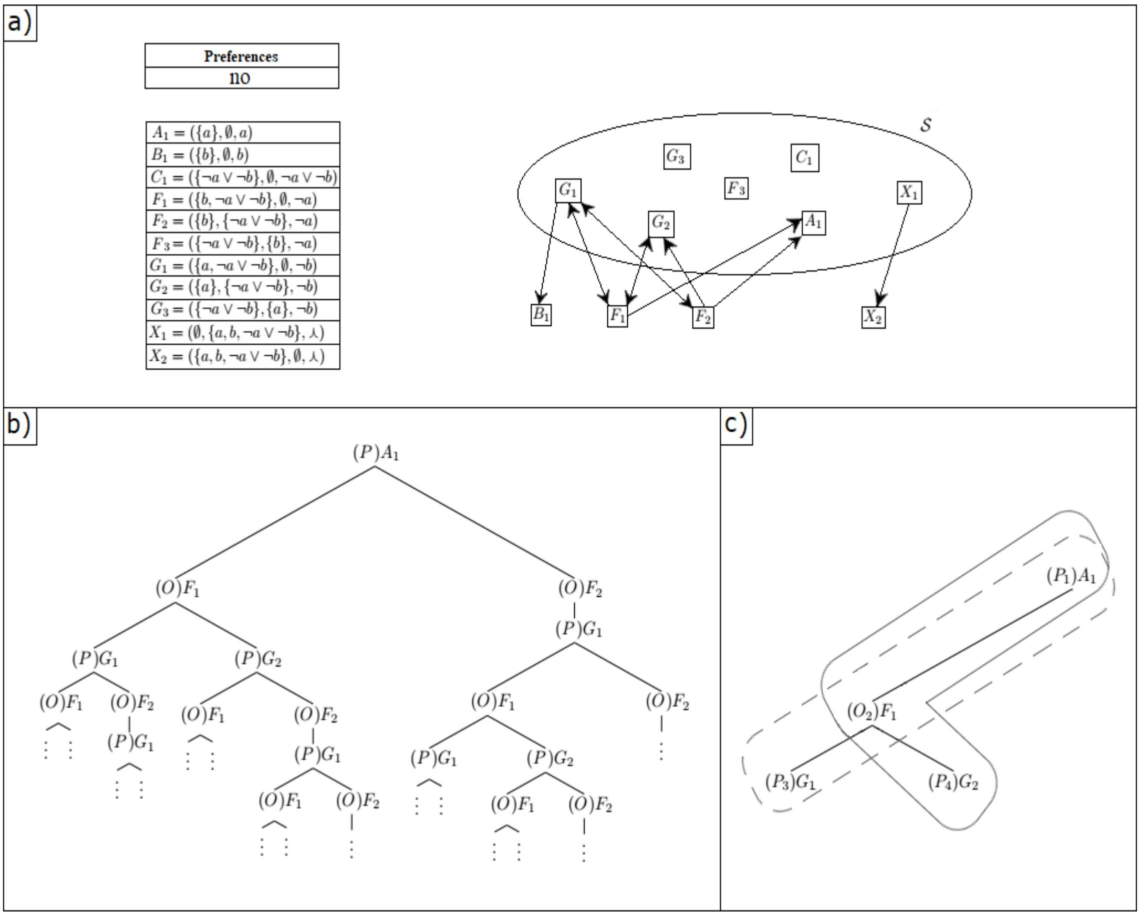

Figure 3.2: Figure a) illustrates a pdAF with a list of its arguments and the set $\mathscr{S}$ that is parametrized by the dialectical defeats. Consider also that X_2 is defeated by all the arguments of the pdAF, except A_1, B_1, and C_1 (the arrows that should have highlighted such defeats have been omitted to avoid unnecessary graphical confusion). Figure b) displays the dialectical dispute tree $\mathscr{D}$ induced by A_1 wrt $\mathscr{S}$ in the pdAF of Figure a). Notice that $\mathscr{D}$ is composed of infinite disputes (the vertical dots represent the endless continuation of the disputes), as such, it does not have a winning strategy. A dialectical dispute tree Φ-$\mathscr{D}^n$, with $n = 4$, is depicted in Figure c) and corresponds to a Φ-dialectical game played up to the nth-move. Observe that the number of each move (next to the label P or O) represents the order in which the arguments have been played in the game. In this example, we are assuming a protocol Φ that licenses legal moves where PRO can play more than one argument per turn, therefore, Φ-$\mathscr{D}^n$ has two winning strategies (both of which are encircled in the figure).

3.4 Main Features of Dialectical Argument Games

Dialectical argument games hold specific features that differentiate them from the standard argument games of [97, 31, 132] and depend upon their protocols and the properties possessed by each pdAF (especially P1, P2 and P3). Although, for convenience, we are going to outline these features using the dialectical admissible/preferred game (Definition

36), notice that the choice of the protocol is irrelevant.

3.4.1 Feature 1 (F1)

(F1) The set of all the arguments moved by PRO in a dialectical winning strategy (i.e., PRO(Φ_P-$\mathscr{W}^n$)), is always conflict-free;

Every pdAF = $\langle$AR, *defeats*$\rangle$ prevents any conflicts existing between arguments in a set E $\subseteq$ AR if each argument in E is acceptable with respect to it. Since this has already been formally proved and shown[ix], here we will try to explain it through an example. Notice also the rationale underpinning *F1*: due to their limited resources, it would be unrealistic to demand that real-world agents actually perform conflict-free checks on every set E of arguments.

Example 6. *Consider a pdAF that includes the arguments listed in Figure 2.4 and such that all the arguments composing the set PRO(Φ_P-$\mathscr{W}^n$) are acceptable wrt it (as it normally is for PRO(Φ_P-$\mathscr{W}^n$)). To simplify the notation, let $E = PRO(\Phi_P\text{-}\mathscr{W}^n)$.*
Among the arguments of E, suppose that there are two conflicting arguments as $G_2 = (\{a\},\{\neg a \vee \neg b\},\neg b)$ and $F_1 = (\{b,\neg a \vee \neg b\},\emptyset,\neg a)$: we are going to show how this will lead to a contradiction. Due to property P1, $A_1 = (\{a\},\emptyset,a) \in AR$. Hence, by property P3, $X_2 = (\{a,b,\neg a \vee \neg b\},\emptyset,\curlywedge) \in AR$ and by property P2, $X_1 = (\emptyset,\{a,b,\neg a \vee \neg b\},\curlywedge) \in AR$. However, if this is the case, $X_1 \Rightarrow_E G_2$ (and, similarly, $X_1 \Rightarrow_E F_1$). Since X_1 is unassailable, $\nexists Z \in E$ such that $Z \Rightarrow_{\{X_1\}} X_1$ and this will contradict the assumption that all the arguments members of E are acceptable wrt to it. Therefore, since all the arguments that compose the set PRO(Φ_P-$\mathscr{W}^n$) are acceptable wrt it, PRO(Φ_P-$\mathscr{W}^n$) must be conflict-free.

3.4.2 Feature 2 (F2)

(F2) The relevance conditions, i.e., the conditions of the protocol that compel both players to change the outcome of the game at the end of every turn, are essential to the unfolding of the dialectical argument games. This also justifies why the set $\mathscr{S}$ cannot be initialized with any set other than PRO(Φ_P-$\mathscr{W}^n$);

[ix]**Lemma 17** of [49] states that: *Let $E \subseteq AR$ such that every argument in E is acceptable wrt E, and AR satisfies $P1, P2$ and $P3$. Then E is conflict free.*
The proof can be found in the same paper.

The relevance conditions (36.1(b) and 36.2(a) of Definition 36) can be summarised as the conditions that force the two players to change the outcome of the game at the end of every turn[x]. These requirements are fundamental for real-world agents that reason with limited availability of resources. Indeed, it would be illogical to allow such players to move arguments useless for the result of the game: this would simply mean wasting valuable resources[xi].

Moreover, the relevance conditions clarify why the set $\mathscr{S}$, referenced in the admissible/preferred protocol, corresponds to the current set of arguments moved by PRO in Φ_P-$\mathscr{W}^n$, that is to say, PRO(Φ_P-$\mathscr{W}^n$). This, in turn, allows avoiding a specific issue that could permanently prevent the victory of PRO, as the following example will show.

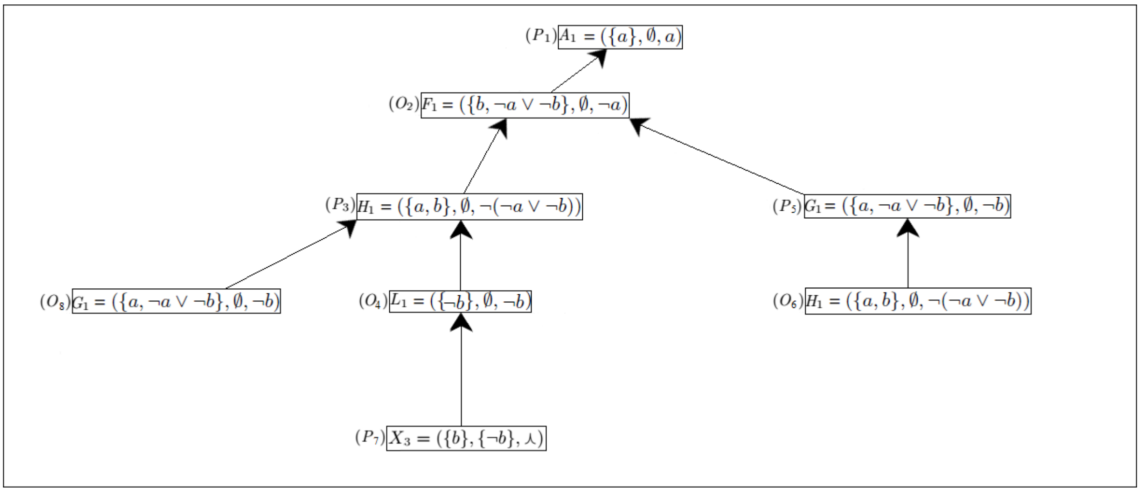

Figure 3.3: The Figure illustrates a dialectical dispute tree Φ_P-$\mathscr{D}^n$, hence generated following the protocol for the dialectical admissible/preferred games. Notice that the arrows indicate the defeats between the arguments. Starting with the root argument A_1, the other arguments are played according to the order highlighted by the numbers near their labels (P or O). The last player to move is OPP, which moves G_1. Since $G_1 \Rightarrow_{\mathscr{S}} H_1$ (where $\mathscr{S} :=$ PRO(Φ_P-$\mathscr{W}^{n\text{-}1}$), i.e., $\mathscr{S} = \{A_1, H_1, X_3\}$) and $G_1 \not\approx (\{b\}, \emptyset, b)$, this ensures OPP invalidates the winning strategy Φ_P-$\mathscr{W}^{n\text{-}1}$. Hence, there is no winning strategy in Φ_P-$\mathscr{D}^n$.

Example 7. *The examples of Figures 3.3, 3.4 and 3.5 depict a dialectical admissible game played using the arguments of Table 2.4, where* $F_1 \not\approx (\{a\}, \emptyset, a)$, $\forall T \in \{G_1, L_1\}$, $T \not\approx (\{b\}, \emptyset, b)$, $\forall V \in \{N_3, X_3\}$, $V \not\approx (\{\neg b\}, \emptyset, \neg b)$, *while* $H_1 \not\approx (\{\neg a \vee \neg b\}, \emptyset, \neg a \vee \neg b)$. *Starting with the root* A_1, *the order in which the arguments are played is outlined*

[x]The research presented in [102] introduces a series of relevant properties for dialogue protocols. Property *R1* seems quite similar to our relevance conditions, although our study concerns argument games rather than dialogues.

[xi]Notice that we are dealing with pdAFs, and so, small subsets of the respective overall set of arguments of the considered framework. As such, positing only relevant arguments is not going to be particularly expensive for agents' resources.

in the brackets, next to the labels P and O. The dialectical dispute tree Φ_P-$\mathscr{D}^n$ (*Figure 3.3*) *has been generated following the protocol for the dialectical admissible/preferred games, however, its extension into* Φ-$\mathscr{D}^{n+1}$ (*Figure 3.4*) *does not take into account PRO's relevance condition* (*36.1*(*b*) *of Definition 36*)*. This immediately raises an issue: without the relevance condition, we could have to face a situation in which PRO is still losing even after its turn has ended* (*Figure 3.4*)*. In this circumstance, during the next turn of OPP, there will be no winning strategy, hence no set of arguments moved by PRO in* Φ_P-$\mathscr{W}^{n+1}$ (*i.e., the set PRO*(Φ_P-$\mathscr{W}^{n+1}$))*, that can be targeted as* $\mathscr{S}$*. Suppose, for*

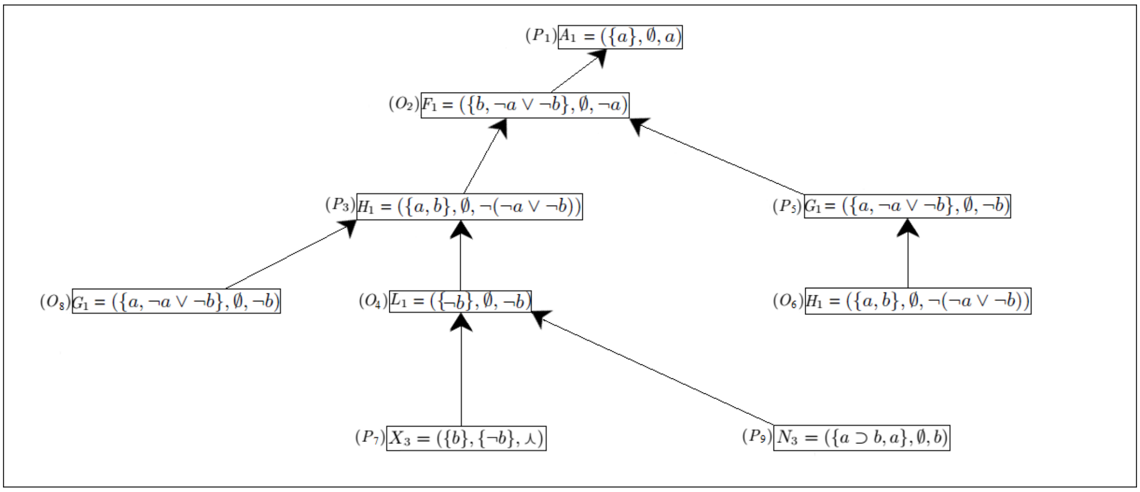

Figure 3.4: The Figure illustrates the extension of the dialectical dispute tree Φ_P-$\mathscr{D}^n$ into Φ_P-$\mathscr{D}^{n+1}$ due to argument N_3 played by PRO. As we can see, if PRO's relevance condition is dropped, then PRO is free to move any argument and not only the ones that will reinstate the winning strategy. $N_3 \Rightarrow_{\{L_1\}} L_1$ and $N_3 \not\rightharpoonup (\{\neg b\}, \emptyset, \neg b)$. However, this implies that, even after PRO moves, there is no winning strategy in Φ_P-$\mathscr{D}^{n+1}$ (because the argument G_1 played by OPP has not yet been defeated).

the sake of the example, that the protocol of the game allows searching for another set $\mathscr{S}$*. What could then be the set* $\mathscr{S}$ *parametrised by the dialectical defeats moved by OPP? Without PRO*(Φ_P-$\mathscr{W}^{n+1}$) *the only reasonable alternative is to consider a different set* $\mathscr{S}$ *initialized in a way such that* $\mathscr{S} \subseteq$ *PRO*(Φ_P-$\mathscr{D}^{n+1}$)*. Nevertheless, notice that if OPP is allowed to suppose the premises of arguments in a non-conflict-free set* $\mathscr{S}$*, then OPP would have enough resources for playing an unassailable argument* (*as* X_1)*. As shown in Figure 3.5, if* H_1*,* $G_1 \in \mathscr{S}$*, then* $B_1 \in AR$ *by property P1 of the pdAF. By P3,* $X_2 \in AR$*, while by property P2, also* $X_1 \in AR$ (*since* X_1 *is the logically equivalent argument of* X_2)*. Argument* X_1 *constitutes the problem: it defeats* A_1 *and has empty premises, which implies it cannot be defeated. This means that, by playing* X_1*, OPP will change the final outcome of the game invalidating any other possible attempt from PRO of reinstating the winning strategy. However, this happened in the example because there*

was no set PRO$(\Phi_P\text{-}\mathscr{W}^{n+1})$ *and OPP had to suppose the premises of the arguments members of a different set* $\mathscr{S} \subseteq$ *PRO*$(\Phi_P\text{-}\mathscr{D}^{n+1})$ *which was not conflict-free. In other words, unassailable arguments as* X_1 *can be moved only when* (i) *arguments that defeat each other or* (ii) *unconditional arguments with conflicting conclusions are in* $\mathscr{S}$*. Moving such arguments will immediately trigger property P3 of the pdAF, which will highlight the inconsistency of their premises, while property P2 will ensure the generation of the corresponding unassailable argument.*
Nevertheless, without requiring a resource-consuming conflict-free check on every $\mathscr{S} \subseteq$ *PRO*$(\Phi_P\text{-}\mathscr{D}^{n+1})$*, how would it be possible to ensure the conflict-freeness of the set* $\mathscr{S}$*? The only set of arguments moved by PRO which satisfies this condition* (*without requiring a conflict-free check*) *in a dialectical argument admissible game is the set PRO*$(\Phi_P\text{-}\mathscr{W}^{n+1})$*, thanks to property F1. Therefore,* $\mathscr{S}$ *has to be initialized to PRO*$(\Phi_P\text{-}\mathscr{W}^{n+1})$.

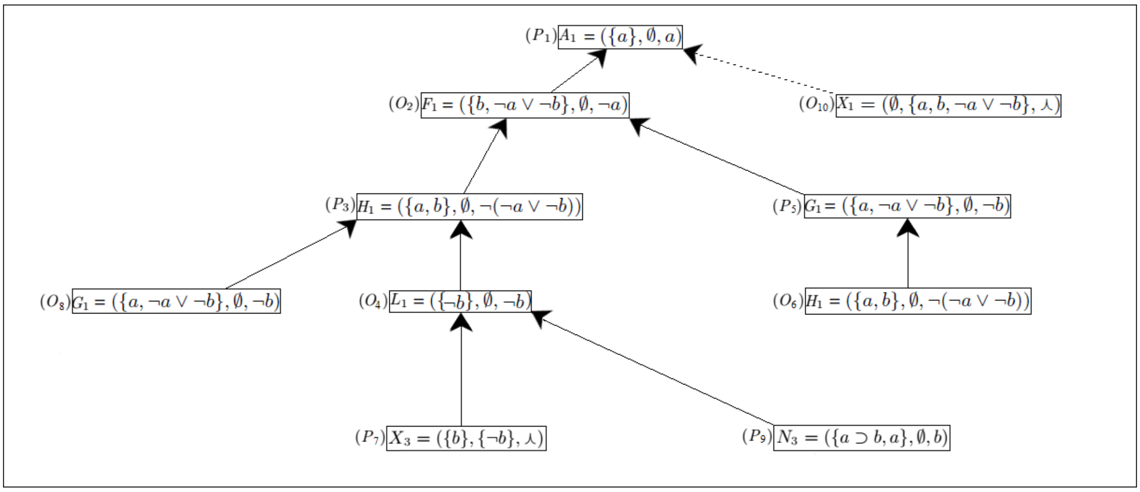

Figure 3.5: The Figure illustrates the extension of the dialectical dispute tree $\Phi_P\text{-}\mathscr{D}^{n+1}$ into $\Phi_P\text{-}\mathscr{D}^{n+2}$ due to argument X_1 played by OPP. It is possible to move X_1 because there is no winning strategy in $\Phi_P\text{-}\mathscr{W}^{n+1}$, hence there is no set PRO$(\Phi_P\text{-}\mathscr{W}^{n+1})$: this forces OPP to target the premises of a different set $\mathscr{S}$, initialized in a way such that $\mathscr{S} \subseteq$ PRO$(\Phi_P\text{-}\mathscr{D}^{n+1})$ (in the case of the example, $\mathscr{S} :=$ PRO$(\Phi_P\text{-}\mathscr{D}^{n+1})$, i.e., $\mathscr{S} = \{A_1, H_1, G_1, X_3, N_3\}$). The danger of arguments such as X_1 lies in their unassailability and the fact that they always succeed as defeats (underlined by the dashed arrow in the picture and explained in Definition 21). That is to say, the final outcome of the game can then be changed if $\mathscr{S} \neq$ PRO$(\Phi_P\text{-}\mathscr{W}^{n+1})$ because it can allow OPP to move arguments as X_1 against the root of the tree (preventing PRO from reinstating any other possible winning strategy).

The implication of what has been shown in Example 7 is that the relevance conditions need to be part of the protocols of any dialectical argument game. Indeed, if this is not the case, we could have to face a situation in which PRO is still losing even after its turn has ended. In this circumstance, during the next turn of OPP, there will be no set

PRO(Φ_P-$\mathscr{W}^n$) that can be used to initialise $\mathscr{S}$. Hence, once again, the issue outlined in Example 7 could arise and change the final outcome of the game by permanently invalidating PRO's winning strategy. This then means that $\mathscr{S}$:= PRO(Φ_P-$\mathscr{W}^n$) and cannot be otherwise.

3.4.3 Feature 3 (F3)

Before the introduction of the third feature (F3) enjoyed by the dialectical admissible/preferred argument games, we need to formally define the *uniqueness* of the dialectical winning strategy regardless of the employed protocol.

Definition 39. [Uniqueness of the dialectical winning strategy] *Let $\mathscr{D}^n$ and let $\mathscr{W}^n$ be defined as in Definition 34. Then $\mathscr{W}^n$ is said to enjoy the uniqueness property if there is no other dialectical winning strategy for* A *wrt $\mathscr{S}$ simultaneously present in $\mathscr{D}^n$.*

Let us consider a dialectical dispute tree $\mathscr{D}^n$ identical (although without the implementation of a specific game protocol) to the one in Figure 3.2(c). This tree has two winning strategies, say $\mathscr{W}_1^n$ and $\mathscr{W}_2^n$, each of which is composed of a single dispute. That is to say: $d_1 = (P_1)A_1$—$(O_2)F_1$—$(P_3)G_1$ and $d_2 = (P_1)A_1$—$(O_2)F_1$—$(P_4)G_2$, such that $\mathscr{W}_1^n$ is composed of d_1, while $\mathscr{W}_2^n$ is composed of d_2. Obviously, $\mathscr{D}^n$ does not enjoy the uniqueness property. Indeed, both G_1 and G_2 defeat the same argument F_1, whereas only one of such defeats is actually needed. This implies that it suffices that either $\mathscr{W}_1^n$ or $\mathscr{W}_2^n$ is present for PRO to win (at least temporarily) the game. For the final outcome of the game, it is pointless to have both winning strategies simultaneously. It is also resource-consuming, meaning that it does not comply well with the Dialectical Cl-Arg purpose of capturing resource-bounded real-world agents' reasoning. Indeed, although there may be multiple winning scenarios in real-world situations, they cannot be realized together due to the agents' limited availability of resources (i.e., it is not possible to actualize multiple dialogue paths concurrently).

(F3) Any dialectical winning strategy Φ_P-$\mathscr{W}^n$ enjoys the uniqueness property.

Uniqueness is a property enforced on a dialectical winning strategy Φ_P-$\mathscr{W}^n$ by the protocol of the dialectical admissible/preferred argument game. Uniqueness is certainly a desirable property since it allows for shorter and simpler games. This ensures a quicker evaluation of the status of the dialectical dispute tree root.

The following Lemma shows that the protocol of the dialectical admissible/preferred game ensures the uniqueness of Φ_P-$\mathscr{W}^n$.

Lemma 3. *Let Φ_P-$\mathscr{D}^n$ identifies a Φ_P-dialectical game for* A. *Then, there exists only one dialectical winning strategy Φ_P-$\mathscr{W}^n$ for* A *wrt $\mathscr{S}$ that is simultaneously present in Φ_P-$\mathscr{D}^n$.*

Proof. Since the protocol of the admissible/preferred game forces the players to move only one argument per turn, the only other way to have multiple winning strategies simultaneously is by having different arguments moved by PRO (in different turns) that defeat the same argument played by OPP. We are going to show how this case cannot occur under the Φ_P protocol.

Let d_1 be a dispute in Φ_P-$\mathscr{W}^n$ and d' a sub-dispute of d_1. Let also $d_1 = d'$—$(O_{n-i})Y$ —$(P_{n-i+1})X$, for $n-i > 1$. As usual, the index near the player labels denotes the order in which the moves have been played. Suppose the game proceeds further and no changes occur in d_1. If the last argument moved is an argument $Z \neq X$ from PRO that dialectically defeats Y and generates $d_2 = d'$—$(O_{n-i})Y$—$(P_n)Z$, which is another dispute in Φ_P-$\mathscr{D}^n$ and d' is a sub-dispute of d_2 as well, then it is easy to see that PRO has played against the protocol Φ_P. That is because:

- If PRO defeats an argument without affecting the existing game status it will violate its relevance condition (Definition 36.1(b)).

Playing argument Z will then be prevented by PRO's relevance condition, ensuring in this way the uniqueness of the dialectical winning strategy Φ_P-$\mathscr{W}^n$. □

3.5 Efficiency Improvements

The protocols thus far developed can benefit from a range of efficiency improvements. They follow from the properties of the dialectical games and Dialectical Cl-Arg in general, which means that they will preserve the already proven soundness and completeness results. In particular, we can obtain shorter games thanks to (I1), which allows us to avoid meaningless repetitions of defeated arguments from OPP. Moreover, (I2) and (I3) show how, due to the features enjoyed by the dialectical games and without additional restrictions on the legal moves available to the players (unlike in [97]), it is possible to obtain other specific efficiency improvements. In the next section, these enhancements will be examined and, when required, also formalised and integrated into the protocols of the dialectical games.

3.5.1 List of Efficiency Improvements for Dialectical Games

In the admissible/preferred dialectical game, OPP is forbidden to repeat any arguments (*and not just in a dispute*) *which have already been defeated, and not defended or indirectly defended by another argument, in the game.*[xii]

Let's assume that OPP's argument Y has been defeated, and not defended, by an argument X moved by PRO in a dispute d. If now OPP repeats Y in a different dispute, then PRO can simply repeat X defeating Y once again.

Example 8. *For instance, let Φ_P-$\mathscr{D}^n$ be a dialectical dispute tree and let d be a dispute in Φ_P-$\mathscr{D}^n$. Suppose also that X is an argument moved by PRO in d, while Y is an argument played by OPP in d such that $X \Rightarrow_{\{Y\}} Y$. Then, if the game goes on* (*up to $n+k$ moves, for $k \geq 1$*)*, whenever Y will 'appear' in a different dispute, PRO can simply play X again. As such, playing argument Y proves to be just a waste of resources.*

We can now formalise this idea by substituting condition $(36.2b)$ from the protocol Φ_P (Definition 36) with the following constraint $(I1)$. The purpose of forbidding such moves is to avoid extending the game by adding useless sequences of arguments to it:

Definition 40 (Improved legal move)**.** *The following additional constraint for OPP* (*where OPP's argument Y is the next move played in the game*) *substitutes* $(36.2b)$ *from the protocol Φ_P:*

$(I1)$ *If $\exists J \in OPP(\Phi$-$\mathscr{D}^n)$ such that J is defeated and not defended* (*neither directly nor indirectly defended*)*, then $Y \neq J$.*

The soundness and completeness results of the dialectical games will not be affected by restriction $(I1)$, as the following lemma will prove:

Lemma 4. *Let Φ_P-$\mathscr{D}^n$ identifies a terminated Φ_P-dialectical game for* A*. Then, there exists a dialectical winning strategy Φ_P-$\mathscr{W}_1^n$ for* A*, iff there exists a dialectical winning strategy Φ_P-$\mathscr{W}_2^n$ for* A *constructed using a protocol that employs* I1.

Proof.

[xii] According to the recursive definition of indirect defence, an argument *X indirectly defends* an argument *A* if: *i*) *X defends A*; *ii*) *X defends Z*, and *Z indirectly defends A*.

[$\rightarrow$] If there exists a dialectical winning strategy Φ_P-$\mathscr{W}_2^n$, then there also trivially exists a dialectical winning strategy Φ_P-$\mathscr{W}_1^n$. Indeed, if OPP cannot repeat its defeated (and not defended) arguments $(I1)$, it cannot as well repeat its arguments in the same disputes ((36.2b) of Definition 36). That is to say, Φ_P-$\mathscr{W}_2^n$ follows every requirement established by protocol Φ_P.

[$\leftarrow$] We are going to show that every dialectical winning strategy Φ_P-$\mathscr{W}_1^n$ can be transformed into a dialectical winning strategy Φ_P-$\mathscr{W}_2^n$. Suppose that there is a dispute d in Φ_P-$\mathscr{W}_1^n$ in which it appears the sequence J—X of arguments such that J is moved by OPP, X is moved by PRO and $X \Rightarrow_{\{J\}} J$. We also know that J is not defended (or indirectly defended) because, being a dispute in the winning strategy, d terminates with a PRO argument. Notice that, since J is an OPP argument moved in a dispute, it must be preceded by a PRO argument. Hence, if we now remove every other J—X—$\cdots$ sequence (including whatever follows after X) from the dialectical winning strategy, we will not affect PRO's victory and we will generate a new dialectical winning strategy, i.e., Φ_P-$\mathscr{W}_2^n$.

$\square$

The following improvements are similar to the ones already introduced in [97], with an important difference. Unlike the standard games, the dialectical games do not need to enforce specific restrictions on their protocols in order to benefit from these efficiency enhancements: they are ensured by the properties enjoyed by any dialectical game.

$(I2)$ *PRO does not move self-defeating arguments* (*i.e., arguments which defeat themselves*)*;*

Whenever a self-defeating argument, say X, is played by PRO, PRO violates property *F1*. Indeed, even if X reinstates a dialectical winning strategy Φ-$\mathscr{W}^n$, the same X will also conflict with an argument member of PRO(Φ-$\mathscr{W}^n$), i.e., X itself.

$(I3)$ *PRO does not play an argument that defeats* (*or is defeated by*) *an argument in PRO*(Φ-$\mathscr{W}^n$)

That is to say, PRO does not move arguments that conflict with the arguments it has already moved in the winning strategy. Indeed, if PRO plays an argument X defeated by (a member of) PRO(Φ-$\mathscr{W}^n$) or that defeats an argument member of PRO(Φ-$\mathscr{W}^n$), the resulting winning strategy will not be conflict-free. This will then violate property *F1*.

Example 9. *Consider the dialectical dispute tree of Figure* 3.1 *and assume that, rather than* G_2*, PRO has moved argument* G_1*. Assume also that OPP makes an additional move* $H_1 = (\{a,b\}, \emptyset, \neg(\neg a \vee \neg b))$ *such that* $H_1 \Rightarrow_{\mathscr{S}} G_1$*. Then, PRO decides to counter its opponent by playing argument* $F_1 = (\{b, \neg a \vee \neg b\}, \emptyset, \neg a)$ *such that* $F_1 \Rightarrow_{\{H_1\}} H_1$ *on* $(\{a\}, \emptyset, a)$*. However, since* F_1 *defeats, hence conflicts, with* $G_1 \in PRO(\Phi\text{-}\mathscr{W}^n)$ (F_1 *is also dialectically defeated by* G_1*) this move will violate property* F1 (*the situation will then be similar to the one described by Example 6*).

Remark 2. *Notice that* (*I3*) *also subsumes the fact that* PRO does not move an argument X in a dispute d if such argument has already been played by OPP in d. *Indeed, playing argument X will reinstate the dialectical winning strategy* $\Phi\text{-}\mathscr{W}^n$*. However, at the same time, X is an argument moved by OPP* (*hence X complies with OPP's relevance condition*)*. As such, playing X will imply defeating once again an argument in PRO*(Φ-$\mathscr{W}^n$)*, violating property F1*[xiii].

As shown, (I2) and (I3) follow directly from the property *F1*, which is enjoyed by any dialectical game. As such, no modifications to the game protocols are needed, meaning that the soundness and completeness results will be preserved.

3.6 Future Work

Initially introduced in [48], the dialectical approach of Dialectical Cl-Arg has been subsequently examined from different perspectives. For example, the investigation concerning argumentative characterisations of Brewka's Preferred Subtheories (PS) [23] showed that, compared with the standard approach, the grounded semantics applied to Dialectical Cl-Arg more closely approximates sceptical inference in PS [50]. In addition, the research presented in [51] provides a full rational account of structured (ASPIC$^+$) arguments under resource bounds by adapting the approach of Dialectical Cl-Arg.

Extending further the study commenced in [32] and continued in [33], we plan to increase the range of dialectical argument games protocols investigating the stable [31], semi-stable [25] and ideal semantics [55, 26]. Another possible research direction that

[xiii]It is interesting to observe that this is not generally the case if PRO (i) repeat an OPP argument or (ii) an already defeated PRO argument, say X, in a different dispute of the dialectical dispute tree. That is because it might be that the opponent cannot suppose anymore the same premises that (ii) allowed it to defeat X the first time or (i) allowed it to defeat an argument in PRO(Φ-$\mathscr{W}^n$). For example, assume that an argument Y moved by OPP dialectically defeated X drawing its suppositions α from Prem(PRO(Φ-$\mathscr{W}^n$)). However, after the game goes on, it might be that $\alpha \notin$ Prem(PRO(Φ-$\mathscr{W}^{n+k}$)). Then Y cannot dialectically defeat X anymore (i.e., Y defeat against X is disqualified), therefore X is now a perfectly viable move for PRO.

will be pursued involves generalising the developed dialectical argument games to dialogues, following the guidelines of the already existing literature in the field (mainly [89, 104, 43]). This would have the interesting consequence of allowing to move from non-monotonic single-agent inference to distributed non-monotonic reasoning.

3.7 Conclusion

The main aspects of the real-world uses of argumentation by resource-bounded agents include: (i) showing inconsistencies in arguments by supposing the premises of the opponent's arguments; (ii) handling only finite subsets of the arguments of the AFs; (iii) reducing the consumption of resources by employing dialectical means (while still satisfying rationality postulates and practical desiderata) [49]. These features would constitute the hallmarks of an argument game based on Dialectical Cl-Arg, thus capable of better approximating non-monotonic single-agent real-world reasoning processes than the standard argument games. In this chapter, we have achieved some important results. We have developed argument game proof theories (denoted as *dialectical argument games*) for the admissible, preferred, and grounded semantics of Dialectical Cl-Arg. Incorporating dialectical defeats in the standard structure of the argument games proved to be a non-trivial process which yielded the discovery of interesting properties that differentiate dialectical games from the standard ones. That is to say, dialectical games enjoy (a) specific relevance conditions that characterise their protocols and yield (b) the uniqueness of their winning strategies, whilst property *F1* ensures (c) the conflict-freeness of the set of arguments moved by the proponent in the winning strategy. The latter is of particular importance since it provides the games with a range of different efficiency improvements. Without the need to perform any additional checks or to enforce additional restrictions in the protocols (unlike in [97]), *F1* allows each dialectical game to prevent the proponent from: playing self-defeating arguments; playing arguments already moved by the opponent (in the same dispute); and playing arguments that defeat (or are defeated by) other arguments already moved by the proponent. Finally, another efficiency improvement can be obtained if the opponent is forbidden to repeat arguments that have already been defeated in the dialectical admissible/preferred game, such that none of them has also been defended or indirectly defended by other arguments.

Chapter 4

Dialectical Labellings

The 3-value (IN, OUT, UNDEC) labellings method is a well-known approach for evaluating the arguments of an argumentation framework (AF). Similarly to the extension-based approach [53], the labelling method assigns a status to each argument in the AF according to its acceptability. Dialectical Classical logic Argumentation (Dialectical Cl-Arg) is a recent approach that provides real-world dialectical characterisations of Cl-Arg arguments by resource-bounded agents while preserving the rational criteria established by the rationality postulates and practical desiderata. This chapter combines both subjects and introduces labellings and labelling procedures for Dialectical Cl-Arg, highlighting the properties enjoyed by these labellings (called 'dialectical labellings') in comparison with the standard ones. Algorithms designed for dialectical labellings will benefit from such properties and the entailed efficiency improvements.

4.1 Developing Dialectical Labellings

The following sections introduce the dialectical labelling approach and the corresponding algorithms, outlining procedures capable of generating each different dialectical labelling according to the considered semantics (admissible, grounded, preferred). On one hand, the characterisation of the new method relies on an adaptation of the dialectical defeats (which will include the reference to a set $\mathscr{S}$ of arguments) to the standard labelling approach illustrated in [97]. On the other hand, the procedural aspect of the dialectical labellings requires more careful considerations. For example, the algorithm that produces the dialectical grounded labellings needs to be substantially modified. In Dialectical Cl-Arg, the iteration of the characteristic function, which yields the constructive definition of the grounded extension, requires the epistemic maximality of the admissible sets generated at each iteration by the function. Since the dialectical algorithm simulates this operation, it is essential to include an epistemically maximal check (and, if needed, a transformation) of the set in its procedure. Another contribution presented in this chapter regards the procedure for enumerating each dialectical preferred labelling, whose innovative idea is to take advantage of the properties characterising Dialectical Cl-Arg arguments.

Finally, soundness and completeness results will be provided for each newly introduced algorithm, along with some efficiency improvements that will enhance their performances.

Definition 41 (Dialectical labelling function)**.** *Let* $\langle AR, defeats \rangle$ *be a finite pdAF.*

- *A dialectical labelling is a total function* $d\mathscr{L}: AR \mapsto \{\text{IN, OUT, UNDEC}\}$*;*
- *We define:* $\mathsf{in}(d\mathscr{L}) = \{X \mid d\mathscr{L}(X) = \text{IN}\}$*;* $\mathsf{out}(d\mathscr{L}) = \{X \mid d\mathscr{L}(X) = \text{OUT}\}$*;* $\mathsf{undec}(d\mathscr{L}) = \{X \mid d\mathscr{L}(X) = \text{UNDEC}\}$.

The set $\mathsf{in}(d\mathscr{L})$ consists of the justified arguments labelled *IN*, whereas according to Dialectical Cl-Arg, a set of justified arguments is identified with the set $\mathscr{S}$. This implies that, when assigning the labels of the dialectical arguments according to the existing dialectical defeats, we have to consider that $\mathscr{S} = \mathsf{in}(d\mathscr{L})$.

Definition 42 (Dialectical legal labelling)**.** [34] *Let* $d\mathscr{L}$ *be a dialectical labelling for the pdAF =* $\langle AR, defeats \rangle$*, let also X, Y* $\in AR$ *and* $\mathscr{S} = \mathsf{in}(d\mathscr{L})$*.*

- *X is dialectically legally* IN *iff X is dialectically labelled* IN *and for every Y that dialectically defeats X, i.e., Y* $\Rightarrow_{\mathscr{S}}$ *X, it implies that Y is dialectically labelled* OUT
- *X is dialectically legally* OUT *iff X is dialectically labelled* OUT *and there is at least one Y that dialectically defeats X, i.e., Y* $\Rightarrow_{\{X\}}$ *X, such that Y is dialectically labelled* IN*;*
- *X is dialectically legally* UNDEC *iff X is dialectically labelled* UNDEC *and not every Y that dialectically defeats X, i.e., Y* $\Rightarrow_{\{X\}}$ *X, is such that Y is dialectically labelled* OUT*, and there is no Y that dialectically defeats X such that Y is dialectically labelled* IN.

Notice that whenever an argument Y (dialectically labelled *IN, OUT* or *UNDEC*) dialectically defeats an argument X dialectically labelled *IN*, it does so wrt a set $\mathscr{S}$ of arguments, i.e., $Y \Rightarrow_{\mathscr{S}} X$. However, no set $\mathscr{S}$ of arguments is targeted if X is dialectically labelled *OUT* or *UNDEC*, i.e. $Y \Rightarrow_{\{X\}} X$.

Definition 43 (Dialectical illegal labelling)**.** *Let pdAF* = $\langle AR, defeats \rangle$*,* $X \in AR$ *and* `l` $\in$ {IN, OUT, UNDEC}. *X is said to be dialectically illegally* `l` *iff X is dialectically labelled* `l`*, and it is not dialectically legally* `l`.

Similarly as presented in [97] a dialectically illegally *IN* argument can also be *super-illegally IN*:

Definition 44 (Dialectically super-illegally *IN* arguments)**.** *An argument X in d*$\mathscr{L}$*, that is dialectically illegally* IN*, is said to be dialectically* super-illegally *IN iff there is at least one Y that dialectically defeats X, i.e.* $Y \Rightarrow_{\mathscr{S}} X$*, such that Y is legally* IN *or* UNDEC.

As in the standard labelling case, it is possible to show the correspondence existing between dialectical extensions and dialectical labellings. The proofs do not contain significant differences from the ones of [25] (except for the use of dialectical labellings and dialectical defeats rather than standard labellings and defeats) and, as such, they will be omitted.

Definition 45 (Dialectical labelling semantics)**.** *A dialectical admissible labelling* $d\mathscr{L}$ *is a dialectical labelling without arguments that are dialectically illegally IN and without arguments that are dialectically illegally OUT;*

- *A dialectical complete labelling* $d\mathscr{L}$ *is a dialectical admissible labelling without arguments that are dialectically illegally UNDEC;*
- $d\mathscr{L}$ *is a dialectical grounded labelling iff* $d\mathscr{L}$ *is a dialectical complete labelling and there does not exist a dialectical complete labelling* $d\mathscr{L}'$ *such that* $\mathsf{in}(d\mathscr{L}') \subset \mathsf{in}(d\mathscr{L})$*;*
- $d\mathscr{L}$ *is a dialectical preferred labelling iff:*
    - $d\mathscr{L}$ *is a dialectical complete labelling and there does not exist a dialectical complete labelling* $d\mathscr{L}'$ *such that* $\mathsf{in}(d\mathscr{L}') \supset \mathsf{in}(d\mathscr{L})$*,*

– *or, equivalently, $d\mathscr{L}$ is a dialectical admissible labelling and there does not exist a dialectical admissible labelling $d\mathscr{L}'$ such that* $\mathsf{in}(d\mathscr{L}') \supset \mathsf{in}(d\mathscr{L})$ [i].

Theorem 12. *Let $\langle AR, defeats\rangle$ be a pdAF and $E \subseteq AR$. For* s $\in$ *{admissible, complete, grounded, preferred}: E is an* s *extension of the pdAF iff there exists a dialectical* s *labelling $d\mathscr{L}$ with* $\mathsf{in}(d\mathscr{L}) = E$.

4.1.1 Main properties of Dialectical Labellings

Dialectical labellings hold specific features that differentiate them from the standard labellings presented in [97]. Such features can be listed as follows. Let $d\mathscr{L}$ be a dialectical labelling for the pdAF = $\langle AR, defeats\rangle$:

$(Q1)$ In every $d\mathscr{L} = all\text{-}IN$, there is always at least one argument dialectically legally *IN*;

$(Q1')$ In every $d\mathscr{L} = all\text{-}IN$, each unassailable argument is dialectically legally *IN*

$(Q2)$ In every $d\mathscr{L} = all\text{-}IN$, whenever there are dialectically illegally *IN* arguments, there is at least one dialectically super-illegally *IN* argument;

$(Q3)$ In every complete $d\mathscr{L}$, if X is an argument dialectically legally *IN*, so are all of its elementary arguments[ii].

$(Q4)$ In every complete $d\mathscr{L}$, if X is an argument dialectically legally *IN*, so are all of its logically equivalent arguments X' if $\text{Prem}(X') \subset \text{Prem}(X)$[iii].

$(Q5)$ Since $\mathscr{S} = \mathsf{in}(d\mathscr{L})$, dialectical labellings and dialectical defeats influence each other.

Properties $Q1$, $Q1'$, $Q2$, $Q3$ and $Q4$ can be formally described and proved as propositions. $Q5$ can be easily illustrated via example since it directly follows from the definitions of dialectical defeats and dialectical labellings.

[i] It is actually easy to see how a maximal (wrt set inclusion) dialectical admissible labelling $d\mathscr{L}_A$ is equal to a maximal (wrt set inclusion) dialectical complete labelling $d\mathscr{L}_C$. The proof is the following. [$\rightarrow$]: if $d\mathscr{L}_A \neq d\mathscr{L}_C$, then $\exists X \notin \mathsf{in}(d\mathscr{L}_A)$ which is acceptable wrt $\mathsf{in}(d\mathscr{L}_A)$. However, this implies that $\mathsf{in}(d\mathscr{L}_A)$ is not maximal (contradiction). [$\leftarrow$] Trivial. A dialectical complete labelling is also a dialectical admissible labelling.

[ii] Let $X = (\Delta, \Gamma, \varphi)$. $\forall \alpha \in \Delta$, elementary arguments are arguments of the kind $X' = (\{\alpha\}, \emptyset, \alpha)$.

[iii] Recall that logically equivalent arguments, say Y and Y', are arguments that differ only for the epistemic distribution of their premises and suppositions. That is to say: $\text{Prem}(Y) \cup \text{Supp}(Y) = \text{Prem}(Y') \cup \text{Supp}(Y')$.

Property Q1 $Q1$ claims that there can never exist a dialectical *all-IN* labelling (i.e., a dialectical labelling where all the arguments of the considered pdAF are *IN*) without a dialectically legally *IN* argument. Which can be formally defined as:

Proposition 5. *Let $d\mathscr{L} = all\text{-}IN$ be a dialectical labelling for the pdAF = $\langle AR, defeats\rangle$. Then $\exists X \in AR$, such that its label is dialectically legally* IN.

Proof. There are two cases we need to consider.

(*a*) For any conflict existing between two arguments $X = (\Delta, \Gamma, \overline{\alpha}) \in AR$ and $Y = (\Pi, \Sigma, \beta) \in AR$ such that $X \Rightarrow_{\mathscr{S}} Y$ on $(\{\alpha\}, \emptyset, \alpha)$, then:

(1) by P1, $\exists A = (\{\alpha\}, \emptyset, \alpha) \in AR$;

(2) by P2, $\exists X' = (\Delta \cup \Gamma, \emptyset, \overline{\alpha}) \in AR$;

(3) by P3, $\exists Z = (\Delta \cup \Gamma \cup \{\alpha\}, \emptyset, \curlywedge) \in AR$;

(4) by P2, $\exists Z' = (\emptyset, \Delta \cup \Gamma \cup \{\alpha\}, \curlywedge) \in AR$.

Therefore, there is always an unassailable argument $Z' \in AR$ which, being undefeated, is dialectically legally *IN*[iv].

(*b*) If there are no conflicting arguments in the pdAF, then, trivially, there is at least one dialectically legally *IN* argument.

□

Property Q1' It follows as a corollary from Property Q1 and Proposition 5.

Property Q2 $Q2$ claims that there can never exist a dialectical *all-IN* labelling without dialectically super-illegally *IN* arguments if there are conflicting arguments in the pdAF. This can be formally defined as:

Proposition 6. *Let $d\mathscr{L} = all\text{-}IN$ be a dialectical labelling for the pdAF = $\langle AR, defeats\rangle$. If $\exists X \in AR$, such that its label is illegally* IN *then $\exists Y \in AR$ such that its label is super-illegally* IN[v].

[iv] Alternatively, by P2, $\exists A' = (\emptyset, \{\alpha\}, \alpha) \in AR$ which suffices to achieve the desired result. Notice also that if $\Gamma = \emptyset$, then $X = (\Delta, \emptyset, \overline{\alpha})$ and we will obtain the same results since: (3′) by P3, $\exists Z = (\Delta \cup \{\alpha\}, \emptyset, \curlywedge) \in AR$ and (4′) by P2, $\exists Z' = (\emptyset, \Delta \cup \{\alpha\}, \curlywedge) \in AR$.

[v] Notice that X and Y could be the same argument.

Proof. Property $Q2$ is entailed by property $Q1$. That is, according to property $Q1$ there always exists at least one dialectically legally *IN* argument in every $d\mathscr{L} = all\text{-}IN$. Then, if there are dialectically illegally *IN* arguments (i.e., conflicting arguments) we are in the same situation described by point (a) of Proof 4.1.1: such conflicting arguments will be defeated by at least an unassailable argument (due to P3) which, being unassailable, is also legally *IN*. Therefore, these illegally *IN* arguments must be super-illegally *IN*. □

Property Q3 $Q3$ claims that, in every complete $d\mathscr{L}$, if X is an argument dialectically legally *IN*, so are all of its elementary arguments.

Proposition 7. *Let $d\mathscr{L}$ be a dialectical complete labelling for the pdAF = $\langle AR, defeats\rangle$. If $\exists X \in AR$ such that its label is dialectically legally* IN*, then $\forall \alpha \in Prem(X)$, argument $(\{\alpha\}, \emptyset, \alpha)$ is dialectically legally* IN *too.*

Proof. Let $Y = (\{\alpha\}, \emptyset, \alpha)$ be an arbitrary elementary argument of the argument X. If X is dialectically legally *IN*, then all the defeats (if any) moved against each one of its premises (recall that dialectical defeats are licensed only on the premises of the dialectical arguments) are made by arguments dialectically *OUT*. Since $\alpha \in$ Prem(X), then α is defeated only by arguments labelled *OUT*, which makes *Y* dialectically legally *IN*. □

Property Q4 $Q4$ claims that, in every complete $d\mathscr{L}$, if X is an argument dialectically legally *IN*, so are all of its logically equivalent arguments X' if Prem$(X') \subset$ Prem(X).

Proposition 8. *Let $d\mathscr{L}$ be a dialectical complete labelling for the pdAF = $\langle AR, defeats\rangle$. If $\exists X \in AR$ such that its label is dialectically legally* IN*, then $\forall X' \in [X]$ such that $Prem(X') \subset Prem(X)$, X' is dialectically legally* IN *too*[vi].

Proof. Similar to Proof 4.1.1. Indeed, if $\forall \alpha \in$ Prem(X), α is such that it is defeated only by arguments dialectically labelled *OUT*, then *X* is dialectically legally *IN*. Therefore, the same must hold also for each one of its logically equivalent arguments X' if Prem$(X') \subset$ Prem(X). □

Property Q5 $Q5$ claims that dialectical labellings and dialectical defeats affect each other. This means that dialectical labels might change their legality at each variation occurring in the set Prem$(\mathscr{S})$ caused by some dialectical defeats. On the other hand, dialectical defeats between two arguments might be 'disqualified' at each variation occurring in the set Prem$(\mathscr{S})$ caused by the change of some dialectical labels. Since this

[vi]Recall that $[X]$ identifies the set of all the logically equivalent arguments of X [49].

mutual influence is a direct consequence of Definition 42 and Definition 43, we are going to show property $Q5$ via an example.

Example 10. *The purpose of this example is to show that, if the label of an argument changes, this can influence the dialectical defeats related to that argument which, in turn, can cause other changes in the labelling. Let us now consider Figure 4.1: although a hypothetical pdAF would be composed of more arguments, for simplicity, let us just focus on the two arguments shown.*

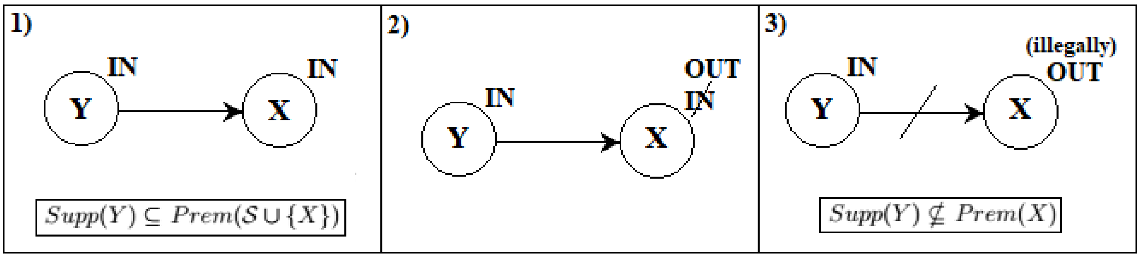

Figure 4.1: The three frames depict only two arguments of the labelling $d\mathscr{L}$ for a pdAF = $\langle \text{AR}, \textit{defeats} \rangle$ and what can happen if the label of one of them changes (2). This can indeed influence the existing dialectical defeats, which, in turn, can generate other changes in $d\mathscr{L}$ (3).

Frame 1) *display only two dialectical arguments, X and Y, of a dialectical labelling $d\mathscr{L}$ for a pdAF = $\langle AR, defeats \rangle$. These arguments are such that Y dialectically defeats X $($assume that $Supp(Y) \neq \emptyset$ and $Conc(Y) \neq \curlywedge)$ and $\{X, Y\} \subset \mathscr{S}$. According to Definition 42, since X is dialectically labelled* IN*, the dialectical defeat will target the set $\mathscr{S} = \mathsf{in}(d\mathscr{L})$, such that $Y \Rightarrow_{\mathscr{S}} X$ and $Supp(Y) \subseteq Prem(\mathscr{S} \cup \{X\})$. However, let us now assume that the label of argument X is changed to dialectically* OUT[vii] *(as shown in frame* 2*). In this circumstance, the dialectical defeat moved by Y cannot target the set $\mathscr{S}$ anymore. Then, if $Supp(Y) \not\subseteq Prem(X)$, the dialectical defeat that was in place before (thanks to the possibility of targeting the premises of the arguments in $\mathscr{S}$) will now be 'disqualified' due to the absence of the formulae that Y need to suppose to draw its conclusion. The situation will then be the one displayed in frame* 3*). Since argument Y is no more capable of dialectically defeating argument X, and since there are no other arguments in the pdAF that defeats X, the label of X will change to dialectically illegally* OUT*. This change can have important implications during the evaluation of an argumentation framework: for example, the presence of an argument legally or illegally* OUT *is a difference that can make a labelling admissible or not.*

[vii]The changing of a label is a very common step in many labelling-based algorithms.

4.2 Algorithm procedures for generating dialectical labellings

To compute dialectical labellings, we have to outline procedures capable of generating such labellings according to the semantics we want to capture. These procedural aspects will be handled by specifically designed algorithms. Intuitively, the general idea behind every algorithm is to start with a given dialectical labelling (that will mostly be empty or composed of arguments, all of which are dialectically labelled *IN*). Then, at each step of the procedure, add labels to unlabelled arguments or modify the existing labels. Recall also that dialectical labellings can influence dialectical defeats and vice versa (Example 10). The final result of these algorithms will be a dialectical labelling for the pdAF which complies with the desired semantics and takes into account the dialectical defeats.

Computational procedures for the generation of dialectical admissible, preferred and ground labellings are inspired by the algorithms already existing in the literature (the main reference will be to [97]). However, some of these procedures will also introduce original strategies which take advantage of the Dialectical Cl-Arg main features and properties, especially $Q3$ and $Q4$.

4.2.1 Preliminary notions for algorithms generating admissible/preferred dialectical labellings

Let us now focus on developing an algorithm that will generate dialectical admissible/preferred labellings. This will be possible by adapting the procedure introduced in [25] and [97] such that it will accommodate dialectical labellings instead. The idea is to start with the *all-IN* dialectical labelling and then to perform a *dialectical transition step* that will change an argument dialectically illegally *IN* to dialectically *OUT* (which changes to dialectically *UNDEC* if it has become dialectically illegally *OUT*). This procedure will then be iterated (generating in this way a *dialectical transition sequence*) until there will be no more dialectically illegally *IN* arguments. It can be shown that the result constitutes a dialectical admissible (or preferred) labelling.

Definition 46 (Dialectical transition step)**.** *Let $d\mathscr{L}$ be a dialectical labelling for the finite pdAF = $\langle AR, defeats\rangle$ and X an argument that is illegally* IN *in $d\mathscr{L}$ (according to Definition 43). A* dialectical transition step *on X in $d\mathscr{L}$ consists of the following:*

(1) *The label of X is changed to* OUT*;*

(2) *For every* $Y \in \{X\} \cup \{Z \mid X \Rightarrow_{\{Z\}} Z\}$ *if* Y *is dialectically illegally* OUT, *then its label is changed from* OUT *to* UNDEC.

That is to say, a dialectical transition step can be considered as a function that takes the tuple $(d\mathscr{L}, X)$ *as input and applies* (1) *and* (2) *to return a dialectical labelling* $d\mathscr{L}'$ *as output. Formally:* $\mathsf{d\_step}(d\mathscr{L}, X) = d\mathscr{L}'$.

Notice that (2) means that any argument made dialectically illegally *OUT* by (1) is turned to *UNDEC*.

Definition 47 (Dialectical transition sequence)**.** *A* dialectical transition sequence *is a list* $[d\mathscr{L}_0, X_1, d\mathscr{L}_1, \ldots, X_n, d\mathscr{L}_n]$ *(with* $n \geq 0$*), where for* $i = 1, \ldots, n$*,* X_i *is dialectical illegally* IN *in* $d\mathscr{L}_{i-1}$*, and* $d\mathscr{L}_i = \mathsf{d\_step}(d\mathscr{L}_{i-1}, X_i)$*.*

A dialectical transition sequence is said to be *terminated* iff $d\mathscr{L}_n$ does not contain any more dialectically illegally *IN* arguments.

Proposition 9. *Let* $[d\mathscr{L}_0, X_1, d\mathscr{L}_1, \ldots, X_n, d\mathscr{L}_n]$ *(with* $n \geq 0$*) be a terminated dialectical transition sequence where* $d\mathscr{L}_0$ *is the* all-IN *dialectical labelling. Then the resulting* $d\mathscr{L}_n$ *is a dialectical admissible labelling.*

Proof. The proof is a straightforward adaptation to the dialectical labellings approach of the proof in [25]. □

Notice that, at the end of the dialectical transition sequence, there could be arguments dialectically illegally *UNDEC*. Nevertheless, this will not prevent the generation of a dialectical admissible labelling since we just need to have arguments dialectically legally *IN* and *OUT* (Definition 45). However, if our purpose had been the generation of a complete labelling, then we would have required also dialectically legally *UNDEC* arguments (if any). [25] and [97] show how to help avoid non-complete labelling by modifying the transition step such that the super-illegally *IN* arguments are prioritized over the illegally *IN*[viii].

Furthermore, [25] showed that for every preferred labelling of an AF, there exists a terminated transition sequence that yields it.

Proposition 10. *Let* $d\mathscr{L}$ *be a dialectical preferred labelling. There exists a dialectical transition sequence* $[d\mathscr{L}_0, X_1, d\mathscr{L}_1, \ldots, X_n, d\mathscr{L}_n]$ *(with* $n \geq 0$*) where* $d\mathscr{L}_0$ *is the* all-IN *dialectical labelling and* $d\mathscr{L}_n = d\mathscr{L}$*.*

[viii] See Definition 44.

Proof. Once again, the proof is a straightforward adaptation to the dialectical labellings approach of the proof in [25]. □

Now, we are going to see how a dialectical transition sequence works. The following example makes use of the arguments of Figure 2.4. Observe that the pdAF should contain more logically equivalent arguments (by property P2) which, instead, have been disregarded in order to provide a shorter step-by-step example. Notice also that this choice is only to provide an intuitive understanding of the approach and should not be intended as attesting against the possibility of scaling up the procedure by inputting larger pdAFs. A more in-depth general analysis of the (best and worst cases) computational complexity of such a procedure will be presented in Section 4.4.

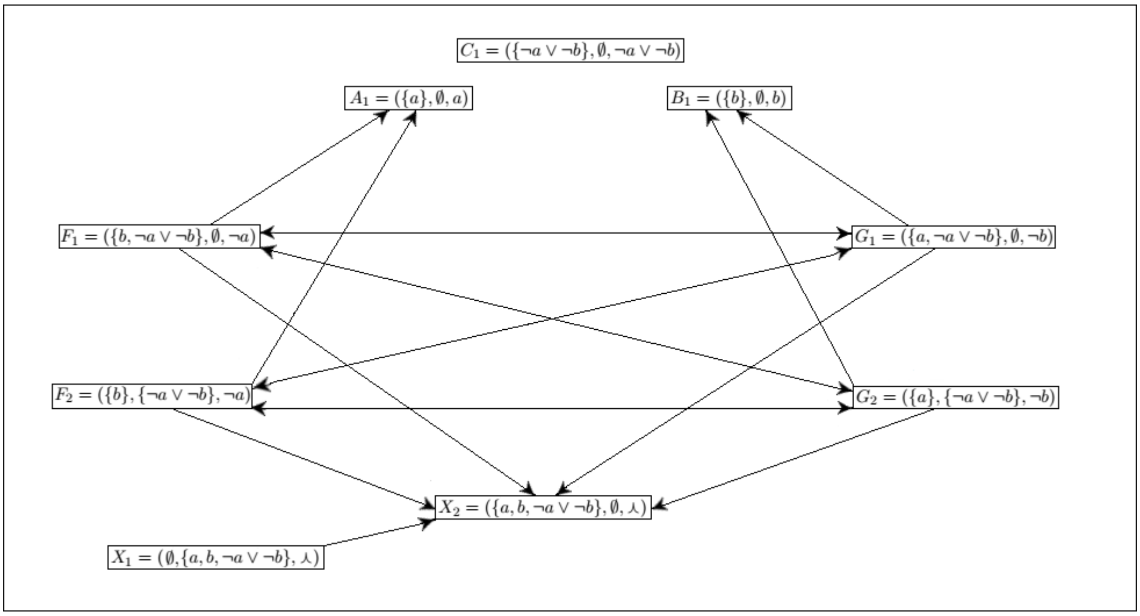

Figure 4.2: Example of an (incomplete) pdAF based on the dialectical arguments of Figure 2.4. The arrows identify the defeats existing among the arguments of the pdAF.

Example 11. *Let us take the framework displayed in Figure 4.2 and start a dialectical transition sequence on it. First of all, let us suppose that the procedure begins with the* all-IN *initial dialectical labelling* $d\mathscr{L}_0$*, that is to say:* $d\mathscr{L}_0 = (\{A_1, B_1, C_1, F_1, F_2, G_1, G_2, X_1, X_2\}, \emptyset, \emptyset)$[ix]*. Now we have to select an argument dialectically illegally* IN *and turn its label to* OUT*. Notice that the dialectical defeats moved by the arguments* F_2 *and* G_2 *are not 'disqualified'* (*as explained by property* $Q5$) *since the required formula* $(\neg a \vee \neg b) \in$ *Prem*$(\mathscr{S})$, (*i.e., Prem*$(\mathsf{in}(d\mathscr{L}))$. *We are going to perform the dialectical transition step*

[ix]From here onwards, a dialectical labelling $d\mathscr{L}$ will be represented as a tuple $(\mathsf{in}(d\mathscr{L}),\ \mathsf{out}(d\mathscr{L}),\ \mathsf{undec}(d\mathscr{L}))$.

on the dialectically illegally IN *argument* B_1*, changing its label to* OUT*. No argument has been made dialectically illegally* OUT *after this step, hence:* $d\mathscr{L}_1 = (\{A_1, C_1, F_1, F_2, G_1, G_2, X_1, X_2\}, \{B_1\}, \emptyset)$*. Next, we are going to perform the dialectical transition step on* X_2 *and* F_1*, with the following outcome:* $d\mathscr{L}_2 = (\{A_1, C_1, F_1, F_2, G_1, G_2, X_1\}, \{B_1, X_2\}, \emptyset)$ *and* $d\mathscr{L}_3 = (\{A_1, C_1, F_2, G_1, G_2, X_1\}, \{B_1, F_1, X_2\}, \emptyset)$*. We still have four arguments in the pdAF which are dialectically illegally* IN*:* A_1*,* F_2*,* G_1*,* G_2*. Let us choose argument* G_2 *and perform the dialectical transition step on it. Not only the label of* G_2 *will be changed to* OUT*, but it will also change to dialectically illegally* OUT*. This is because the dialectical defeat moved by* F_2 *will now be disqualified since* F_2 *cannot suppose the required formula* $(\neg a \vee \neg b)$ *from* $Prem(G_2)$*. Also, no other argument dialectically labelled* IN *defeats* G_2*. This means that argument* G_2*, according to the dialectical transition step, will change its dialectical label to* UNDEC*, hence:* $d\mathscr{L}_4 = (\{A_1, C_1, F_2, G_1, X_1\}, \{B_1, F_1, X_2\}, \{G_2\})$*. The performance of the dialectical transition step on* F_2 *will terminate the transition sequence procedure since there will be no more dialectically illegally* IN *arguments. This will generate the dialectical admissible labelling:* $d\mathscr{L}_A = (\{A_1, C_1, G_1, X_1\}, \{B_1, F_1, F_2, X_2\}, \{G_2\})$*. Notice that* $d\mathscr{L}_A$ *is not a dialectical complete labelling since argument* G_2 *is dialectically illegally* UNDEC*.*

4.3 Developing a sound and complete algorithm for dialectical preferred labellings

Since every admissible extension is a subset of a preferred extension (Dung's Fundamental Lemma [53], adapted for Dialectical Cl-Arg in [97]), by generating a maximal (wrt set inclusion) admissible extension, we will automatically hold also a preferred extension. Having shown the equivalence existing between admissible/preferred dialectical labellings and extensions, the strategy employed by the algorithm for the generation of the dialectical preferred labellings will involve, once again, the dialectical transition sequence. Indeed, Proposition 10 has already shown that there exist terminated dialectical transition sequences which hold preferred labellings as results.

The purpose of the whole procedure is to generate terminated dialectical transition sequences which will produce, among others, dialectical preferred labellings. In a nutshell, the idea is to reduce the number of dialectical transition sequences that need to be investigated in order to get dialectical preferred labellings. As a first step, we will adapt the naive preferred algorithm, presented in [97], to dialectical labellings. Afterwards, we will design a procedure that optimizes the enumeration of dialectical preferred labellings by employing specific Dialectical Cl-Arg properties.

4.3.1 The naive algorithm

Intuitively, we can summarise Algorithm 1 as follows. It takes as input the dialectical labelling $d\mathscr{L} = \textit{all-IN}$ for the pdAF = $\langle \mathrm{AR}, \textit{defeats} \rangle$ and initializes P as the set containing all of the candidate dialectical preferred labellings. Then, it proceeds according to the following stages:

(I) It checks if the considered dialectical labelling has fewer *IN* labelled arguments than other candidate dialectical labellings previously investigated and stored in P. If that is the case, the considered labelling cannot be maximally admissible. As such, the procedure backtracks and looks for new possibilities to examine (lines 9-10).

(II) It checks if the result of a terminated dialectical transition sequence has a higher number of *IN* labelled arguments than other candidate dialectical labellings previously investigated. If that is the case, it modifies P accordingly (lines 11-18).

(III) A dialectical transition step will be performed on a super-illegally *IN* argument (if any) of the considered dialectical labelling. On the resulting new dialectical labelling, function `find_preferred` will be recursively called (lines 20-22).

(IV) Dialectical transition steps will be performed on every remaining illegally *IN* argument of the considered dialectical labelling. On the new dialectical labellings resulting from each such transition step, function `find_preferred` will be recursively called (lines 24-27).

Trivially, if the pdAF = $\langle \mathrm{AR}, \emptyset \rangle$, then Algorithm 1 will immediately output $P = \{d\mathscr{L}\}$.

Notice that performing dialectical transition steps on super-illegally *IN* arguments is a way to increase the efficiency of the algorithm. Indeed, every dialectical super-illegally *IN* argument always changes its label to legally *OUT* at the end of the procedure (the proof is very similar to the one in [25] (Lemma 6)). As such, focussing preemptively on super-illegally *IN* argument (as targets for the performance of dialectical transition steps) will help to prune the search space of the algorithm. Therefore, it will be possible to reach the termination of a transition sequence sooner and without changing its result. From this trivially follows:

Corollary 1. *Let* $[d\mathscr{L}_0, X_1, d\mathscr{L}_1, \ldots, X_n, d\mathscr{L}_n]$ *(with* $n \geq 0$*) be a dialectical transition sequence such that* $d\mathscr{L}_n$ *is a preferred labelling. Performing dialectical transition steps on super-illegally* IN *arguments will still result in the dialectical preferred labelling* $d\mathscr{L}_n$.

Algorithm 1 Enumeration of Dialectical Preferred Labellings

Input: *the dialectical labelling* $d\mathscr{L}$ = *all-IN for the pdAF* = $\langle \mathrm{AR}, \mathit{defeats} \rangle$
Output: *a set* $P = \{d\mathscr{L} \mid d\mathscr{L}$ *is a dialectical preferred labelling for the pdAF*$\}$

```
P := ∅
dL_i := dL
find_preferred(dL_i)
print P
end
·
·
Function find_preferred(dL_i)
if ∃dL'_i ∈ P such that in(dL_i) ⊂ in(dL'_i) then
    return # backtrack and try next possibility
else if dL_i has no dialectically illegally IN argument then
    for each dL'_i ∈ P
        if in(dL'_i) ⊂ in(dL_i) then
            P := P \ {dL'_i}
        endif
    end for each
    P := P ∪ {dL_i}
    return # backtrack and try next possibility
else
    if dL_i has a dialectically super-illegally IN argument X then
        dL_{i+1} = d_step(dL_i, X)
        find_preferred(dL_{i+1})
    else
        for each X which is dialectically illegally IN
            dL_{i+1} = d_step(dL_i, X)
            find_preferred(dL_{i+1})
        end for each
    endif
endif
end Function
```

Theorem 13 (Soundness and completeness of algorithm 1)**.** *Consider a finite pdAF* = $\langle AR, \mathit{defeats} \rangle$. *Then Algorithm 1 returns dialectical preferred labellings iff there exist dialectical preferred labellings in the pdAF.*

Proof. To generate dialectical preferred labellings, the algorithm employs terminated dialectical transition sequences while it filters out every dialectical admissible labelling which is not maximal (hence, not preferred). Notice that preemptively choosing super-illegally *IN* arguments to perform dialectical transition steps will not affect the transition sequences, as ensured by Corollary 1. As such, the soundness of the procedure is guaranteed by Proposition 9 and steps (I) and (II) of the algorithm, while its completeness is ensured by Proposition 10. □

4.3.2 The optimized algorithm

The optimized version of Algorithm 1 will exploit properties $Q3$ and $Q4$ of the dialectical labelling approach to improve the efficiency of the procedure. Their employment will help in reducing the number of dialectical transition steps needed to terminate some of the dialectical transition sequences. Also, at the same time, they will provide a higher probability of generating dialectical complete labellings[x].

The idea is to prune the search space of the algorithm by preventing further examinations of dialectical transition sequences where properties $Q3$ or $Q4$ have been violated. Indeed, if the elementary argument of a dialectically legally *IN* argument X is instead not (or illegally) *IN*, we cannot expect the resulting dialectical labelling $d\mathscr{L}$ to be complete, hence preferred. Similarly, if $X' \in [X]$ is the logically equivalent argument of the dialectically legally *IN* argument X, such that $\text{Prem}(X') \subset \text{Prem}(X)$, then X' must be *IN* as well. Otherwise, the resulting $d\mathscr{L}$ will not be complete, hence not preferred.

Intuitively, we can summarise Algorithm 2 as follows. It takes as input the dialectical labelling $d\mathscr{L} = \textit{all-IN}$ for the pdAF = $\langle \text{AR}, \textit{defeats} \rangle$ and initializes P as the set containing all of the candidate dialectical preferred labellings. Then, it proceeds according to the following stages:

(I) This stage (lines 9-10) behaves identically to the one introduced for Algorithm 1 (i.e., the naive algorithm).

(a) It checks if there is a dialectically legally *IN* argument X such that at least one of its elementary arguments X' is not legally *IN*. If that is the case, the resulting dialectical labelling is violating property $Q3$, meaning that it cannot be a complete (nor preferred) labelling. As such, the procedure backtracks and looks for new possibilities to examine (lines 11-12).

[x] Recall that dialectical preferred labellings can be defined as maximal dialectical admissible labellings as well as maximal dialectical complete labellings (Definition 45).

(b) It checks if there is a dialectically legally *IN* argument X such that (at least one of) its logically equivalent argument X', where $\text{Prem}(X') \subset \text{Prem}(X)$, is not legally *IN*. If that is the case, the resulting dialectical labelling is violating property $Q4$, meaning that it cannot be a complete (nor preferred) labelling. As such, the procedure backtracks and looks for new possibilities to examine (lines 13-14).

(II, III, IV) These stages (lines 15-22, 24-26, 28-31) behave identically to the ones introduced for Algorithm 1 (i.e., the naive algorithm).

It may be possible that an argument dialectically legally IN will turn illegally *IN* (hence, *OUT* or *UNDEC* after additional dialectical transition steps) once the dialectical transition sequence has moved forward. Performing the pruning outlined by (a) and (b) could then possibly prevent the generation of a dialectical labelling which could have ended up being preferred. Notice, however, that the `optimized_find_preferred` function performs `d_step` on each illegally *IN* argument at every one of its recursive calls, meaning that a situation where both arguments X and X' (of either scenario (a) or (b)) are dialectically legally *OUT* (or *UNDEC*) will also be contemplated. That is to say, among the various branches of the algorithm search space, a situation where argument X will turn from dialectically illegally *IN* to *OUT* or *UNDEC* (without transiting from a dialectically legally *IN* status) will be considered as well, along with any resulting dialectical labelling. Therefore, (a) and (b) prove to be efficient improvements to the procedure without affecting its completeness property.

Algorithm 2 Optimized Enumeration of Dialectical Preferred Labellings

Input: *the dialectical labelling* $d\mathscr{L}$ = *all-IN for the pdAF* = $\langle \mathrm{AR}, defeats \rangle$

Output: *a set* $P = \{d\mathscr{L} \mid d\mathscr{L}$ *is a dialectical preferred labelling for the pdAF*$\}$

$P := \emptyset$

$d\mathscr{L}_i := d\mathscr{L}$

`optimized_find_preferred`$(d\mathscr{L}_i)$

print P

end

·

·

Function `optimized_find_preferred`$(d\mathscr{L}_i)$

if $\exists d\mathscr{L}'_i \in P$ *such that* $\mathsf{in}(d\mathscr{L}_i) \subset \mathsf{in}(d\mathscr{L}'_i)$ **then**

return # backtrack and try next possibility

else if $\exists X$ *legally IN*, *but one of its elementary argument is not* **then**

return # backtrack and try next possibility

else if $\exists X$ *legally IN*, *but* $X' \in [X]$ $(Prem(X') \subset Prem(X))$ *is not* **then**

return # backtrack and try next possibility

else if $d\mathscr{L}_i$ *has no dialectically illegally* IN *argument* **then**

for each $d\mathscr{L}'_i \in P$

if $\mathsf{in}(d\mathscr{L}'_i) \subset \mathsf{in}(d\mathscr{L}_i)$ **then**

$P := P \setminus \{d\mathscr{L}'_i\}$

endif

end for each

$P := P \cup \{d\mathscr{L}_i\}$

return # backtrack and try next possibility

else

if $d\mathscr{L}_i$ *has a dialectically super-illegally* IN argument X **then**

$d\mathscr{L}_{i+1} =$ `d_step`$(d\mathscr{L}_i, X)$

`optimized_find_preferred`$(d\mathscr{L}_{i+1})$

else

for each X *which is dialectically illegally* IN

$d\mathscr{L}_{i+1} =$ `d_step`$(d\mathscr{L}_i, X)$

`optimized_find_preferred`$(d\mathscr{L}_{i+1})$

end for each

endif

endif

end Function

Theorem 14 (Soundness and completeness of algorithm 2)**.** *Consider a finite pdAF =* $\langle AR, defeats \rangle$. *Then Algorithm 2 returns dialectical preferred labellings iff there exist dialectical preferred labellings in the pdAF.*

Proof. Consider that Corollary 1 holds also for Algorithm 2. The completeness property is then guaranteed by Proposition 10. Soundness, instead, will be ensured by Proposition 9, properties $Q3$, $Q4$ and steps (I) and (II) of the procedure. □

4.4 Algorithms analysis

The features that differentiate Algorithms 1 and 2 are the additional filtering steps that involve properties $Q3$ and $Q4$. As we have seen, the advantage of employing such features lies in the possibility of shortening the dialectical transition sequences in their process of searching for dialectical preferred labellings. This also contributes to enhancing the efficiency of the procedure. Indeed, the algorithm can save time (while still preserving the completeness property) by preventing the performance of dialectical transition steps on dialectical labelling that, with a high likelihood, will not become complete (nor preferred) labelling.

4.4.1 Time complexity

Let us now analyse how Algorithms 1 and 2 behave in their two extreme cases:

Best case In the best-case scenario, the pdAF = $\langle \text{AR}, \emptyset \rangle$ and no arguments defeat each other. In this circumstance, the time complexity of both Algorithms 1 and 2 is *constant*. Since there is no dialectical transition step to perform, the algorithm will immediately output the (only) preferred extension as a result of its procedure, i.e., $d\mathcal{L} = \textit{all-IN}$, therefore, $P = \{d\mathcal{L}\}$.

Worst case Let $d\mathcal{L} = \textit{all-IN}$ be a dialectical labelling for the pdAF = $\langle \text{AR}, \textit{defeats} \rangle$. In the worst-case scenario for Algorithm 1, the procedure will have to perform `d_steps` on every dialectically illegally *IN* argument (excluding super-illegally *IN* arguments, already handled by step (III)). The worst-case time complexity will then be bounded by operation (IV), i.e., the maximum number of times the function `find_preferred` can be recursively called), hence:

$$O((n-k)!)$$

where n is the number of arguments of the pdAF (i.e., $|\text{AR}| = n$) used as input in Algorithm 1 and k identifies the number of arguments dialectically legally *IN* of the input labelling $d\mathcal{L}$[xi]. Despite the fact that Algorithm 2 has a similar factorial upper bound,

[xi]More precisely, k denotes the number of dialectically legally *IN* arguments that cannot change their labels throughout the dialectical transition sequence, i.e., the undefeated arguments.

it is also capable of pruning its search space taking advantage of properties $Q3$ and $Q4$. This allows the shortening of some of the dialectical transition sequences performed by the procedure. The worst-case time complexity will then be equal to:

$$O((n-k-c)!)$$

where c represents the average number of prevented dialectical transition steps due to stages (a) and (b) of Algorithm 2. Let $k+c=k'$, we can then simplify the previous notation and write $O((n-k')!)$.

	Best Case	Worst Case
Algorithm 1	$\Omega(1)$	$O((n-k)!)$
Algorithm 2	$\Omega(1)$	$O((n-k')!)$

Figure 4.3: Table describing the time complexity of the developed algorithms. Notice that $k' > k$.

Remark 3. *The three-labels standard enumerating procedure for the preferred labellings (as presented in [41] and [97]) have a worst-case time complexity of $n!$. This is due to the fact that, in abstract argumentation, there can be AFs where every argument attacks/defeats each other. However, the same situation would never happen in Dialectical Cl-Arg since, whenever there are conflicting arguments, there are also unassailable arguments that cannot be defeated (as shown by properties Q1 and Q1′). This means that, in comparison with the three-labels standard enumerating procedure, preferred algorithms involving dialectical labellings perform more efficiently by default.*

4.4.2 Comparison of Algorithm 1 and Algorithm 2

Let us now compare the performances of the two developed algorithms. For simplicity, we are going to run the procedures on the incomplete pdAF depicted in Figure 4.2.

Example 12. *Recall that both algorithms take as input the dialectical labelling $d\mathscr{L}=$* all-IN *for the pdAF and begin by assuming $P=\emptyset$, where P is the set of dialectical preferred labellings. As such, the initial dialectical labelling is $d\mathscr{L}_0=(\{A_1,B_1,C_1,F_1,F_2,G_1,G_2,X_1,X_2\},\emptyset,\emptyset)$:*

1) *Following the procedure of Algorithm 1, we have to start by comparing the initial dialectical admissible labelling with the ones already members of P. Since P is empty and there are dialectically illegally* IN *arguments in the pdAF, we can skip stages (I)-(II) and move to stage (III) performing dialectical transition steps on dialectically super-illegally* IN *arguments, if any. X_2 is such an argument because*

it is defeated by the dialectically legally IN X_1. *Changing the label of* X_2 *to dialectically* OUT *will yield* $d\mathscr{L}_1 = (\{A_1, B_1, C_1, F_1, F_2, G_1, G_2, X_1\}, \{X_2\}, \emptyset)$. *Since there are no more dialectically super-illegally* IN *arguments, we can proceed with stage* (IV). *The subsequent dialectical transition steps will choose random arguments until the termination of a dialectical transition sequence. The outcome will be a dialectical admissible labelling that will update P after the checks of steps* (I)-(II). *Then, Algorithm 1 will will keep backtracking* (*to the other dialectical transition sequences generated by recursively calling function* `find_preferred`) *and revising set P until each possibility has been investigated.*

2) *Following the procedure of Algorithm 2, we have to check if, during a dialectical transition sequence, the resulting dialectical labelling violates properties* $Q3$ *and/or* $Q4$. *If so, we can backtrack and look for a different possibility. The remaining procedure is identical to Algorithm 1* (*stages* (I)-(IV)).

   To provide an example of the shortcuts granted by Algorithm 2, we will examine one of the dialectical transition sequences generated by the procedure. The algorithm begins by performing a dialectical transition step on the super-illegally IN *argument* X_2 (*this corresponds to stage* (III)), *transforming* $d\mathscr{L}_0$ *in* $d\mathscr{L}_1 = (\{A_1, B_1, C_1, F_1, F_2, G_1, G_2, X_1\}, \{X_2\}, \emptyset)$. *We should now move to stage* (IV) *and perform* `d_steps` *on the remaining dialectically illegally* IN *arguments. In the following, let us suppose that the resulting dialectical labelling always passes the check of stage* (I). *This will lead to the generation of* $d\mathscr{L}_2 = (\{A_1, C_1, F_1, F_2, G_1, G_2, X_1\}, \{B_1, X_2\}, \emptyset)$, $d\mathscr{L}_3 = (\{A_1, C_1, F_1, F_2, G_1, X_1\}, \{B_1, G_2, X_2\}, \emptyset)$ *and* $d\mathscr{L}_4 = (\{A_1, C_1, F_1, F_2, X_1\}, \{G_1, G_2, B_1, X_2\}, \emptyset)$ *by respectively choosing arguments* B_1, G_2, G_1 *as targets for the* `d_steps`. *Observe that* $d\mathscr{L}_4$ *violates* $Q3$ *since* F_1 (*alternatively,* F_2) *is dialectically labelled* IN, *whereas its elementary argument* B_1 *is labelled* OUT. *Although not terminated, the dialectical transition sequence will then be interrupted* (*stage* (a)) *and the procedure will backtrack. This outcome ensured a reduction of the considered dialectical transition sequence of one* `d_steps`, *proving to be a valid improvement on the efficiency of the algorithm.*

Example 12 highlights the efficiency advantages ensured by the shortcuts of Algorithm 2 over Algorithm 1. Indeed, we can see how the number of steps needed to reach the termination of a dialectical transition sequence might differ.

To the best of my knowledge, the three-labels enumerating procedure presented in

Algorithm 1 (especially its optimized version, Algorithm 2) constitutes a novelty. The literature mainly focuses on developing (or improving) algorithms for abstract AFs[xii], addressing the main reasoning problems (i.e., enumeration, verification, sceptical and credulous acceptance of arguments) for a variety of semantics. Also, many of such procedures employ more than three labels to identify the arguments of an AF: besides *IN*, *OUT* and *UNDEC* they add labels such as *BLANK*, *MUST_OUT* and *ATT*. Although some works have been undertaken in structured argumentation (for example, [73] for ABA and [123] for ASPIC$^+$), to my recollection, no original algorithm has been developed for Dialectical Cl-Arg thus far. The procedures presented in this chapter are the first attempt at computing and enumerating admissible and preferred extensions for Dialectical Cl-Arg and accomplish so by exploiting the properties of such a dialectical approach. For instance, the optimized dialectical preferred algorithm guides the dialectical transition sequences and enhances the procedure simply by making use of the properties $Q1$, $Q1'$, $Q3$ and $Q4$. Finally, notice that despite the factorial worst-time complexity of both Algorithm 1 and Algorithm 2 their performance is far more efficient than $n!$, which is the worst-case time complexity of the three-labels enumerating procedure for the preferred labellings as presented in [41] and [97].

4.5 Developing a sound and complete algorithm for dialectical grounded labellings

As in the standard case described in [97], the strategy for designing an algorithm capable of generating the dialectical grounded labelling will employ the emulation of the characteristic function $\mathscr{F}$. However, recall that for Dialectical Cl-Arg the monotonicity of the characteristic function has been proven only for epistemically maximal admissible sets of arguments [49]. Since the goal is to reach the least fixed point of $\mathscr{F}$, which yields the grounded extension according to its constructive definition, we require $\mathscr{F}$ to be monotonic and the pdAF to be finitary. By Definition 41, the considered pdAF is finite, hence finitary[xiii]. Instead, in order to achieve the monotonicity of the characteristic function, at every step of the iteration of $\mathscr{F}$, we will have to invoke another function that will change

[xii] As shown by the proceedings of the latest 'Systems and Algorithms for Formal Argumentation (SAFA) 2016 [121], 2018 [122], 2020 [64] and the 'International Competition on Computational Models of Argumentation' (ICCMA) 2015 [124], 2017 [63], 2019 [19], 2021 [81].

[xiii] Since the pdAF is finite, there is only a finite number of defeating arguments for each argument of the framework (which corresponds to the definition of finitary).

the current set of *IN* arguments into an epistemically maximal admissible set. Such function, called Cl_{em}, has already been introduced in [49] and now needs to be adapted to dialectical labellings.

Definition 48 (Epistemically maximal step). *Let consider a pdAF* = $\langle AR, defeats \rangle$. *The function* $Cl_{em} : 2^{AR} \mapsto 2^{AR}$ *maps any* $\mathsf{in}(d\mathscr{L})$ *sets to its epistemically maximal set such that:* $Cl_{em}(\mathsf{in}(d\mathscr{L})) = \mathsf{in}(d\mathscr{L}) \cup$

$$\{X' \mid X \in \mathsf{in}(d\mathscr{L}),\ X' \in [X],\ Prem(X) \subseteq Prem(X'),\ Prem(X') \subseteq Prem(\mathsf{in}(d\mathscr{L}))\};$$

We are going to write $\mathtt{em\_step}(\mathsf{in}(d\mathscr{L}))$ *to denote the performance of* Cl_{em} *on* $\mathsf{in}(d\mathscr{L})$*.*

In the following algorithm, the resulting $d\mathscr{L}_G$ will identify the dialectical grounded labelling. Recall also that $\mathscr{S} = \mathsf{in}(d\mathscr{L})$.

Algorithm 3 Dialectical Grounded Labelling Algorithm

Input: *the dialectical labelling* $d\mathscr{L} = (\emptyset, \emptyset, \emptyset)$ *for the pdAF* = $\langle \mathrm{AR}, defeats \rangle$
Output: *the dialectical grounded labelling* $d\mathscr{L}_G$

$d\mathscr{L}_i := d\mathscr{L}$
$\mathtt{find\_grounded}(d\mathscr{L}_i)$
print $d\mathscr{L}_G$
end
·
·
Function $\mathtt{find\_grounded}(d\mathscr{L}_i)$
repeat
    $\mathsf{in}(d\mathscr{L}_{i+1}) = \mathsf{in}(d\mathscr{L}_i) \cup \{X \mid X$ *is not dialectically labelled in* $d\mathscr{L}_i$
    and $\forall Y$: *if* $Y \Rightarrow_{\mathscr{S}} X$ *then* $Y \in \mathsf{out}(d\mathscr{L}_i)\ \}$
        if $\mathsf{in}(d\mathscr{L}_{i+1}) \neq Cl_{em}(\mathsf{in}(d\mathscr{L}_{i+1}))$
            then $\mathsf{in}(d\mathscr{L}_{i+1}) = \mathtt{em\_step}(\mathsf{in}(d\mathscr{L}_{i+1}))$
        endif
    $\mathsf{out}(d\mathscr{L}_{i+1}) = \mathsf{out}(d\mathscr{L}_i) \cup \{X \mid X$ *is not dialectically labelled in* $d\mathscr{L}_i$
    and $\exists Y : Y \Rightarrow_{\{X\}} X$ *and* $Y \in \mathsf{in}(d\mathscr{L}_{i+1})\ \}$
until $d\mathscr{L}_{i+1} = d\mathscr{L}_i$
$d\mathscr{L}_G = (\mathsf{in}(d\mathscr{L}_i), \mathsf{out}(d\mathscr{L}_i), AR - (\mathsf{in}(d\mathscr{L}_i) \cup \mathsf{out}(d\mathscr{L}_i)))$
end Function

Intuitively, the algorithm begins by (I) assigning *IN* to each dialectically undefeated argument of an empty dialectical labelling $d\mathscr{L} = (\emptyset, \emptyset, \emptyset)$ and (II), if needed, proceed further by epistemically maximally closing the set of arguments just made *IN*. The following step (III) assigns *OUT* to every argument dialectically defeated by the arguments turned *IN* by (I) and (II), hence (IV) assigns the dialectical label *IN* to those arguments all of whose defeaters are dialectically labelled *OUT*. This procedure will be iterated until (V) no more *IN*/*OUT* assignments can be done, while any remaining unlabelled arguments will be dialectically labelled *UNDEC*.

Finally, having successfully emulated the iteration of the characteristic function $\mathscr{F}$ and reached the least fixed point of it, we can conclude that $d\mathscr{L}_G$ (printed by the procedure) corresponds to the dialectical grounded labelling.

Proposition 11 (Soundness and completeness of the algorithm)**.** *Let consider a finite pdAF =* $\langle AR, defeats \rangle$*. Then Algorithm 3 returns the dialectical grounded labelling iff there exists the dialectical grounded labelling for the pdAF.*

Proof. The purpose of Algorithm 3 is to construct the dialectical grounded labelling (which corresponds to the dialectical grounded extension by Theorem 12) by emulating the iteration of $\mathscr{F}$, i.e., the characteristic function of the framework. As such, the proof of soundness and completeness are already given in [49], since they are identical to the same proofs for $\mathscr{F}$. □

As we have seen, the only difference between Algorithm 3 and the algorithm in [97] for generating the grounded labelling is the presence of the `em_step`. However, is this step really necessary? The following section will investigate such a question.

4.5.1 Epistemically preserved admissible set

Recall once again that in Dialectical Cl-Arg, the monotonicity of the framework characteristic function $\mathscr{F}$ is shown under the domain of epistemically maximal admissible sets [49]. Epistemic maximality allows preventing a situation where, given two admissible set $\mathscr{S}$, $\mathscr{S}'$ (such that $\mathscr{S}' \supset \mathscr{S}$), an argument X is acceptable with respect to $\mathscr{S}$, but not with respect to $\mathscr{S}'$ (i.e., X $\in \mathscr{F}(\mathscr{S})$, but $X \notin \mathscr{F}(\mathscr{S}')$). This situation can be better captured with the introduction of *epistemically preserved* (*ep*) sets. In this section, we are going to show that, without the need to employ the epistemically maximal closure Cl_{em}, we can achieve the same result by focusing on epistemically preserved admissible sets.

Definition 49 (Epistemically preserved sets)**.** *Let* $\langle AR, defeats \rangle$ *be a pdAF and let* E, E' *be two extensions of AR such that* $E' \supset E$*. Then E is an epistemically preserved* (ep) *set iff:*

- *If X is acceptable with respect to E, then* $\nexists Z \in AR$ *such that* $Z \Rightarrow_{E'} X$ *and X is not defended by an argument in E.*

Lemma 5. *Let E be an epistemically preserved and epistemically maximal admissible set and let $\mathscr{S} = \mathscr{F}(E)$ be the* ep *set whose members are arguments acceptable wrt E. Then $Cl_{em}(\mathscr{S}) = \mathscr{S}$.*

Proof. We are going to show that if $X = (\Delta, \Gamma, \alpha) \in \mathscr{S}$ and $\Gamma' \subseteq (\Gamma \cap \text{Prem}(\mathscr{S}))$, then $X' = (\Delta \cup \Gamma', \Gamma \setminus \Gamma', \alpha) \in \mathscr{S}$.

Assume the contrary: $X \in \mathscr{S}$, $\Gamma' \subseteq (\Gamma \cap \text{Prem}(\mathscr{S}))$, but $X' \notin \mathscr{S}$. This means that $\exists Z = (\Pi, \Sigma, \varphi) \in AR$ such that $Z \Rightarrow_E X'$ on β ($\varphi = \overline{\beta}$ or $\varphi = \curlywedge$) and $\nexists W \in E$ such that $W \Rightarrow_{\{Z\}} Z$. There are two cases to consider:

- Suppose $Z \Rightarrow_E X$ (if $\beta \in \Delta$) or $Z \Rightarrow_E Q$ (if $\beta \in \Gamma'$) where $\Gamma' \subseteq \text{Prem}(Q)$. However, we know that $X, Q \in \mathscr{S}$ and $\mathscr{S} = \mathscr{F}(E)$. As such, $\exists W \in E$ where $W \Rightarrow_{\{Z\}} Z$ contradicting the assumptions.
- Suppose:
  - $\beta \in \Delta$ but $Z \not\Rightarrow_E X$. This means that $Z \Rightarrow_{E \cup \Gamma'} X$. Since $\mathscr{S} \supset E \cup \Gamma'$, then $Z \Rightarrow_{\mathscr{S}} X$, contradicting the fact that E is *ep*.
  - $\beta \in \Gamma'$ but $Z \not\Rightarrow_E Q$, where $\Gamma' \subseteq \text{Prem}(Q)$. This means that $Z \Rightarrow_{E \cup \Delta} X$. Since $\mathscr{S} \supset E \cup \Delta$, then $Z \Rightarrow_{\mathscr{S}} Q$, contradicting the fact that E is *ep*.

Finally, the Fundamental Lemma (adapted for Dialectical Cl-Arg in [49]) and its implied properties (especially Lemma 21 of [49]) ensures that $Cl_{em}(\mathscr{S})$ is also admissible. □

As we have seen, the only difference between Algorithm 4 and the algorithm in [97] for generating the grounded labelling is the presence of the `em_step`. However, is this step really necessary? If we can prove that each iteration of the algorithm sub-routine outputs an *ep* admissible set of dialectically *IN* arguments (i.e., $\mathsf{in}(d\mathscr{L}_i)$ is *ep*), we can also ensure that such a set is *em* admissible by Lemma 5. This would allow us to eschew the `em_step`.

Lemma 6. *Each iteration of Algorithm 3 sub-routine generates a set of dialectically* IN *arguments which is epistemically preserved.*

Proof. Every iteration of Algorithm 3's sub-routine generates a set $\mathsf{in}(d\mathscr{L}_{i+1})$ which is composed of the set $\mathsf{in}(d\mathscr{L}_i)$, generated by the previous iteration, in addition to all the new arguments made dialectically legally *IN* (lines 9-10). By Definition 42, an argument X is dialectically legally *IN* iff all of its defeaters are dialectically *OUT*. This means that if an unlabelled argument Z dialectically defeats X, argument X cannot be considered legally *IN*, even if all of its other defeaters are *OUT*. As such, X will not be a member of $\mathsf{in}(d\mathscr{L}_{i+1})$. This corresponds with the idea of the epistemically preserved set: if X is acceptable wrt $\mathsf{in}(d\mathscr{L}_i)$ then it can be defended against any possible defeaters $Z \Rightarrow_E X$ (where $E = \mathsf{in}(d\mathscr{L}_{i+k})$, for $k \geq 0$). □

Lemma 5 and Lemma 6 entail the following Corollary:

Corollary 2. *The purpose of Algorithm 3 sub-routine is to emulate the iterations of the framework characteristic function $\mathscr{F}$. Each iteration of the sub-routine produces an* ep *admissible set which is also* em[xiv]. *This makes the* `em_step` *unnecessary.*

4.5.2 The optimized algorithm

Thanks to Corollary 2, we can optimize the procedure for the generation of the dialectical grounded labelling by eschewing the `em_step`. This means that the following algorithm can be used instead of Algorithm 3 for generating the dialectical grounded labelling. Notice that Algorithm 4 works almost identically to Algorithm 3 (it is indeed the same procedure without the `em_step`), hence, also the results of soundness and completeness will remain the same.

Algorithm 4 Optimized Dialectical Grounded Labelling Algorithm

Input: *the dialectical labelling* $d\mathscr{L} = (\emptyset, \emptyset, \emptyset)$ *for the pdAF* = $\langle \mathrm{AR}, defeats \rangle$
Output: *the dialectical grounded labelling* $d\mathscr{L}_G$

$d\mathscr{L}_i := d\mathscr{L}$
`find_grounded`$(d\mathscr{L}_i)$
print $d\mathscr{L}_G$
end
·
·
Function `find_grounded`$(d\mathscr{L}_i)$
repeat
    $\mathsf{in}(d\mathscr{L}_{i+1}) = \mathsf{in}(d\mathscr{L}_i) \cup \{X \mid X$ *is not dialectically labelled in* $d\mathscr{L}_i$
    and $\forall Y$: *if* $Y \Rightarrow_{\mathscr{S}} X$ *then* $Y \in \mathsf{out}(d\mathscr{L}_i)\ \}$
    $\mathsf{out}(d\mathscr{L}_{i+1}) = \mathsf{out}(d\mathscr{L}_i) \cup \{X \mid X$ *is not dialectically labelled in* $d\mathscr{L}_i$
    and $\exists Y : Y \Rightarrow_{\{X\}} X$ *and* $Y \in \mathsf{in}(d\mathscr{L}_{i+1})\ \}$
until $d\mathscr{L}_{i+1} = d\mathscr{L}_i$
$d\mathscr{L}_G = (\mathsf{in}(d\mathscr{L}_i), \mathsf{out}(d\mathscr{L}_i), AR - (\mathsf{in}(d\mathscr{L}_i) \cup \mathsf{out}(d\mathscr{L}_i)))$
end Function

[xiv] Notice that even the initial $\mathsf{in}(d\mathscr{L}_i) = \emptyset$ is, trivially, both *em* and *ep*.

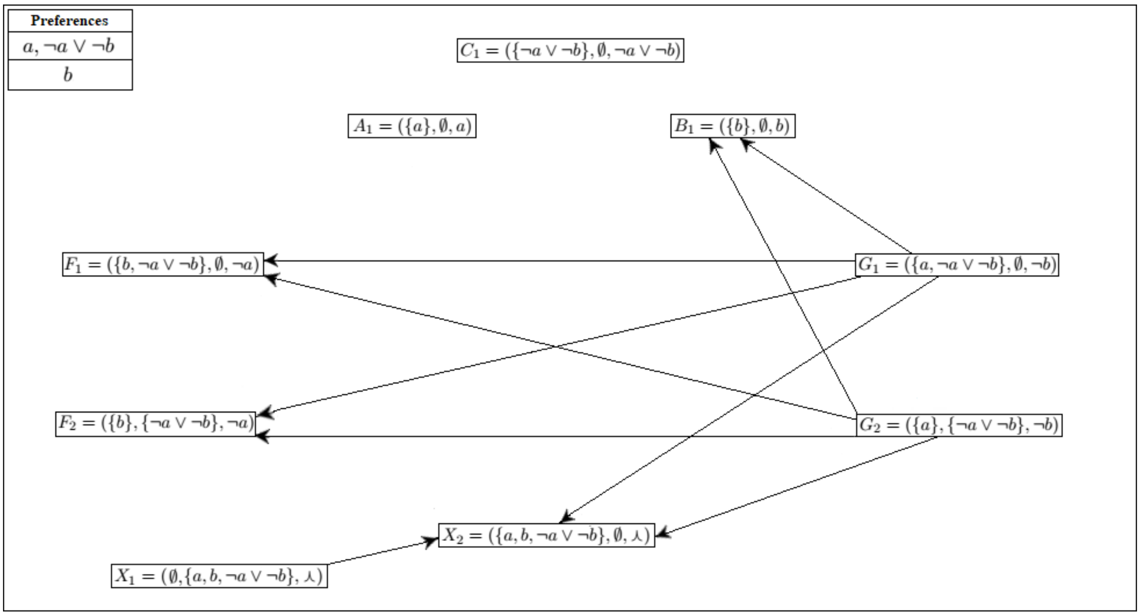

Figure 4.4: This pdAF is similar to the one in Figure 4.2 except for a different preference ordering among the arguments (highlighted by the box in the top-left corner) and the entailed defeats.

Example 13. *Let us now consider the framework displayed in Figure 4.4 and implement Algorithm 4 on it. The procedure will start with* $d\mathscr{L} = (\emptyset, \emptyset, \emptyset)$. *Now we continue by assigning the dialectical label* IN *to every undefeated argument of the pdAF, that is to say,* $\mathsf{in}(d\mathscr{L}_1) = \{A_1, C_1, G_1, G_2, X_1\}$. *Having labelled* G_1, G_2, X_1 *dialectically* IN *entails changing to dialectically* OUT *the arguments* B_1, F_1, F_2 *and* X_2. *Now there are no more arguments that can be labelled* IN *or* OUT, *as such, to proceed further, we have to update all the remaining unlabelled arguments to dialectically* UNDEC. *Since there are no such arguments, the algorithm will output the dialectical grounded labelling:* $d\mathscr{L}_G = (\{A_1, C_1, G_1, G_2, X_1\}, \{B_1, F_1, F_2, X_2\}, \emptyset)$.

4.6 Dialectical labellings & dialectical games

In the previous chapter, we have shown the correlation existing between the victory of a dialectical argument game for an argument A and the membership of A into the extension the game was meant to capture. In this chapter, instead, we have seen the equivalence existing between dialectical labellings and extensions (Theorem 12). As a result, the following propositions will hold[xv]:

[xv] We are going to assume familiarity with the notion of dialectical dispute tree $\mathscr{D}$ along with the other notations and definitions of the same chapter.

Proposition 12 (Dialectical preferred game & labelling)**.** *Let* Φ_P-$\mathscr{D}^n$ *identifies a terminated* Φ_P*-dialectical game for* A*. Then, there exists a dialectical winning strategy* Φ_P-$\mathscr{W}^n$ *for* A*, such that the set* $PRO(\Phi_P\text{-}\mathscr{W}^n)$ *of arguments moved by PRO in* Φ_P-$\mathscr{W}^n$ *is conflict-free, iff there exists a dialectical preferred labelling* $d\mathscr{L}_P$ *with* $d\mathscr{L}_P(A) = \text{IN}$[xvi].

Proposition 13 (Dialectical grounded game & labelling)**.** *Let* Φ_G-$\mathscr{D}^n$ *identifies a terminated* Φ_G*-dialectical game for* A*. Then, there exists a dialectical winning strategy* Φ_G-$\mathscr{W}^n$ *for* A*, such that the em closure* $Cl_{em}(PRO(\Phi_G\text{-}\mathscr{W}^n))$ *of the set of arguments moved by PRO in* Φ_G-$\mathscr{W}^n$ *is conflict-free, iff there exists a dialectical grounded labelling* $d\mathscr{L}_G$ *with* $d\mathscr{L}_G(A) = \text{IN}$.

4.7 Future Work

The way in which the chapter has been designed allows for two possible lines of future research. The first (i) will focus on extending the introduced dialectical labellings procedure to other semantics (for example, stable, semi-stable and ideal semantics). Whereas the second (ii) line of potential investigation would provide an optimization of the developed algorithms. Indeed, Algorithm 2 might be further improved if we move to a four-labels (as in [41]) or a five-labels [99] enumerating procedure. Concerning (i), notice that both Algorithms 1 and 2 can be adapted to enumerate all the dialectical stable or semi-stable labellings rather than just the preferred ones. For the stable labellings, we only need to introduce a check in the procedure that will exclude all the dialectical labellings composed by at least one dialectically *UNDEC* argument (this corresponds to the definition of stable labellings as given in [25]). For example, we could replace line 9 of code (whether we are dealing with Algorithm 1 or Algorithm 2) declaring:

if $\mathsf{undec}(d\mathscr{L}_i) \neq \emptyset$ **then return**

In addition, we can remove lines 12-16 (or lines 16-20 if we are dealing with Algorithm 2): there is no need to make any comparison about sets of *IN* arguments.

Whereas for the semi-stable labellings, we just need to introduce a check in the procedure that will exclude all the dialectical labellings which are not composed by a minimal amount of *UNDEC* arguments (this corresponds to the definition of semi-stable labellings as given in [25]). For example, we could replace line 9 of code (whether we are dealing with Algorithm 1 or Algorithm 2) declaring:

[xvi] Similarly for the admissible extension. Recall indeed that the dialectical admissible and preferred argument games are identical.

if $\exists d\mathscr{L}'_i \in P : \mathsf{undec}(d\mathscr{L}'_i) \subset \mathsf{undec}(d\mathscr{L}_i)$ **then return**

In addition, we can replace line 13 (or line 17 if we are dealing with Algorithm 2) declaring:

if $\mathsf{undec}(d\mathscr{L}_i) \subset \mathsf{undec}(d\mathscr{L}'_i)$ **then**

4.8 Conclusion

In this chapter, we have produced three main results: (1.) adapted labelling definitions and procedures to Dialectical Cl-Arg; devised algorithms for the generation of (2.) the dialectical preferred labellings, and (3.) the dialectical grounded labelling. Both these latter two achievements constitute novel contributions (to the best of our knowledge, no previous attempt on accommodating labellings and implementing them via algorithms had been pursued for Dialectical Cl-Arg). We discovered that dialectical labellings enjoy specific properties, which have then been used to improve the efficiency of the algorithms. For example, the proposed optimized dialectical preferred algorithm guides the procedure simply by employing the properties of Dialectical Cl-Arg. Finally, concerning the dialectical grounded labelling, another important result has been established: thanks to the notion of *epistemically preserved* admissible sets, we were able to prove that, against expectations, the procedure did not require an epistemically maximal closure at the end of each of its iteration.

Dialectical argument game proof theories and labellings appear to be the two faces of the same 'semantic' coin, thus providing different formal instruments to achieve the same result: the computation of dialectical semantics (Definition 23). We shall now proceed by trying to extend the single-agent reasoning process modelled by dialectical argument games into a more distributed form of interaction. Specifically, we are going to focus on explanation dialogues since we envisage them as being one of the kind of dialectical exchanges that occurs more often among real-world resource-bounded agents.

Chapter 5

Explanation-Question-Response (EQR) Dialogue and Scheme

Argument schemes (AS) represent structures of inferences that model presumptive reasoning in different circumstances. The AS Over Proposal for Action, for example, formalises the rational pattern for proposing an action, whereas the AS from Expert Opinion serves to model claims from experts. By merging these two schemes, however, it is possible to obtain a novel kind of template (hereafter EQR) capable of modelling the reasoning, and the entailed consequences, that follow from an expert opinion. This chapter presents the EQR scheme, an AS designed to be the core element of the Explanation-Question-Response dialogue, whose protocol is specialized in the assessment of explanations. This dialogue (and the related EQR scheme) offers three different variants, depending on the focus the protocol is meant to emphasise: assertion, endorsement or specific endorsement.

5.1 Argument Scheme from Expert Opinion

Argument schemes model stereotypical patterns of presumptive reasoning [138]. They exemplify forms of default reasoning where a claim is supported by default and its validity is assessed by specific critical questions. Since these days it is not unusual to rely on the opinions, or suggestions, of experts (humans or AI) [145], particular interest assumes the scheme representing an *Argument from Expert Opinion* [139]. Indeed, it might be argued that an individual will never be able to understand every aspect of our society, based on complex and ever-changing technology. As such, we all need to put (some degree of) our trust in the assertions made by authorities and experts [145].

Following the notation of [140], we are going to outline the general structure of an argument scheme from expert opinion. We can identify E as an expert in a specific field of knowledge F composed of a finite collection of propositions (among which there is A):

Argument Scheme from Expert Opinion
Premise : Source E is an expert in field F containing proposition A *Premise* : E asserts that proposition A (in field F) is true (false)
Conclusion : A may plausibily be taken to be true (false)

This reasoning pattern is always accompanied by a series of critical questions (**CQs**) whose purpose is to assess the validity of the conveyed argument. The following is the complete list of such questions:

CQ1. (*Expertise Question*) How knowledgeable is E as an expert source?

CQ2. (*Field Question*) Is E an expert in the field F that A is in?

CQ3. (*Opinion Question*) What did E assert that implies A?

CQ4. (*Trustworthiness Question*) Is E personally reliable as a source?

CQ5. (*Consistency Question*) Is A consistent with what other experts assert?

CQ6. (*Evidence Question*) Is E's assertion based on evidence?

In [90] another question is added to the previous six:

CQ7. (*Self-interest Question*) Is it the case that E does not stand to gain by our endorsement of proposition A?

Argument scheme from expert opinion and its related critical questions aim at modelling explanations from experts. Evaluation and assessment of explanations are particularly suited to be modelled as dialogical interactions between an *explainer*, i.e., an agent capable and willing to answer questions concerning the explanation and an *explainee*, i.e., an agent seeking to determine the validity of such answers. In [90], this dialogue is called Explanation-Question-Response (EQR) and its protocol can be seen as halfway between persuasion, information-giving/seeking and query dialogues, avoiding in this way the employment of a more complicated formalism (as the Control Layer) that would account for different simultaneous discussions taking place[i]. As such, the EQR dialogue proves to be an efficient way to capture multiple kinds of dialectical interactions that might occur when the topic revolves around the explanation of a statement. The purpose of this chapter is to develop three fully-fledged variants of the EQR dialogue protocol (considered as a new dialogue type that may be termed *'explanation'*) and the core elements (i.e., the EQR *schemes*) upon which they are based. Such protocols envisage different locutions, according to specific pre and post-conditions, capable of updating the commitment store associated with each agent participating in the dialogue.

To begin with, we can distinguish three different types of formalisations that will lead to the creation of three kinds of EQR dialogues (and corresponding EQR *schemes*) called: *'claim'*, *'endorsement'* and *'endorsed-by-whom'*.

1. The *'claim'* dialogue will consider how (acting upon) the expert's assertion will affect the current epistemic state of the world;
2. The *'endorsement'* dialogue will not only take into account the expert's assertion, but its endorsement too. It is now (acting upon) such endorsement that will affect the current epistemic state of the world;
3. The *'endorsed-by-whom'* dialogue is an *endorsement* dialogue that will also model and make explicit the agent(s) that is(are) *endorsing* the expert's assertion.

5.2 The *EQR claim* Dialogue

Before delving into the creation of the actual protocol, let us begin by formalising the scheme upon which the *EQR claim* dialogue will be based (i.e., the AS that will be employed as the starting point of the dialogue). Its logical structure can be seen as halfway between the *Argument Schemes Over Proposal for Action* (ASOPA) [7] and the *Argument Scheme from Expert Opinion* (ASEO) [139]. The underlying idea is to merge the knowledge elicited by those two formal patterns in a single scheme that would then yield the advantage of concentrating and synthesizing the same amount of information in a unique data structure that may be queried more conveniently. Indeed, the architecture of the EQR scheme has been engineered having in mind its concrete implementation. By combining theoretical and practical bits of information, the presented structure serves as a placeholder capable of conveying enough data to clarify the topic of interest. This proves particularly effective in explanatory contexts, where the retrieval of specific information

[i]Refer to Chapter 2 for a brief overview of dialogue types and their interactions.

may result in a time-consuming search through a plethora of different sources. EQR scheme instantiations can then behave as small-scale (space-saving) database units that can be interrogated by users. Orderly storing information, which would require double the number of ASOPA and ASEO instantiations to be collected, results in a (potential) reduction of memory and time consumption, proving thus suited for the scarce resource availability issue of real-world agents.

More specifically, the purpose of the EQR scheme is to formalise the consequences arising (and the presumptive reasoning leading to them) by acting upon a specific expert opinion. A reference to such authority provides the rationale that justifies the conclusion of the (instantiated) argument, also leaving chances of inquiry for more detailed explanations.

***EQR claim* Scheme**
Premise : In the current state R *Premise* : asserting α (from an expert E in a field F) *Premise* : will result in a new state S *Premise* : which will make proposition A true (alternatively, false) *Premise* : which will promote some value v
Conclusion : Acting upon α should make proposition A true (false) and entail value v

Being the core element of *EQR claim* dialogues, we call it the *EQR claim* scheme to distinguish it from the other EQR formalisations that will be introduced in the next sections. The proposed scheme assumes the existence of:

- A finite set of knowledgeable experts, called *Experts*, denoted with elements E, E', E'', etc. Experts are deemed knowledgeable if they can somehow prove their competencies (e.g., years of experience, professional achievements, research publications).

- A finite set of relevant disciplinary fields, called *Fields*, denoted with elements F, F', F'', etc.

- A finite set of propositions, called *Opinions*, denoted with elements α, β, γ, etc. Each member represents the viewpoint of an expert with regard to a specific topic.

- A finite set of *Opinions* × *Experts* × *Fields* tuples, called *Competences*, denoted with elements $\alpha_{\langle E,F \rangle}$, $\beta_{\langle E',F' \rangle}$, $\gamma_{\langle E'',F'' \rangle}$, etc. Each element describes an opinion α from an expert E in a field F[ii].

- A finite set of propositions and their negations, called *Prop*, denoted with elements *A, ¬A, B, ¬B,* etc.

- A finite set of states, called *States*, denoted with elements *R, S, T,* etc. Each element describes a specific state of the world and corresponds to an assignment of truth values $\{T,F\}$ to every element of *Prop*.

- A finite set of *Values* denoted with elements *v, w,* etc. This category includes both positive (i.e., constructive, such as wellbeing, altruism, integrity, etc.) and negative (i.e., non-constructive, such as dishonesty, manipulation, greed, etc.) values.

- A function *value* mapping each element of *Prop* to a pair $(v, sign)$, where $v \in$ *Values* and $sign \in \{+,-,=\}$.

- A ternary relation `assert` on *Competences* × *States* × *States*.

Intuitively, starting from the current circumstance *R* and acting upon the opinion asserted by a competent expert in the relevant field, the agent instantiating the scheme wishes to attain *A* (or not *A*) and the actual reason for it (value *v*), along with the entailed consequences, whether they are desired or not (new state *S*). As an example of expert opinion, consider an architect asserting that, according to her recent evaluation, the nearby bridge requires immediate maintenance to prevent its collapse. In this case, by acting upon such an opinion, the practical intervention of specialized workers will change the state of the world into a new state where the bridge is no longer precarious (promoting the safety value).

In addition, four statements need to be satisfied if the argument represented by the formalisation is to be valid:

Statement 1: *R* is the case.

Statement 2: $\texttt{assert}(\alpha_{\langle E,F \rangle}, R, S) \in \texttt{assert}$.

[ii]Notice that the same expert can express various opinions in the same field (i.e. $\beta_{\langle E,F \rangle}$). Alternatively, different experts can express the same opinions in the same field (i.e. $\alpha_{\langle E''',F \rangle}$) or even different opinions (i.e. $\gamma_{\langle E''',F \rangle}$). Finally, consider also that the same expert can be an expert in multiple fields (i.e. $\delta_{\langle E,F' \rangle}$).

Statement 3: $S \models A$ (i.e., "A is true in the state S"); alternatively, $S \models \neg A$ (i.e., "$\neg A$ is true in the state S").

Statement 4: $value(A) = (v, +)$

We can represent the *EQR claim* scheme following the diagrammatic form of [6]:

$$R \xrightarrow{\alpha_{\langle E,F \rangle}} S \models A \uparrow v \qquad (\star)$$

The intended meaning is: *"Asserting the expert's opinion* $\alpha_{\langle E,F \rangle}$ *(and acting upon it) in the current state R, results in a new state S in which proposition A (alternatively,* $\neg A$*) is true, and this promotes value v".*

5.2.1 Syntax

To assess the validity of an *EQR claim* scheme instantiation we are going to primarily harness the critical questions of the Argument Scheme from Expert Opinion, although we first need to convert each of them into specific attacks recognized by the *EQR* dialogue protocol (following an approach akin to [7]). The definition of such attacks will additionally consider two more critical questions:

CQ8. (*Disciplinary-relevant Question*) Is F a relevant disciplinary field?

As an example, we can consider an expert (E) in the field of Wiccan witchcraft[iii] (F). Thanks to its expertise, E's opinion *"Wiccan healing rituals can cure every kind of disease"* supports the proposition *"There exists a cure against COVID-19"* (A). Observe that it could be possible to build an EQR scheme instantiation such that, using the above information, all the critical questions (**CQ1-CQ6**) will be addressed positively. Nevertheless, the majority of the rational agents will not actually believe that Wiccan magic has any disciplinary relevance to A. As such, healing rituals cannot be considered a valid alternative to medical sciences to develop a cure for the COVID-19 virus.

Besides **CQ8**, we can introduce another (self-explanatory) critical question:

CQ9. (*Value Question*) Is E's assertion promoting a negative value?

[iii] Wicca, or Pagan Witchcraft, is a movement halfway between a neo-pagan religion and an occultist stream among the Western esotericism. It was first introduced to the public in 1954 by Gerald Gardner, a British Civil Servant.

It is also important to highlight that the *EQR claim* dialogue does not admit each ASEO critical question. **CQ7** relates to the endorsement of the expert's assertion which leads to the endorsement of the evaluated proposition. Since the *EQR claim* dialogue does not take into account such endorsement, **CQ7** will not be considered. Finally, for a more suitable formalisation, we are also going to rephrase some of the already established critical questions:

CQ2. *Is E an expert in the field F that* α *is in?*

CQ3. *Did E's assertion imply A?*

CQ4. *Is E's assertion entailing contradictory propositions?*

CQ6. *Is E's assertion based on the (facts expressed by) state R?*

List of possible attacks Having in mind the *EQR claim* scheme structure of ($\star$), we can introduce a series of attacks and their variants:

Attack 1a.(CQ1) It is not the case that $E \in \mathit{Experts}$.

Attack 1b.(CQ1) It is not the case that E, and there is an $E' \in \mathit{Experts}$ where $E' \neq E$ such that E' is the case.

Attack 2a.(CQ2) [*rephrased as "Is E an expert in the field F that* α *is in?"*] It is not the case that $\alpha_{\langle E,F\rangle} \in \mathit{Competences}$.

Attack 2b.(CQ2) [*rephrased as "Is E an expert in the field F that* α *is in?"*] It is not the case that $\alpha_{\langle E,F\rangle}$, and there is a $\beta_{\langle E',F'\rangle} \in \mathit{Competences}$ where $\beta_{\langle E',F'\rangle} \neq \alpha_{\langle E,F\rangle}$ such that $\beta_{\langle E',F'\rangle}$ is the case.

Attack 3a.(CQ3) [*rephrased as "Did E's assertion imply A?"*] It is not the case that $\mathtt{assert}(\alpha_{\langle E,F\rangle}, R, S) \in \mathtt{assert}$ and $S \models A$.

Attack 3b.(CQ3) [*rephrased as "Did E's assertion imply A?"*] It is the case that $\mathtt{assert}(\alpha_{\langle E,F\rangle}, R, S) \in \mathtt{assert}$, but $S \not\models A$, $S \models B$, $B \in \mathit{Prop}$ and $B \neq A$.

Attack 3c.(CQ3) [*rephrased as "Did E's assertion imply A?"*] It is not the case that $\mathtt{assert}(\alpha_{\langle E,F\rangle}, R, S) \in \mathtt{assert}$, and it is the case that $\mathtt{assert}(\alpha_{\langle E,F\rangle}, R, T) \in \mathtt{assert}$, where $T \in \mathit{States}$, $T \neq S$, but it is not the case that $T \models A$.

Attack 4.(CQ4) [*rephrased as "Is E's assertion entailing contradictory propositions?"*] It is the case that $\mathtt{assert}(\alpha_{\langle E,F\rangle}, R, S) \in \mathtt{assert}$, $S \models A$ and $S \models \neg A$.

Attack 5.(CQ5) It is the case that there is an $\texttt{assert}(\beta_{\langle E',F\rangle}, R, T) \in \texttt{assert}$, $\beta_{\langle E',F\rangle} \in \mathit{Competences}$, $\beta_{\langle E',F\rangle} \neq \alpha_{\langle E,F\rangle}$, $T \in \mathit{States}$, $T \models \neg A$ and $T \neq S$.

Attack 6a.(CQ6) [*rephrased as "Is E's assertion based on the (facts expressed by) state R?"*] It is not the case that R.

Attack 6b.(CQ6) [*rephrased as "Is E's assertion based on the (facts expressed by) state R?"*] It is not the case that $R \in \mathit{States}$.

Attack 6c.(CQ6) [*rephrased as "Is E's assertion based on the (facts expressed by) state R?"*] It is not the case that R, and there is a $Q \in \mathit{States}$ where $Q \neq R$ such that Q is the case.

Attack 7a.(CQ8) It is not the case that $F \in \mathit{Fields}$.

Attack 7b.(CQ8) It is not the case that F, and there is an $F' \in \mathit{Fields}$, where $F' \neq F$, such that F' is the case.

Attack 8a.(CQ9) It is the case that $\texttt{assert}(\alpha_{\langle E,F\rangle}, R, S) \in \texttt{assert}$, $S \models A$, but $\mathit{value}(A) = (v, -)$.

Attack 8b.(CQ9) It is the case that $\texttt{assert}(\alpha_{\langle E,F\rangle}, R, S) \in \texttt{assert}$, $S \models A$, but $\mathit{value}(A) = (w, -)$, $w \in \mathit{Values}$ and $w \neq v$.

Attack 8c.(CQ9) It is the case that $\texttt{assert}(\alpha_{\langle E,F\rangle}, R, S) \in \texttt{assert}$, but $S \not\models A$, $S \models B$, $B \in \mathit{Prop}$, $\mathit{value}(B) = (v, -)$, and $B \neq A$.

Attack 8d.(CQ9) It is the case that $\texttt{assert}(\alpha_{\langle E,F\rangle}, R, S) \in \texttt{assert}$, $S \models A$. Also $S \models B$, $B \in \mathit{Prop}$, $\mathit{value}(B) = (w, -)$, $w \in \mathit{Values}$, where $B \neq A$ and $w \neq v$.

The following attacks (and their variants) target specific elements of the *EQR claim* scheme and are not related to any particular critical question:

Attack 9a. It is not the case that $\alpha \in \mathit{Opinions}$.

Attack 9b. It is not the case that α, and there is a $\beta \in \mathit{Opinions}$, where $\beta \neq \alpha$, such that β is the case.

Attack 10a. It is not the case that $\texttt{assert}(\alpha_{\langle E,F\rangle}, R, S) \in \texttt{assert}$.

Attack 10b. It is not the case that $\texttt{assert}(\alpha_{\langle E,F\rangle}, R, S)$, and it is the case that $\texttt{assert}(\alpha_{\langle E,F\rangle}, R, T) \in \texttt{assert}$, where $T \in \mathit{States}$ and $T \neq S$.

Attack 11a. It is not the case that $S \in States$.

Attack 11b. It is not the case that S, and there is a $T \in States$, where $T \neq S$, such that T is the case.

Attack 12a. It is not the case that $S \models A$.

Attack 12b. It is not the case that $S \models A$, and there is a $B \in Prop$, where $B \neq A$, such that $S \models B$ is the case

Attack 13a. It is not the case that $A \in Prop$.

Attack 13b. It is not the case that A, and there is a $B \in Prop$, where $B \neq A$, such that B is the case.

Attack 14a. It is not the case that $value(A) = (v, +)$.

Attack 14b. It is not the case that $value(A) = (v, +)$, and there is a value $w \in Values$, where $w \neq v$, such that $value(A) = (w, +)$ or $value(A) = (w, -)$.

Attack 15a. It is not the case that $v \in Values$.

Attack 15b. It is not the case that v, and there is a $w \in Values$, where $w \neq v$, such that w is the case.

The purpose of the above attacks formalisation is twofold: (I) identifies, as they are, the specific categories of challenges that can be moved against the *EQR claim* scheme to test the validity of its instantiations; (II) serves as templates that can be instantiated to generate every possible kind of attacks in an actual dialogue implementation. That is to say, an example of (I) might be any attack (**1a-15b**) as it is, since it is challenging one (or more) aspect of the *EQR claim* scheme as in:

> *It is not the case that S, and there is a T ∈ States, where T ≠ S, such that T is the case.* (**Attack 11b**).

An example of (II) might be an attack generated by leveraging **Attack 11b** as a template and substituting S with its instance S_1 and T with its instance T_1:

> *It is not the case that S_1, and there is a $T_1 \in States$, where $T_1 \neq S_1$, such that T_1 is the case.*

The *EQR claim* dialogue presents two syntactic layers: (i) an innermost layer in which the contents of the utterances are expressed in a formal way through propositional logic; (ii) an outermost layer which expresses the illocutionary force of the single utterances [80][iv]. This outermost, wrapping layer can be represented by listing all the possible locutions of the dialogue as detailed in the first column of Table A.1, Table A.2, Table A.3, Table A.4, and Table A.5[v].

Resolution of attacks Since there are different types of attacks that entail different possible resolutions, we should start by grouping the attacks into four main classes [7]:

- **Attacks concerning factual disagreement** This kind of attacks involves the nature of the current state of the world (including causal relations).
- **Attacks concerning representation** This kind of attacks involves issues related to the language and logic being used.
- **Attacks concerning different preferences** This kind of attacks involves the different ranking of the players' preferences.
- **Attacks concerning clarification of a position** This kind of attacks involves questioning a specific position of the contender. The answer will give a clarification useful for moving an attack.

Assuming that every participant of the *EQR claim* dialogue pre-emptively agrees on the involved ontology (independently from the kind of attacks moved by the players), it is then possible to identify two forms of resolution: (a) value-preferred defeat or (b) rational disagreement. Both types of resolution require a means for evaluating defeats according to the ranked-value order of the players. For this purpose, it is possible to employ any argumentation theory capable of handling defeats. The rational disagreement is then formalised via the generation of two different (and conflicting) admissible extensions, each of which is related to the preference of one (team of) player.

[iv]Notice, that in [80] the authors presented an additional middle layer in which other elements of the dialogue, such as the sender and the recipient, are identified. EQR dialogues, however, do not necessitate such an additional layer since the protocol already establishes these roles among the players.

[v]Once again, we are following an approach similar to the one adopted in [7] for the PARMA protocol.

5.2.2 Semantics

An axiomatic semantics for the *EQR claim* dialogue presents the pre-conditions necessary for the legal utterance of each locution under the protocol, and any post-conditions arising from their legal utterance (Table A.1, Table A.2, Table A.3, Table A.4, and Table A.5). Such pre and post-conditions influence the *commitment store* of each agent participating in the dialogue. These commitment stores are public statements that the agents have to defend in the dialogue (unless they are withdrawn), but they might not correspond to the agent's real beliefs or intentions.

5.2.3 Turns structure and Winning Conditions

Having in mind the semantic pre- and post-conditions of each locution of the *EQR claim* dialogue protocol, we can informally identify the ordered sequence of locutions that distinguishes every turn of a player. We can determine two parties of agents playing the dialogue (which can also be composed of one element each): the proponent team (PRO), i.e., the *explainer* agent/s that claim(s) the truth of proposition A yielded by (acting upon) the expert opinion (which results in the instantiation of an EQR scheme); the opponent team (OPP), i.e., the *explainee* agent/s trying to challenge the proponent claim. The goal of OPP is to successfully attack the argument schemes instantiated by PRO which, in turn, have to counter every such attack. Notice that the purpose of the explainee is to retrieve information and understand the rationale behind the received explanation (recall that an EQR dialogue is a mixture of persuasion, information-giving/seeking[vi] and query dialogues) rather than suggesting its own view on the subject. The ordered sequence of locutions can then be summarized as follows:

- PRO is the first to play. Its first turn will consist only in the utterance of the locution **state 'something'**, where 'something' comprises all the information needed to instantiate the initial EQR argument scheme.
- A turn can finish only after:
    - (*PRO's turn*) an attack has been moved, a question has been asked or a statement has been uttered. That is to say, if a locution **deny 'something'**, **ask 'something'** (where 'something' refers to the target of the attack or the query) or **state 'something'** (where 'something' comprises all the information needed to answer a question) is uttered by the team of players. Recall that PRO's task

[vi]Such that the explainer gives information and the explainee seeks information.

is to counter every challenge against the initial EQR argument scheme instantiation it moved, as such, *it must* utter locutions that serve this purpose.

  - (*OPP's turn*) an attack has been moved, a question has been asked or a statement has been uttered. That is to say, if a locution **deny 'something'**, **deny initial 'something'**, **ask 'something'** (where 'something' refers to the target of the attack or the query) or **state 'something'** (where 'something' comprises all the information needed to answer a question) is uttered by the team of players[vii]. Recall that OPP's task is to challenge the initial EQR argument scheme instantiation moved by PRO, as such, *it must* utter locutions that serve this purpose.

- The team to whom the attack, question or response of the previous turn was addressed must begin its current turn with the locution **concede**, **reject** (if the party is, respectively, accepting the result of an attack, accepting the response to its previous query, claiming it does not know an answer, or disagreeing with the result of an attack or a response) or the locution **state 'something'** (if the party is answering the 'something' asked by the other team's previous question). Notice that OPP is the only one *not obliged* to counter the previous move from PRO: even if it could do otherwise, it can simply concede it (uttering **concede**) and move another attack or ask another question.

- No player can perform more than one locution per turn except for the ones designed for controlling the dialogue (Table A.1).

- **enter dialogue** (as well as **leave dialogue**) is a locution performed only one time during the whole dialogue.

Winning conditions In an EQR dialogue, where we need to assess the reliability of an expert opinion and how (acting upon) this entails a specific proposition, the 'burden of proof' lies on PRO. Indeed, it is the proponent who needs to show the validity of its initial argument scheme instantiation and persuade its contender, while for OPP it suffices to successfully attack or question it (i.e., to attack or to question such that PRO cannot respond to it with other than a **concede** locution). We can then informally define the winning condition of the two teams of players as:

[vii]Observe that **deny initial 'something'** identifies an attack that directly targets the initial EQR scheme instantiation moved by PRO.

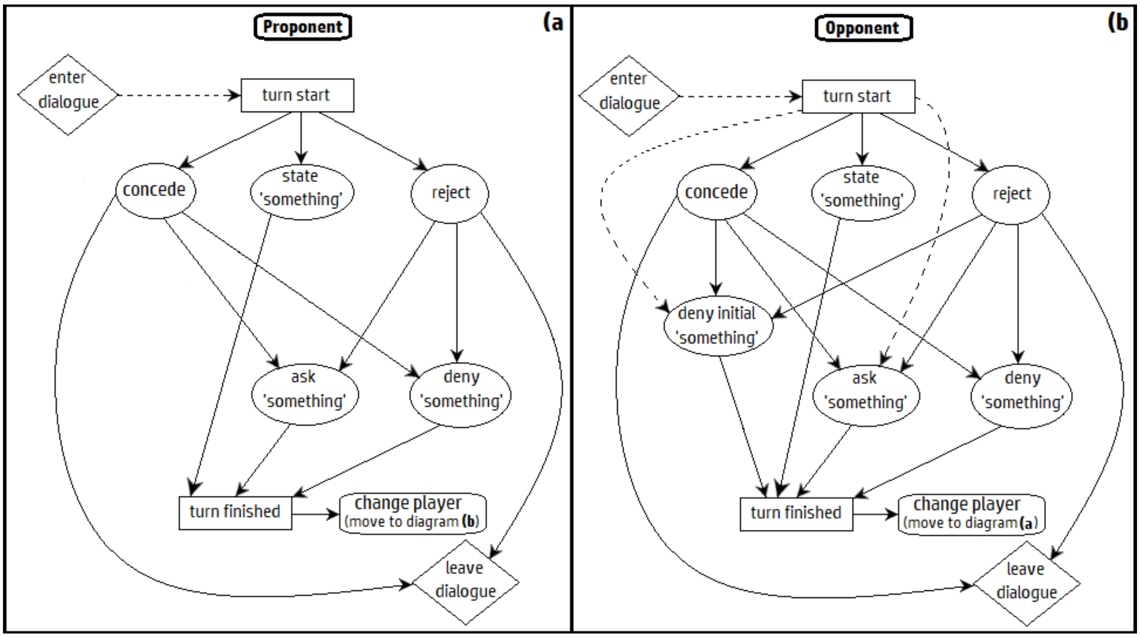

Figure 5.1: Ordered sequences of locutions describing the turns of each player. The dashed arrows denote that the corresponding moves must be performed during the first turn only (e.g., in all the subsequent turns, the player will start from the locution **turn start** rather than **enter dialogue**). Notice also that the opponent will always prefer to utter a **deny initial 'something'** rather than a **deny 'something'** or **ask 'something'**. This preference is emphasized in the graph by the different positions of the locutions.

(PRO) *The proponent wins if the opponent leaves the dialogue*[viii]. *PRO has countered every possible attack/answered every possible question moved by the contender party which is now persuaded about the validity of the initial argument scheme instantiation. This means that OPP has uttered the locution **concede** before the locution **leave dialogue**.*

(OPP) *The opponent wins if the proponent leaves the dialogue. OPP successfully attacks/inquires the initial argument scheme instantiation of the contender party raising at least one doubt about its validity. This means that PRO has uttered the locution **concede** before the locution **leave dialogue**.*

(Draw) *The utterance of **reject** before **leave dialogue** (by either contender) implies that the last locutions of each player have the same level of preference. As a result, this might create two different and conflicting admissible extensions (according to Dung's semantics* [53]*). Such extensions would represent the rational disagreement*

[viii]A similar idea of winning conditions is proposed by Krabbe: "[...] whosoever abandons a chain of arguments has lost that chain of arguments, and that who loses the last chain of arguments, loses the discussion as a whole" [78].

reached by the two parties of agents and the respective positions[ix].

5.3 The *EQR endorsement* Dialogue

Before delving into the creation of the actual protocol, let us begin by formalising the scheme upon which the *EQR endorsement* dialogue will be based. The formalisation and underlying idea will represent an extended version of the *EQR claim* scheme where the focus is now on the proposition (and related value) promoted by (acting upon) the *endorsement* of the expert assertion.

***EQR endorsement* Scheme**
Premise : In the current state R *Premise* : endorsing assertion α (from an expert E in a field F) *Premise* : will result in a new state S *Premise* : which will make proposition A true (alternatively, false) *Premise* : which will promote some value v
Conclusion : Acting upon the endorsement of α should make proposition A true (false) and entail value v

Being the core element of *EQR endorsement* dialogues, we call it the *EQR endorsement* scheme to distinguish it from the other *EQR* formalisations. The introduced scheme assumes the existence of:

- A finite set of knowledgeable experts, called *Experts*, denoted with elements E, E', E'', etc. Experts are deemed knowledgeable if they can somehow prove their competencies (e.g., years of experience, professional achievements, research publications).
- A finite set of disciplinary relevant fields, called *Fields*, denoted with elements F, F', F'', etc.
- A finite set of propositions, called *Opinions*, denoted with elements α, β, γ, etc. Each member represents the viewpoint of an expert with regard to a specific topic.
- A finite set of *Opinions* $\times$ *Experts* $\times$ *Fields* tuples, called *Competences*, denoted with elements $\alpha_{\langle E,F \rangle}$, $\beta_{\langle E',F' \rangle}$, $\gamma_{\langle E'',F'' \rangle}$, etc. Each element describes an opinion α from an expert E in a field F.

[ix] Consider that a draw can be solved by an additional inquiry dialogue to adjust the preference ordering between players.

- A unary relation assert on *Competences* with $\texttt{assert}(\alpha_{\langle E,F \rangle})$ to be read as *"E, which is an expert in the field F, asserts opinion α"*. Notice that, unlike *Competences*, assert emphasizes the public act of expressing (asserting) the expert's opinion.

- A finite set of propositions and their negations, called *Prop*, denoted with elements *A, ¬A, B, ¬B,* etc.

- A finite set of states, called *States*, denoted with elements *R, S, T,* etc. Each element describes a specific state of the world and corresponds to an assignment of truth values $\{T,F\}$ to each element of *Prop*.

- A finite set of *Values* denoted with elements *v, w,* etc. This category includes both positive (i.e., constructive, such as wellbeing, altruism, integrity, etc.) and negative (i.e., nonconstructive, such as dishonesty, manipulation, greed, etc.) values.

- A function *value* mapping each element of *Prop* to a pair $(v, sign)$, where $v \in Values$ and $sign \in \{+,-,=\}$.

- A ternary relation endorsement on assert $\times$ *States* $\times$ *States*.

In addition, four statements need to be satisfied if the argument represented by the formalisation is to be valid:

Statement 1: *R* is the case.

Statement 2: $\mathsf{endorsement}(\texttt{assert}(\alpha_{\langle E,F \rangle}), R, S) \in \mathsf{endorsement}$.

Statement 3: $S \models A$ (i.e., "*A* is true in the state *S*"); alternatively, $S \models \neg A$ (i.e., "$\neg A$ is true in the state *S*").

Statement 4: $value(A) = (v, +)$.

We can represent the *EQR endorsement* scheme following the diagrammatic form of [6]:

$$R \xrightarrow{\texttt{assert}(\alpha_{\langle E,F \rangle})} S \models A \uparrow v \qquad (*)$$

The intuitive meaning is: *"(acting upon) the endorsement of the expert's opinion* $\texttt{assert}(\alpha_{\langle E,F \rangle})$ *in the current state R, results in a new state S in which proposition A (alternatively, ¬A) is true, and this promotes value v"*.

5.3.1 Syntax

Before defining the possible attacks allowed by the protocol we can consider adapting (and replacing) some of the critical questions for the *EQR claim* scheme:

CQ3a. (*Endorsement Opinion Question*) What did E assert such that its endorsement implies A?

CQ4a. (*Endorsement Trustworthiness Question*) Is E personally reliable as a source for an endorsement?

CQ5a. (*Endorsement Consistency Question*) Is the endorsement of E's assertion consistent with the endorsement of other experts' assertions?

CQ6a. (*Endorsement Evidence Question*) Is the endorsement of E's assertion based on evidence?

CQ9a. (*Endorsement Value Question*) Is the endorsement of E's assertion promoting a negative value?

Question **CQ3a** will substitute **CQ3** since the implication of proposition *A* passes now via the endorsement of the expert's assertion rather than its assertion alone. Similar reasoning can be applied also for **CQ4a**, **CQ5a**, **CQ6a** and **CQ9a**, which will replace, respectively, **CQ4**, **CQ5**, **CQ6** and **CQ9**. Observe that, since the *EQR endorsement* dialogue formalises endorsement, we should expect **CQ7** to be present among the possible attacks. However, being **CQ9a** the generalisation of **CQ7** (therefore, subsuming it) we do not need to consider **CQ7**.

Finally, for a more suitable formalisation, we are also going to rephrase some of the already established critical questions:

CQ3a. *Did the endorsement of E's assertion imply A?*

CQ4a. *Is the endorsement of E's assertion entailing contradictory propositions?*

CQ6a *Is the endorsement of E's assertion based on the (facts expressed by) state R?*

List of possible attacks Having in mind the *EQR endorsement* scheme structure of ($*$), we can introduce a series of attacks and their variants:

Attack 1a.(CQ1) It is not the case that $E \in Experts$.

Attack 1b.(CQ1) It is not the case that E, and there is an $E' \in Experts$ where $E' \neq E$ such that E' is the case.

Attack 2a.(CQ2) [*rephrased as "Is E an expert in the field F that α is in?"*] It is not the case that $\alpha_{\langle E,F \rangle} \in Competences$.

Attack 2b.(CQ2) [*rephrased as "Is E an expert in the field F that α is in?"*] It is not the case that $\alpha_{\langle E,F \rangle}$, and there is a $\beta_{\langle E',F' \rangle} \in Competences$ where $\beta_{\langle E',F' \rangle} \neq \alpha_{\langle E,F \rangle}$ such that $\beta_{\langle E',F' \rangle}$ is the case.

Attack 3a.(CQ3a) [*rephrased as "Did the endorsement of E′s assertion imply A?"*] It is not the case that $\mathsf{endorsement}(\texttt{assert}(\alpha_{\langle E,F \rangle}), R, S) \in \mathsf{endorsement}$ and $S \models A$.

Attack 3b.(CQ3a) [*rephrased as "Did the endorsement of E′s assertion imply A?"*] It is the case that $\mathsf{endorsement}(\texttt{assert}(\alpha_{\langle E,F \rangle}), R, S) \in \mathsf{endorsement}$, but $S \not\models A$, $S \models B$, $B \in Prop$ and $B \neq A$.

Attack 3c.(CQ3a) [*rephrased as "Did the endorsement of E′s assertion imply A?"*] It is not the case that $\mathsf{endorsement}(\texttt{assert}(\alpha_{\langle E,F \rangle}), R, S) \in \mathsf{endorsement}$, and it is the case that $\mathsf{endorsement}(\texttt{assert}(\alpha_{\langle E,F \rangle}), R, T) \in \mathsf{endorsement}$, where $T \in States$, $T \neq S$, but it is not the case that $T \models A$.

Attack 4.(CQ4a) [*rephrased as "Is the endorsement of E's assertion entailing contradictory propositions?"*] It is the case that $\mathsf{endorsement}(\texttt{assert}(\alpha_{\langle E,F \rangle}), R, S) \in \mathsf{endorsement}$, $S \models A$ and $S \models \neg A$.

Attack 5.(CQ5a) It is the case that there is an $\mathsf{endorsement}(\texttt{assert}(\beta_{\langle E',F \rangle}), R, T) \in \mathsf{endorsement}$, $\texttt{assert}(\beta_{\langle E',F \rangle}) \in \texttt{assert}$, $\texttt{assert}(\beta_{\langle E',F \rangle}) \neq \texttt{assert}(\alpha_{\langle E,F \rangle})$, $T \in States$, $T \models \neg A$ and $T \neq S$.

Attack 6a.(CQ6a) [*rephrased as "Is the endorsement of E's assertion based on the (facts expressed by) state R?"*] It is not the case that R.

Attack 6b.(CQ6a) [*rephrased as "Is the endorsement of E's assertion based on the (facts expressed by) state R?"*] It is not the case that $R \in States$.

Attack 6c.(CQ6a) [*rephrased as "Is the endorsement of E's assertion based on the (facts expressed by) state R?"*] It is not the case that R, and there is a $Q \in States$ where $Q \neq R$ such that Q is the case.

Attack 7a.(CQ8) It is not the case that $F \in Fields$.

Attack 7b.(CQ8) It is not the case that F, and there is an $F' \in Fields$, where $F' \neq F$, such that F' is the case.

Attack 8a.(CQ9a) It is the case that $\mathsf{endorsement}(\mathtt{assert}(\alpha_{\langle E,F\rangle}),R,S) \in \mathsf{endorsement}$, $S \models A$, but $value(A) = (v,-)$.

Attack 8b.(CQ9a) It is the case that $\mathsf{endorsement}(\mathtt{assert}(\alpha_{\langle E,F\rangle}),R,S) \in \mathsf{endorsement}$, $S \models A$, but $value(A) = (w,-)$, $w \in Values$ and $w \neq v$.

Attack 8c.(CQ9a) It is the case that $\mathsf{endorsement}(\mathtt{assert}(\alpha_{\langle E,F\rangle}),R,S) \in \mathsf{endorsement}$, but $S \not\models A$, $S \models B$, $B \in Prop$, $value(B) = (v,-)$, and $B \neq A$.

Attack 8d.(CQ9a) It is the case that $\mathsf{endorsement}(\mathtt{assert}(\alpha_{\langle E,F\rangle}),R,S) \in \mathsf{endorsement}$, $S \models A$. Also $S \models B$, $B \in Prop$, $value(B) = (w,-)$, $w \in Values$, where $B \neq A$ and $w \neq v$.

The following attacks (and their variants) target specific elements of the *EQR endorsement* scheme and are not related to any particular critical question:

Attack 9a. It is not the case that $\alpha \in Opinions$.

Attack 9b. It is not the case that α, and there is a $\beta \in Opinions$, where $\beta \neq \alpha$, such that β is the case.

Attack 10a. It is not the case that $\mathtt{assert}(\alpha_{\langle E,F\rangle}) \in \mathtt{assert}$.

Attack 10b. It is not the case that $\mathtt{assert}(\alpha_{\langle E,F\rangle})$, and there is a $\mathtt{assert}(\beta_{\langle E',F'\rangle}) \in \mathtt{assert}$ where $\mathtt{assert}(\beta_{\langle E',F'\rangle}) \neq \mathtt{assert}(\alpha_{\langle E,F\rangle})$ such that $\mathtt{assert}(\beta_{\langle E',F'\rangle})$ is the case.

Attack 11a. It is not the case that $\mathsf{endorsement}(\mathtt{assert}(\alpha_{\langle E,F\rangle}),R,S) \in \mathsf{endorsement}$.

Attack 11b. It is not the case that $\mathsf{endorsement}(\mathtt{assert}(\alpha_{\langle E,F\rangle}),R,S)$, and it is the case that $\mathsf{endorsement}(\mathtt{assert}(\alpha_{\langle E,F\rangle}),R,T) \in \mathsf{endorsement}$, where $T \in States$ and $T \neq S$.

Attack 12a. It is not the case that $S \in States$.

Attack 12b. It is not the case that S, and there is a $T \in States$, where $T \neq S$, such that T is the case.

Attack 13a. It is not the case that $S \models A$.

Attack 13b. It is not the case that $S \models A$, and there is a $B \in Prop$, where $B \neq A$, such that $S \models B$ is the case.

Attack 14a. It is not the case that $A \in Prop$.

Attack 14b. It is not the case that A, and there is a $B \in Prop$, where $B \neq A$, such that B is the case.

Attack 15a. It is not the case that $value(A) = (v, +)$.

Attack 15b. It is not the case that $value(A) = (v, +)$, and there is a value $w \in Values$, where $w \neq v$, such that $value(A) = (w, +)$ or $value(A) = (w, -)$.

Attack 16a. It is not the case that $v \in Values$.

Attack 16b. It is not the case that v, and there is a $w \in Values$, where $w \neq v$, such that w is the case.

The syntax of the *EQR endorsement* dialogue is exactly like the *EQR claim* one. The few differences between protocols are listed in Table A.6, Table A.7, Table A.8, and Table A.9.

5.3.2 Semantics

An axiomatic semantics for the *EQR endorsement* dialogue presents the pre-conditions necessary for the legal utterance of each locution under the protocol, and the post-conditions arising from their legal utterance, along with any influence this might have on the *commitment store*. The conditions characterising the *EQR endorsement* dialogue are the same as the *EQR claim* dialogue with the differences listed in Table A.6, Table A.7, Table A.8, and Table A.9.

5.3.3 Turns structure and Winning Conditions

The ordered sequence of locutions that describes each player's turn in an *EQR endorsement* dialogue is identical to the one presented in the *EQR claim* dialogue (Section 5.2.3). The only differences comprise the semantical pre and post-conditions established by the *EQR endorsement* protocol (as highlighted in Section 5.3.2).

Winning conditions Similarly, the winning conditions of an *EQR endorsement* dialogue are identical to the ones already introduced for the *EQR claim* dialogue (Section 5.2.3).

5.4 The *EQR endorsed-by-whom* Dialogue

Before delving into the creation of the actual protocol, let us begin by formalising the scheme upon which the *EQR endorsed-by-whom* dialogue will be based. The formalisation and underlying idea will represent an extended version of the *EQR endorsement* scheme in which we additionally consider *who is endorsing* the expert assertion.

***EQR endorsed-by-whom* Scheme**
Premise : In the current state R
Premise : endorsing assertion α (from an expert E in a field F)
Premise : by endorser(s) X
Premise : will result in a new state S
Premise : which will make proposition A true (alternatively, false)
Premise : which will promote some value v
Conclusion : Acting upon the endorsement of α by X should make proposition A true (false) and entail value v

Being the core element of *EQR endorsed-by-whom* dialogues, we call it the *EQR endorsed-by-whom* scheme to distinguish it from the other *EQR* formalisations. The introduced scheme assumes the existence of:

- A finite set of knowledgeable experts, called *Experts*, denoted with elements E, E', E'', etc. Experts are deemed knowledgeable if they can somehow prove their competencies (e.g., years of experience, professional achievements, research publications).

- A finite set of disciplinary relevant fields, called *Fields*, denoted with elements F, F', F'', etc.

- A finite set of propositions, called *Opinions*, denoted with elements α, β, γ, etc. Each member represents the viewpoint of an expert with regard to a specific topic.

- A finite set of *Opinions* $\times$ *Experts* $\times$ *Fields* tuples, called *Competences*, denoted with elements $\alpha_{\langle E,F \rangle}$, $\beta_{\langle E',F' \rangle}$, $\gamma_{\langle E'',F'' \rangle}$, etc. Each element describes an opinion α from an expert E in a field F.

- A unary relation `assert` on *Competences* with $\texttt{assert}(\alpha_{\langle E,F \rangle})$ to be read as *"E, which is an expert in the field F, asserts opinion* α*"*. Notice that, unlike *Competences*, `assert` emphasizes the public act of expressing (asserting) the expert's opinion.

- A finite set of propositions and their negations, called *Prop*, denoted with elements *A,* $\neg$*A, B,* $\neg$*B,* etc.

- A finite set of states, called *States*, denoted with elements *R, S, T,* etc. Each element describes a specific state of the world and corresponds to an assignment of truth values $\{T, F\}$ to each element of *Prop*.

- A finite set of *Values* denoted with elements *v, w,* etc. This category includes both positive (i.e., constructive, such as wellbeing, altruism, integrity, etc.) and negative (i.e., nonconstructive, such as dishonesty, manipulation, greed, etc.) values.

- A function *value* mapping each element of *Prop* to a pair (*v*, *sign*), where $v \in$ *Values* and $sign \in \{+, -, =\}$.

- A finite family of sets of endorsers, called *Endorsers*, denoting each group of endorsers with elements X, Y, Z, etc[x].

- A quaternary relation endorsement on `assert` $\times$ *Endorsers* $\times$ *States* $\times$ *States*

In addition, four statements need to be satisfied if the argument represented by the formalisation is to be valid:

Statement 1: *R* is the case.

Statement 2: $\mathsf{endorsement}(\texttt{assert}(\alpha_{\langle E,F \rangle}), X, R, S) \in \mathsf{endorsement}$.

[x]Notice that there can also be sets composed of one element only, i.e., $|X| = 1$.

Statement 3: $S \models A$ (i.e., "A is true in the state S"); alternatively, $S \models \neg A$ (i.e., "$\neg A$ is true in the state S").

Statement 4: $value(A) = (v, +)$.

We can represent the *EQR endorsed-by-whom* scheme following the diagrammatic form of [6]:

$$R \xrightarrow{\texttt{assert}(\alpha_{\langle E,F\rangle}),\, X} S \models A \uparrow v \qquad (\dagger)$$

The intuitive meaning is: *"(acting upon) the endorsement of the expert's opinion* $\texttt{assert}(\alpha_{\langle E,F\rangle})$ *by X in the current state R, results in a new state S in which proposition A (alternatively, ¬A) is true and this promotes value v".*

5.4.1 Syntax

Before defining the possible attacks allowed by the protocol we can consider adapting (and replacing) some of the existing critical questions for the *EQR claim* scheme:

CQ3b. (*X-Endorsement Opinion Question*) What did E assert such that its endorsement by X implies A?

CQ4b. (*X-Endorsement Trustworthiness Question*) Is E personally reliable as a source for X's endorsement?

CQ5b. (*X-Endorsement Consistency Question*) Is the endorsement of E's assertion by X consistent with the endorsement of X for other experts' assertions?

CQ6b. (*X-Endorsement Evidence Question*) Is the endorsement of E's assertion by X based on evidence?

CQ9b. (*X-Endorsement Value Question*) Is the endorsement of E's assertion by X promoting a negative value?

Question **CQ3b** will substitute **CQ3** since the implication of proposition A passes now via the endorsement (by an endorser X) of the expert's assertion rather than its assertion alone. Similar reasoning can be applied also for **CQ4b**, **CQ5b**, **CQ6b** and **CQ9b**, which will replace, respectively, **CQ4**, **CQ5**, **CQ6** and **CQ9**. Observe that, since the *EQR endorsed-by-whom* dialogue formalises endorsers, we should expect **CQ7** to be present among the possible attacks. However, being **CQ9b** the generalisation of **CQ7** (therefore, subsuming it) we do not need to consider **CQ7**.

Finally, for a more suitable formalisation, we are also going to rephrase some of the already established critical questions:

CQ3b. *Did the endorsement of E's assertion by X imply A?*

CQ4b. *Is the endorsement of E's assertion by X entailing contradictory propositions?*

CQ6b *Is the endorsement of E's assertion by X based on the (facts expressed by) state R?*

List of possible attacks Having in mind the *EQR endorsed-by-whom* scheme structure of (†), we can introduce a series of attacks and their variants:

Attack 1a.(CQ1) It is not the case that $E \in \mathit{Experts}$.

Attack 1b.(CQ1) It is not the case that E, and there is an $E' \in \mathit{Experts}$ where $E' \neq E$ such that E' is the case.

Attack 2a.(CQ2) [*rephrased as "Is E an expert in the field F that α is in?"*] It is not the case that $\alpha_{\langle E,F \rangle} \in \mathit{Competences}$.

Attack 2b.(CQ2) [*rephrased as "Is E an expert in the field F that α is in?"*] It is not the case that $\alpha_{\langle E,F \rangle}$, and there is a $\beta_{\langle E',F' \rangle} \in \mathit{Competences}$ where $\beta_{\langle E',F' \rangle} \neq \alpha_{\langle E,F \rangle}$ such that $\beta_{\langle E',F' \rangle}$ is the case.

Attack 3a.(CQ3b) [*rephrased as "Did the endorsement of E′s assertion by X imply A?"*] It is not the case that $\mathsf{endorsement}(\mathtt{assert}(\alpha_{\langle E,F \rangle}),X,R,S) \in \mathsf{endorsement}$ and $S \models A$.

Attack 3b.(CQ3b) [*rephrased as "Did the endorsement of E′s assertion by X imply A?"*] It is the case that $\mathsf{endorsement}(\mathtt{assert}(\alpha_{\langle E,F \rangle}),X,R,S) \in \mathsf{endorsement}$, but $S \not\models A$, $S \models B$, $B \in \mathit{Prop}$ and $B \neq A$.

Attack 3c.(CQ3b) [*rephrased as "Did the endorsement of E′s assertion by X imply A?"*] It is not the case that $\mathsf{endorsement}(\mathtt{assert}(\alpha_{\langle E,F \rangle}),X,R,S) \in \mathsf{endorsement}$, and it is the case that $\mathsf{endorsement}(\mathtt{assert}(\alpha_{\langle E,F \rangle}),X,R,T) \in \mathsf{endorsement}$, where $T \in \mathit{States}$, $T \neq S$, but it is not the case that $T \models A$.

Attack 4.(CQ4b) [*rephrased as "Is the endorsement of E's assertion by X entailing contradictory propositions?"*] It is the case that $\mathsf{endorsement}(\mathtt{assert}(\alpha_{\langle E,F \rangle}),X,R,S) \in \mathsf{endorsement}$, $S \models A$ and $S \models \neg A$.

Attack 5.(CQ5b) It is the case that there is an endorsement$(\texttt{assert}(\beta_{\langle E',F\rangle}), X, R, T)$ $\in$ endorsement, $\texttt{assert}(\beta_{\langle E',F\rangle}) \in \texttt{assert}$, $\texttt{assert}(\beta_{\langle E',F\rangle}) \neq \texttt{assert}(\alpha_{\langle E,F\rangle})$, $T \in$ *States*, $T \models \neg A$ and $T \neq S$.

Attack 6a.(CQ6b) [*rephrased as "Is the endorsement of E's assertion by X based on the* (*facts expressed by*) *state R?"*] It is not the case that R.

Attack 6b.(CQ6b) [*rephrased as "Is the endorsement of E's assertion by X based on the* (*facts expressed by*) *state R?"*] It is not the case that $R \in$ *States*.

Attack 6c.(CQ6b) [*rephrased as "Is the endorsement of E's assertion by X based on the* (*facts expressed by*) *state R?"*] It is not the case that R, and there is a $Q \in$ *States* where $Q \neq R$ such that Q is the case.

Attack 7a.(CQ8) It is not the case that $F \in$ *Fields*.

Attack 7b.(CQ8) It is not the case that F, and there is an $F' \in$ *Fields*, where $F' \neq F$, such that F' is the case.

Attack 8a.(CQ9b) It is the case that endorsement$(\texttt{assert}(\alpha_{\langle E,F\rangle}), X, R, S) \in$ endorsement, $S \models A$, but $value(A) = (v, -)$.

Attack 8b.(CQ9b) It is the case that endorsement$(\texttt{assert}(\alpha_{\langle E,F\rangle}), X, R, S) \in$ endorsement, $S \models A$, but $value(A) = (w, -)$, $w \in$ *Values* and $w \neq v$.

Attack 8c.(CQ9b) It is the case that endorsement$(\texttt{assert}(\alpha_{\langle E,F\rangle}), X, R, S) \in$ endorsement, but $S \not\models A$, $S \models B$, $B \in Prop$, $value(B) = (v, -)$, and $B \neq A$.

Attack 8d.(CQ9b) It is the case that endorsement$(\texttt{assert}(\alpha_{\langle E,F\rangle}), X, R, S) \in$ endorsement, $S \models A$. Also $S \models B$, $B \in Prop$, $value(B) = (w, -)$, $w \in$ *Values*, where $B \neq A$ and $w \neq v$.

The following attacks (and their variants) target specific elements of the *EQR endorsed-by-whom* scheme and are not related to any particular critical question:

Attack 9a. It is not the case that $\alpha \in$ *Opinions*.

Attack 9b. It is not the case that α, and there is a $\beta \in$ *Opinions*, where $\beta \neq \alpha$, such that β is the case.

Attack 10a. It is not the case that $\texttt{assert}(\alpha_{\langle E,F\rangle}) \in \texttt{assert}$.

Attack 10b. It is not the case that $\texttt{assert}(\alpha_{\langle E,F\rangle})$, and there is a $\texttt{assert}(\beta_{\langle E',F'\rangle}) \in$ $\texttt{assert}$ where $\texttt{assert}(\beta_{\langle E',F'\rangle}) \neq \texttt{assert}(\alpha_{\langle E,F\rangle})$ such that $\texttt{assert}(\beta_{\langle E',F'\rangle})$ is the case.

Attack 11a. It is not the case that $X \subseteq Endorsers$.

Attack 11b. It is not the case that X, and there is a $Y \subseteq Endorsers$, where $Y \neq X$, such that Y is the case.

Attack 12a. It is not the case that $\mathsf{endorsement}(\texttt{assert}(\alpha_{\langle E,F\rangle}), X, R, S) \in \mathsf{endorsement}$.

Attack 12b. It is not the case that $\mathsf{endorsement}(\texttt{assert}(\alpha_{\langle E,F\rangle}), X, R, S)$, and it is the case that $\mathsf{endorsement}(\texttt{assert}(\alpha_{\langle E,F\rangle}), X, R, T) \in \mathsf{endorsement}$, where $T \in$ *States* and $T \neq S$.

Attack 13a. It is not the case that $S \in States$.

Attack 13b. It is not the case that S, and there is a $T \in States$, where $T \neq S$, such that T is the case.

Attack 14a. It is not the case that $S \models A$.

Attack 14b. It is not the case that $S \models A$, and there is a $B \in Prop$, where $B \neq A$, such that $S \models B$ is the case.

Attack 15a. It is not the case that $A \in Prop$.

Attack 15b. It is not the case that A, and there is a $B \in Prop$, where $B \neq A$, such that B is the case.

Attack 16a. It is not the case that $value(A) = (v, +)$.

Attack 16b. It is not the case that $value(A) = (v, +)$, and there is a value $w \in Values$, where $w \neq v$, such that $value(A) = (w, +)$ or $value(A) = (w, -)$.

Attack 17a. It is not the case that $v \in Values$.

Attack 17b. It is not the case that v, and there is a $w \in Values$, where $w \neq v$, such that w is the case.

The syntax of the *EQR endorsed-by-whom* dialogue is exactly like the *EQR endorsement* one. The few differences between protocols are listed in Table A.10, Table A.11, Table A.12, and Table A.13.

5.4.2 Semantics

An axiomatic semantics for the *EQR endorsed-by-whom* dialogue presents the pre-conditions necessary for the legal utterance of each locution under the protocol, and the post-conditions arising from their legal utterance, along with any influence this might have on the *commitment store*. The conditions characterising the *EQR endorsed-by-whom* dialogue are the same as the *EQR endorsement* dialogue with the differences listed in Table A.10, Table A.11, Table A.12, and Table A.13.

5.4.3 Turns structure and Winning Conditions

The ordered sequence of locutions that describes each player's turn in an *EQR endorsed-by-whom* dialogue is identical to the one presented in the *EQR claim* dialogue (Section 5.2.3). The only differences comprise the semantical pre and post-conditions established by the *EQR endorsed-by-whom* protocol (as highlighted in Section 5.4.2).

Winning conditions Similarly, the winning conditions of an *EQR endorsed-by-whom* dialogue are identical to the ones already introduced for the *EQR claim* dialogue (Section 5.2.3).

5.5 Future Work

The EQR schemes have been created in three versions due to the relevance of the endorsement/endorsed-by-whom focus. This served to highlight the importance that such endorsement can have in the evaluation of the whole argument scheme from expert opinion. An example concerning the application of the *EQR* (*claim*) scheme in a medical scenario will be presented in Chapter 7. However, further investigations might also lead to interesting applications involving endorsement in fields such as social sciences, law and politics, i.e., in every area where it is not essential to focus on the truth-value of something. For example, let us consider a legal trial for a case of tax fraud where the jury has to decide (hence, act) upon the negligence of the indicted person. The defence may call to witness many professionals, such as a financial expert and a psychologist. The financial expert could testify about the execution of several suspicious transactions that are also very difficult to track completely. On the other hand, the psychologist could attest to a double personality disorder of the suspect and the possibility that these frauds have been accomplished without full awareness by the indicted person. In this circumstance, the final outcome of the trial depends on the endorsement given by the jury rather than the actual truth associated with each expert assertion. A similar example can be provided for a political campaign, where the election among two candidates relies on the endorsement provided by the voters rather than the actual truth claimed by the politicians. Sometimes, it is instead the identity of the endorsers that matters. For instance, in a board of shareholders, it may be convenient to discern who is endorsing a proposed idea. Indeed, the endorsement provided by the majority shareholder is more influential than the one given by any other member of the board.

For these reasons, *EQR endorsement* and *endorsed-by-whom* schemes may promote specific studies that concern the logical nature of the endorsement. This could lead to a formalisation that renders endorsements as a particular kind of argument labellings (different from the standard 3, 4 or 5-labellings approaches). Notice that this may involve a consideration of the illocutionary force conveyed by the EQR scheme instantiations, thus listing 'endorse' among the other pragmatical meanings (such as 'command' and 'promise') associated with speech acts. Alternatively, it could extend the research regarding bipolar AFs, denoted as BAFs [39, 131], enhancing the involved support relation[xi]. That is to say, BAFs could update their semantics by considering only the set of supports (representing the endorsements) rather than the set of defeats, thus mimicking the idea envisaged by the EQR (*endorsement/endorsed-by-whom*) schemes.

5.6 Conclusion

Two novel contributions arise from this chapter. Moving from the brief outline sketched in [90], a (1.) fully-fledged Explanation-Question-Response (EQR) dialogue protocol has been developed. This dialogue (considered as a new type that may be termed 'explanation') is halfway between persuasion, information-giving/seeking and query and already incorporates locutions for handling each of these tasks without the need for adopting additional tools. As such, it provides a model tailored for supplying explanations to agents eschewing formalism (as the Control Layer) that would unnecessarily complicate the protocol. Employed as the starting point of the dialogue, the (2.) EQR argument scheme constitutes the second result of this chapter. The underlying idea is to merge the knowledge elicited by the Argument Schemes Over Proposal for Action [7] and the Argument Scheme from Expert Opinion (ASEO) [139] in a single pattern that would then yield the advantage of concentrating and synthesizing the same amount of information in a unique data structure that may be queried more conveniently. In a nutshell, the purpose of the

[xi]Core element of BAFs, the support relation accompanies the defeat relation and the set of arguments to constitute the framework: BAFs = $\langle \text{AR}, \textit{defeat}, \textit{support} \rangle$ [39].

EQR scheme is to formalise the consequences arising (and the presumptive reasoning leading to them) by acting upon a specific expert opinion focusing on either the: assertion (*EQR claim* scheme), the endorsement (*EQR endorsement* scheme) or the specific endorsement (*EQR endorsed-by-whom* scheme). Each of these variations informs a corresponding version of the EQR dialogue protocol.

Chapter 6

EQR dialogue characterisations for resource-bounded real-world agents

Ideally, in order to provide more accurate dialectical interactions involving real-world agents, we would want to incorporate the convenient properties of Dialectical Cl-Arg (including the satisfaction of the rationality postulates and the practical desiderata under minimal resource consumption) into each EQR dialogue variants. However, the simple machinery of Dialectical Cl-Arg does not suffice in dealing properly with the inference rules encompassed by the argument schemes. A rule-based general framework as D-ASPIC$^+$ constitutes a better choice instead. Indeed, not only it can handle default reasoning via its embedded defeasible rules, but it can also preserve all the dialectical features of Dialectical Cl-Arg. As such, in this chapter, we present a D-ASPIC$^+$ model of the EQR argument scheme, denoted as *D-scheme*, and its instantiations. Such instantiations will provide EQR dialogue implementations that will characterise exchanges of arguments between real-world resource-bounded agents with a better approximation compared to the dialogues devised in Chapter 5.

6.1 D-ASPIC$^+$ EQR Scheme

D-scheme can be identified as the *EQR* scheme whose logical structure has been modelled as a D-ASPIC$^+$ argument. The advantages of this operation consist mainly in the possibility of generating *EQR* scheme instantiations that accounts for resource-bounded real-world agents. As such, these models enjoy properties P1-P4 (Proposition 1) which ensure to satisfy the rationality postulate under minimal resource consumption. In other words, we are going to design formalisations that render *EQR* scheme variants as D-ASPIC$^+$ arguments. Being both patterns of non-monotonic reasoning, their connection follows naturally.

6.1.1 D-ASPIC$^+$ *EQR claim* scheme

Let us examine once again the intuitive diagrammatic representation of the *EQR claim* scheme that we are going to call *Arg*:

$$R \xrightarrow{\alpha_{\langle E,F\rangle}} S \models A \uparrow v \qquad (Arg)$$

Arg is composed of *states* (i.e., R,S), logical formulae (i.e., α, E, F, A, v) and defeasible rules (which can be approximately identified with the symbols $\longrightarrow$, $\models$, $\uparrow$ along with the corresponding antecedents and consequents). Observe also that we can regard *states* as sets of logical formulae[i] (that is to say, collections of logical formulae mapped to the truth-value assignment that each state is supposed to encode). These preliminary remarks allow us to move to the formal definition:

Definition 50 (D-ASPIC$^+$ *claim* scheme)**.** *Let Arg be the* EQR claim *scheme. Assume that states R and S are sets of logical formulae. Also, let α, E, F, A, and v, be logical formulae such that α is defeasibly derived from E and F. Similarly, let S be defeasibly derived from R and α, A from S and v from A. Then, the D-ASPIC$^+$* EQR claim *scheme (henceforth D-claim scheme) is composed of the following elements:*

- $\mathscr{R}_s = \emptyset$
- $\mathscr{R}_d = \{(R,\alpha,S,A,v) \rightsquigarrow A; (E,F) \rightsquigarrow \alpha; (R,\alpha) \rightsquigarrow S; S \rightsquigarrow A; A \rightsquigarrow v\}$
- $n((R,\alpha,S,A,v) \rightsquigarrow A) = d_0$, $n((E,F) \rightsquigarrow \alpha) = d_1$, $n((R,\alpha) \rightsquigarrow S) = d_2$, $n(S \rightsquigarrow A) = d_3$, $n(A \rightsquigarrow v) = d_4$
- $\mathscr{K}_n = \emptyset$
- $\mathscr{K}_p = \{R, E, F\}$

Where $Arg1, \ldots, Arg7$ equals to:

- $Arg1 = (\{R\}, \emptyset, R)$;

[i]Notice that if the underlying language that instantiates the *EQR claim* scheme corresponds to classical logic, then we are abusing the notation of Dialectical Cl-Arg. Indeed, the sets of classical logical formulae should be denoted by uppercase greek letters (that is the standard we have followed in the previous chapters involving Dialectical Cl-Arg arguments).

- $Arg2 = (\{E\}, \emptyset, E)$;

- $Arg3 = (\{F\}, \emptyset, F)$;

- $Arg4 = (\{Arg2, Arg3, d_1\}, \emptyset, \alpha)$;

- $Arg5 = (\{Arg1, Arg4, d_2\}, \emptyset, S)$;

- $Arg6 = (\{Arg5, d_3\}, \emptyset, A)$;

- $Arg7 = (\{Arg6, d_4\}, \emptyset, v)$.

Finally, the actual D-claim *scheme will be represented as:*

- $Arg = (\{Arg1, Arg4, Arg5, Arg6, Arg7, d_0\}, \emptyset, A)$

The general structure introduced by the *D-claim* scheme can be seen as a template that, when instantiated by logic formulae, generates a fully-fledged D-ASPIC$^+$ argument instantiation. Multiple D-ASPIC$^+$ arguments can then inform a pdAF that may be evaluated according to a Dung's semantics. As already emphasized by Prakken in [106],

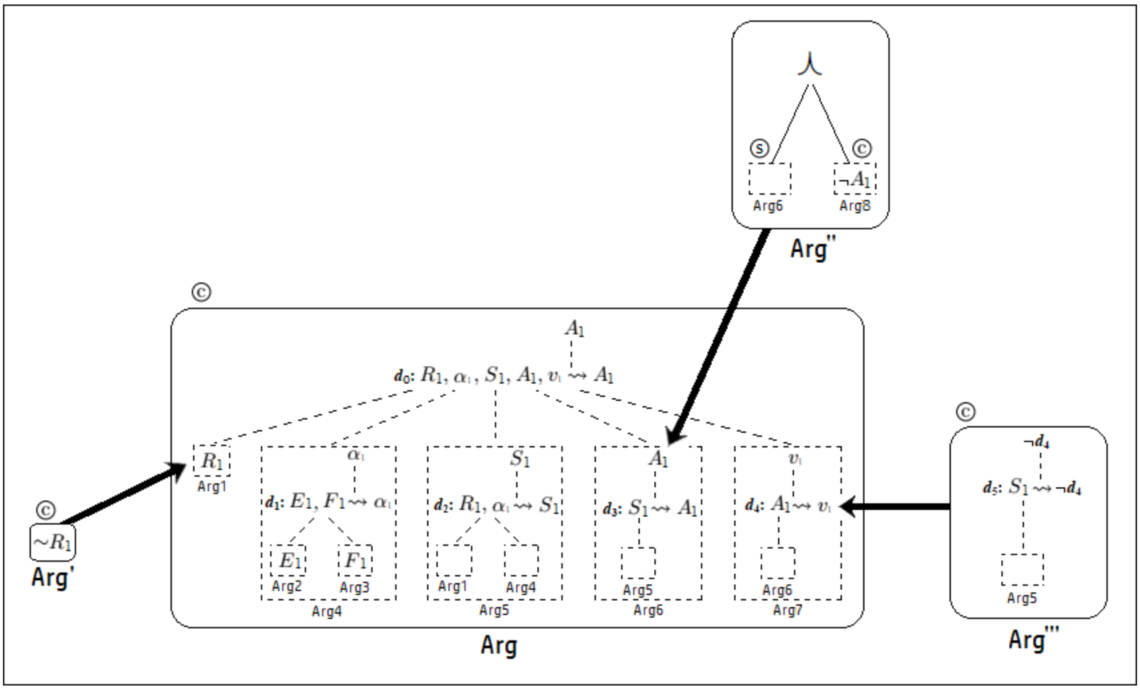

Figure 6.1: *D-claim* scheme instantiation (Arg) and other conflicting D-ASPIC$^+$ arguments. Every such D-ASPIC$^+$ argument (Arg, Arg', Arg'', Arg''', i.e., the solid line boxes) and their respective sub-arguments ($Arg1, \ldots, Arg8$, i.e., the dashed line boxes) are depicted as upside-down trees, whose leaves are premises, yielding the arguments' claim (i.e., the root node) via application of defeasible (dashed lines) and strict (solid lines) rules. The straight arrows identify attacks between arguments highlighting the specific target.

every critical question introduced in [139] corresponds to precise ASPIC types of attacks (especially undercut and rebut). Here we can move a step forward and provide the corresponding ASPIC-attack for each possible attack available in the *EQR claim* dialogue protocols (Table 6.1). Notice that such an attack might fall under more than one category (i.e., undercut, rebut or undermine). In a similar situation, the classification will be rendered according to the first element targeted. In general, if it is a defeasible rule, a conclusion of a defeasible rule or a premise of the *D-claim* scheme, the attack will be categorized as, respectively, undercut, rebut or undermine. Such classification will prove useful also when dealing with defeats. Indeed, exactly as in ASPIC$^+$ [98], even undercuts in D-ASPIC$^+$ are preference independent, hence they always succeed as defeats. That is to say, half of the available attacks of the *EQR claim* dialogue protocol (represented as D-ASPIC$^+$ argument instantiations of the *D-claim* scheme and its respective critical questions) automatically succeed as defeats.

EQR *claim* dialogue attacks	**Undercut**	**Rebut**	**Undermine**
$1a, 1b$			✓
$2a, 2b$	✓		
$3a, 3b, 3c$	✓		
4	✓		
5	✓		
$6a, 6b, 6c$			✓
$7a, 7b$			✓
$8a, 8b, 8c, 8d$	✓		
$9a, 9b$		✓	
$10a, 10b$	✓		
$11a, 11b$		✓	
$12a, 12b$	✓		
$13a, 13b$		✓	
$14a, 14b$	✓		
$15a, 15b$		✓	

Table 6.1: Classification of *EQR claim* dialogue attacks as D-ASPIC$^+$ attacks.

Example 14. *Let* $Arg = (\{Arg1, Arg4, Arg5, Arg6, Arg7, d_0\}, \emptyset, A_1)$*, along with its sub-arguments* $Arg1$-$Arg7$*, be the D-ASPIC*$^+$ *arguments obtained by instantiating the* D-claim *scheme of Definition 50. Let* $Arg' = (\{\sim R_1\}, \emptyset, \sim R_1)$, $Arg'' = (\{\neg A_1\}, \{Arg6\}, \curlywedge)$, $Arg''' = (\{Arg5, S_1 \rightsquigarrow \neg d_4\}, \emptyset, \neg d_4)$ *be D-ASPIC*$^+$ *arguments too. Let also* $\mathcal{R}_s = \vdash_{CL}$ *(i.e., the available strict rules correspond to the classical logic entailment),* $\mathcal{R}_d = \{(R_1, \alpha_1, S_1, A_1, v_1) \rightsquigarrow A_1; (E_1, F_1) \rightsquigarrow \alpha_1; (R_1, \alpha_1) \rightsquigarrow S_1; S_1 \rightsquigarrow A_1; A_1 \rightsquigarrow v_1; S_1 \rightsquigarrow \neg d_4\}, n((R_1, \alpha_1, S_1, A_1, v_1) \rightsquigarrow A_1) = d_0,\ n((E_1, F_1) \rightsquigarrow \alpha_1) = d_1, n((R_1, \alpha_1) \rightsquigarrow S_1) = d_2, n(S_1 \rightsquigarrow A_1) = d_3, n(A_1 \rightsquigarrow v_1) = d_4, n(S_1 \rightsquigarrow \neg d_4) = d_5$*,* $\mathcal{K}_n = \emptyset$ *and* $\mathcal{K}_p = \{E_1, F_1, R_1, \sim R_1, \neg A_1\}$*. Figure 6.1 represents the defined arguments and the attacks moved against each other. That* *is to say, Arg' contrary undermines Arg on $R_1 \in \mathsf{prem}_p(Arg)$ conveying **CQ6** (either via **Attack 6a** or **Attack 6b**). Arg''', instead, undercuts Arg by attacking the (top) defeasible rule d_4 of argument $Arg7$ (via **Attack 14a**). Finally, Arg'' rebuts Arg on $Arg6 \in \mathsf{sub}(Arg)$ conclusion, i.e., A_1 (via **Attack 13a**). Indeed, Arg'' shows how supposing $Arg6$ and committing to $\neg A_1$ (i.e., $Arg8 = (\{\neg A_1\}, \emptyset, \neg A_1)$) will lead to a contradiction (rendered by $\curlywedge$).*

6.1.2 D-ASPIC$^+$ *EQR endorsement* scheme

Following the same reasoning of the previous section, we can translate the *EQR endorsement* scheme into a D-ASPIC$^+$ argument. Let us first recall the *EQR endorsement* scheme's diagrammatic representation:

$$R \xrightarrow{\texttt{assert}(\alpha_{\langle E,F\rangle})} S \models A \uparrow v \qquad (Arg^*)$$

Observe that the logical structure of Arg^* and the one of the *EQR claim* scheme (i.e., *Arg*) are almost identical[ii]. Their only differences concern: (i) the alternative meaning given to the defeasible rule $R \xrightarrow{\texttt{assert}(\alpha_{\langle E,F\rangle})} S$, which now denotes the endorsement of an assertion rather than a simple assertion; (ii) the additional label *assert* that specifies the illocutionary act of publicly uttering the expert's opinion α.

Definition 51 (D-ASPIC$^+$ endorsement scheme)**.** *Let the D-ASPIC$^+$* EQR endorsement *scheme (henceforth D-endorsement scheme) be formally defined as in Definition 50 with the following exception:* $\texttt{assert}\left(\alpha_{\langle E,F\rangle}\right)$ *differs from $\alpha_{\langle E,F\rangle}$ only for the additional specification of the illocutionary act* `assert` *undertaken when α is defeasibly entailed by E and F.*

That is to say, from the formal point of view only, the inner structures of *D-endorsement* and *D-claim* schemes are equivalent. However, we can find a difference in the classification of *EQR endorsement* dialogues attacks as D-ASPIC$^+$ attacks (Table 6.2). In addition, despite the identical organization and name, recall that these attacks derive from modified versions of the critical questions introduced for the *EQR claim* scheme (as such, they semantically diverge from the ones presented in Table 6.1).

[ii]For this reason, we are going to omit the graphical representation of Arg^*.

EQR *endorsement* dialogue attacks	Undercut	Rebut	Undermine
$1a, 1b$			✓
$2a, 2b$	✓		
$3a, 3b, 3c$	✓		
4	✓		
5	✓		
$6a, 6b, 6c$			✓
$7a, 7b$			✓
$8a, 8b, 8c, 8d$	✓		
$9a, 9b$		✓	
$10a, 10b$	✓		
$11a, 11b$	✓		
$12a, 12b$		✓	
$13a, 13b$	✓		
$14a, 14b$		✓	
$15a, 15b$	✓		
$16a, 16b$		✓	

Table 6.2: Classification of *EQR endorsement* dialogue attacks as D-ASPIC$^+$ attacks.

6.1.3 D-ASPIC$^+$ *EQR endorsed-by-whom* scheme

Similarly to the *D-claim* and the *D-endorsement* schemes, we can also translate the *EQR endorsed-by-whom* scheme into a D-ASPIC$^+$ argument.

$$R \xrightarrow{\texttt{assert}(\alpha_{\langle E,F \rangle}),\, X} S \models A \uparrow v \qquad (Arg^{\dagger})$$

We can observe a correspondence between the logical structure of $Arg^{\dagger}$ and the one of the *EQR endorsement* scheme (i.e., Arg^{*}), which, indeed, are almost identical. Their only difference concerns the additional element X denoting a group of endorsers that upholds the endorsement of the assertion of the expert's opinion $\texttt{assert}(\alpha_{\langle E,F \rangle})$.

Definition 52 (D-ASPIC$^+$ *endorsed-by-whom scheme*)**.** *Let D-ASPIC$^+$* EQR endorsed-by-whom *scheme* (*henceforth D-endorsed-by-whom scheme*) *be formally defined as in Definition 51 with the supplement of the element X. This modifies the* D-endorsement *scheme as follows:*

- $\mathscr{R}_s = \emptyset$,
- $\mathscr{R}_d = \{(R, \alpha, X, S, A, v) \rightsquigarrow A; (E, F) \rightsquigarrow \alpha; (R, \alpha, X) \rightsquigarrow S; S \rightsquigarrow A;$
  $A \rightsquigarrow v\}$
- $n((R, \alpha, X, S, A, v) \rightsquigarrow A) = d_0$, $n((E, F) \rightsquigarrow \alpha) = d_1$, $n((R, \alpha, X) \rightsquigarrow S) = d_2$, $n(S \rightsquigarrow A) = d_3$, $n(A \rightsquigarrow v) = d_4$

- $\mathscr{K}_n = \emptyset$
- $\mathscr{K}_p = \{R, E, F, X\}$

Where $Arg^{\dagger}1, \ldots, Arg^{\dagger}8$ *equals to:*

- $Arg^{\dagger}1 = (\{R\}, \emptyset, R)$;
- $Arg^{\dagger}2 = (\{E\}, \emptyset, E)$;
- $Arg^{\dagger}3 = (\{F\}, \emptyset, F)$;
- $Arg^{\dagger}4 = (\{X\}, \emptyset, X)$;
- $Arg^{\dagger}5 = (\{Arg^{\dagger}2, Arg^{\dagger}3, d_1\}, \emptyset, \alpha)$;
- $Arg^{\dagger}6 = (\{Arg^{\dagger}1, Arg^{\dagger}5, Arg^{\dagger}4, d_2\}, \emptyset, S)$;
- $Arg^{\dagger}7 = (\{Arg^{\dagger}6, d_3\}, \emptyset, A)$;
- $Arg^{\dagger}8 = (\{Arg^{\dagger}7, d_4\}, \emptyset, v)$.

Finally, the actual D-endorsed-by-whom *scheme will be represented as:*

- $Arg^{\dagger} = \left(\{Arg^{\dagger}1, Arg^{\dagger}5, Arg^{\dagger}4, Arg^{\dagger}6, Arg^{\dagger}7, Arg^{\dagger}8, d_0\}, \emptyset, A\right)$

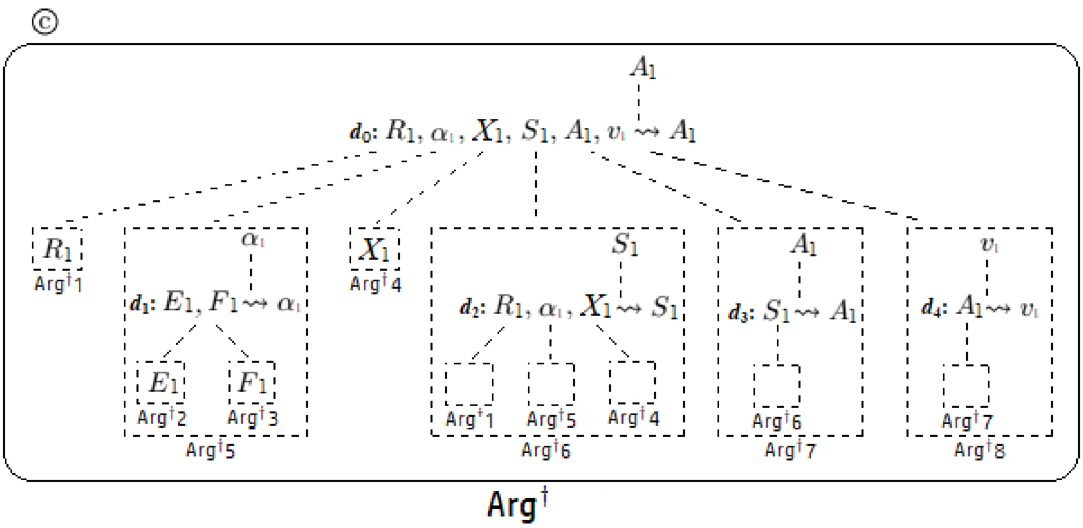

Figure 6.2: *D-endorsed-by-whom scheme* instantiation.

As for the previous *D*-scheme variants, we can detect and organize the *EQR endorsed-by-whom* dialogues attacks as D-ASPIC$^+$ attacks (Table 6.3). Once again, consider that the critical questions upon which these attacks are based slightly differ from the ones introduced for the other *EQR* scheme variants.

EQR *endorsed-by-whom* dialogue attacks	Undercut	Rebut	Undermine
$1a, 1b$			✓
$2a, 2b$	✓		
$3a, 3b, 3c$	✓		
4	✓		
5	✓		
$6a, 6b, 6c$			✓
$7a, 7b$			✓
$8a, 8b, 8c, 8d$	✓		
$9a, 9b$		✓	
$10a, 10b$	✓		
$11a, 11b$			✓
$12a, 12b$	✓		
$13a, 13b$		✓	
$14a, 14b$	✓		
$15a, 15b$		✓	
$16a, 16b$	✓		
$17a, 17b$		✓	

Table 6.3: Classification of *EQR endorsed-by-whom* dialogue attacks as D-ASPIC$^+$ attacks.

6.2 Implementing the dialogue

Dialogues concerning *D-scheme* instantiations and their critical questions (rendered as D-ASPIC$^+$ arguments too) have a tighter bond with the structured argumentation formalism than standard EQR dialogue variants. This means that if we can find a way for simplifying the EQR dialogue locutions into the more basic attacks/defeats relation between arguments, we will be able to easily accommodate an evaluation according to one of Dung's semantics. However, we first have to identify a way of determining the overall preference ordering among arguments and how such ordering is rendered in the dialogue. Finally, we also have to account for the set $\mathscr{S}$ of arguments parametrized by the dialectical defeats and pinpoint its corresponding element in the EQR dialogue structure employing *D-scheme* instantiations.

6.2.1 Preference ordering

Dialectical defeats provide a way to establish whether an attack succeeds or fails in its attempt to challenge the targeted argument. To do so, the defeats make use of a preference ordering over the D-ASPIC$^+$ arguments. Before starting the actual dialogue, the idea would then be to stipulate such preference ordering (i.e., $\prec$) by requesting the agents to perform a preliminary inquiry dialogue[iii] on this matter. Notice that, although the em-

[iii]Consider that this type of dialogue can benefit from the employment of *metalevel* arguments that allow expressing the rationale for preferences and reason about conflicting *object-level* arguments. Indeed,

ployment of Value-based argumentation framework (VAF)[iv] seems to be the best choice to rely on this circumstance, the lack of arguments with a specific value element (indeed, not every D-ASPIC$^+$ argument moved in the dialogue is a *D-scheme* instantiation) renders less convenient the use of VAF. That is because we would still need to perform an inquiry dialogue to determine the overall mapping between the arguments conveyed by the EQR dialogue's locutions and the existing values.

6.2.2 Parametrizing the set $\mathscr{S}$

Dialectical defeats parametrize a set $\mathscr{S}$ from which committed sub-arguments can be supposed to draw specific conclusions. However, we still have to establish the relationship existing between $\mathscr{S}$ and the EQR dialogue that employs D-ASPIC$^+$ arguments.

Intuitively, the proponent of the dialogue topic *Arg* (represented by the first *D-scheme* instantiation played and meant to be tested) must defend from all the attacks/defeats moved against it (including the critical questions converted into D-ASPIC$^+$ arguments) in order to assess its validity. In doing so, the proponent will probably deploy different arguments which, although capable of defeating *Arg*'s defeaters, might require to be defended as well. This leads to the construction of a set of committed PRO's arguments revolving around *Arg*, which seems to be the appropriate candidate to identify $\mathscr{S}$. Indeed, there should be no other extension representing a suitable candidate. The opponent's primary task is to invalidate the initial *D-scheme* moved by the proponent: it has no interest in suggesting its own view on the topic or supporting different expert's opinions[v], which might generate an alternative set $\mathscr{S}$. Unlike the proponent's team, it could be said that the opponent's team does not have the burden of defending the arguments it commits to.

Before introducing a formal definition for $\mathscr{S}$ in the *EQR* dialogue, let us first present the notation we are going to use throughout the remainder of the chapter.

Metalevel Argumentation Frameworks (MAF) can, for example, be used to formalise Preference-based argumentation (PAF) [96].

[iv]Value-based argumentation framework (VAF) (extended as Audience-specific value-based argumentation framework (AVAF)) acknowledges several possible classifications of the given values, identified as 'audiences'. Each abstract argument is mapped to a value and every defeat, against which the semantics of the frameworks is established, parametrized a single audience [14].

[v]In the EQR dialogue, the explainee (OPP) party's purpose is to retrieve and assess the required information rather than moving counter proposals, although this might happen when trying to invalidate the explainer party (PRO) move. In this case, however, it would be just a consequence of an attack, not the opponent's main goal. To focus on proposing its own view, the explainee party should perform a dialectical shift towards another dialogue protocol, hence abandoning (or pausing) the current EQR dialogue.

Remark 4 (Notation)**.** *In the next sections, we are going to employ the following notation. The dialogue locutions and their conveyed* (*D-ASPIC*$^+$) *arguments will be formally denoted by Pl*[***locution***]*:*argument*, where Pl represents the team of players* (*either P or O, abbreviations for PRO and OPP*) *that utters them.*
(Pl_n)`turn` *identifies, instead, the set of locutions and conveyed* (*D-ASPIC*$^+$) *arguments uttered by either team as the* (*last*) *nth-move of the unfolding dialogue*[vi]*. In addition, we are going to use* $\mathscr{A}, \mathscr{A}', \mathscr{A}''$*, etc. as variables denoting arbitrary D-ASPIC*$^+$ *arguments.*

Finally, $\mathscr{A}^\star = \mathscr{A}_1\text{-}\mathscr{A}_2\text{-}\cdots\text{-}\mathscr{A}_i$ *designates a sequence of alternating PRO and OPP arguments* (*each of which is part of a locution, i.e., Pl*[***locution***]*:*argument) *ending with argument* $\mathscr{A}_i$ (*for* $i > 0$)*. The sequence is such that* $\mathscr{A}_1\text{-}\mathscr{A}_2$ *identifies the defeat occurring from argument* $\mathscr{A}_1$*, moved by one team, to* $\mathscr{A}_2$*, moved by the other team. Writing* $\mathscr{A}_1 \Rightarrow_{\mathscr{S}} \mathscr{A}^\star$ *means that* $\mathscr{A}_1$ *indirectly dialectically defeats* $\mathscr{A}_i$*, i.e., the last argument of the sequence* $\mathscr{A}^\star$*.*

Definition 53 (Set $\mathscr{S}_n$)**.** *PRO commits to each locution and conveyed D-ASPIC*$^+$ *arguments* (*e.g.,* $\mathscr{A}$) *it moves during the dialogue.* $\mathscr{S}_n$ *identifies an extension that includes all of such arguments up to the nth-move of the unfolding dialogue. The only exception will be represented by the undefended arguments, that occur in the same sequence, before PRO utterance of a* **concede** *locution. Formally:*

$\mathscr{S}_n = \{\mathscr{A} \mid$

1) $\mathscr{A}$ *is an argument moved by PRO;*
2) *If P*[***concede***]*:*$\mathscr{A}' \in (P_k)$`turn` (*with* $k \leq n$)*, for every* $\mathscr{A}''$ *moved by PRO before* (P_k)`turn` *such that* $\mathscr{A}' \Rightarrow_{\mathscr{S}_j} \mathscr{A}''$ *or* $\mathscr{A}' \Rightarrow_{\mathscr{S}_j} \mathscr{A}^\star$ (*with* $j < k$ *and* $\mathscr{A}^\star = \mathscr{A}'\text{-}\cdots\text{-}\mathscr{A}''$) *and* $\nexists \mathscr{A}'''$ *moved by PRO such that* $\mathscr{A}''' \Rightarrow_{\{\mathscr{A}'\}} \mathscr{A}'$*, then* $\mathscr{A} \neq \mathscr{A}''\}$

That is to say if PRO **concede**s an argument $\mathscr{A}'$ (as its $k \leq n$ move), then all of the other PRO's arguments (presented before its kth-move) that are dialectically (directly or indirectly) defeated by $\mathscr{A}'$ (with respect to $\mathscr{S}_j$, where $j < k \leq n$) and not defended, will not be part of the set $\mathscr{S}_n$. Consider also that the elements of $\mathscr{S}_n$ might change after the utterance of specific locutions, leading $\mathscr{S}_n$ to be updated throughout the dialogue. However, unlike in the dialectical argument games of Chapter 3, there will be no issue concerning 'disqualified defeats'. With these terms we mean the dialectical defeats that will be invalidated due to the updates of the specific parametrized set $\mathscr{S}$. Since in an *EQR*

[vi]That is to say, (P_n)`turn` and (O_n)`turn` identify the set of locutions and conveyed (D-ASPIC$^+$) arguments moved by PRO, respectively OPP, as the (last) nth-move of the unfolding dialogue.

dialogue, it is PRO that determines the members of the set $\mathscr{S}_n$ by uttering the locution **concede**, then changes in the dialectical defeats moved by OPP will not affect $\mathscr{S}_n$ and its arguments. Observe indeed that a series of locutions can terminate (i.e., there are no more available locutions to utter) either after a (i) PRO or an (ii) OPP move. In the latter case (ii), PRO has uttered **concede**, thus meaning that it 'surrenders' to its counterpart locution and conveyed argument. This yields that, independently from any potentially disqualified defeats of OPP arguments, PRO has resigned its arguments in that dialogue's line of defence (unless a future move will manage to defend some of them iff they are part of a new line of defence). On the other hand (i), if it is the proponent team that utters the last locution in a series, then, regardless of the presence of any disqualified defeats of OPP arguments, PRO will still 'win' that dialogue's line of defence.

6.2.3 Dialogue formal protocol

In order to provide a formalisation of the EQR dialogue protocol that employs *D-scheme* instantiations, we will have to first establish a form of conversion between the interactions among locutions of the standard EQR dialogue and the basic attacks/defeats relation of D-ASPIC$^+$.

ask This locution aims at asking the ground on which it is the case that 'something' (say a) and not otherwise. As such, it can be seen as an argument attacking/defeating another argument on a.

state Except with PRO's first turn (in which **state** corresponds just with the instantiation of the initial *D-scheme*), the purpose of the locution is to answer the query moved by the **ask** locution. This will establish the reason why the questioned 'something' is the case. **state** can then be seen as identifying such rationale via an argument attacking/defeating the argument that posed the previous question.

deny Intuitively, this locution denotes a refutation against the 'something' (say a) it is addressed. As such, it can be straightforwardly seen as an argument attacking/defeating another argument on a. The same reasoning can be applied to **deny initial** (which is merely a locution *deny* that directly targets the initial *D-scheme* moved by PRO).

The role of locutions such as **enter/leave dialogue**, **turn start/finished**, **concede**, **reject** and **change player** is, instead, to administer the dialogue in a more 'structural' way that is not directly connected to arguments attacks/defeats. Their purpose is to identify specific phases of each turn and different stages of the overall dialogue.

Example 15. *To clarify, let us consider a brief conversation as an example (avoiding D-scheme instantiations for simplicity) and label each utterance with the corresponding location according to the notation of Remark 4.*

*P[**state**]:"It is going to rain soon"*

*O[**ask**]:"It is actually sunny, why do you say so?"*

*P[**state**]:"I heard it on the forecast of Channel 7"*

*O[**deny**]:"Channel 7 is a very unreliable source of information"*

*P[**ask**]:"Even for the weather forecast?"*

*O[**state**]:"Especially for the weather forecast..."*

*Suppose the proponent decides to believe the opponent's assertion about Channel 7 (hence uttering the **concede** locution). Given that PRO has now resigned from her initial argument, OPP can be deemed to have successfully defeated its counterpart's argument, thus winning this particular dispute.*

We now have all the elements to formally introduce the *unique* dialogue protocol. Indeed, there is only one protocol that includes *D-claim, D-endorsement* and *D-endorsed-by-whom* schemes[vii]. That is because the introduced formalism affects all the EQR dialogue versions in the same way:

Definition 54 (Protocol)**.** *Let Arg be the* D-scheme *instantiation that needs to be tested and $\mathscr{S}_n$ the set characterized as in Definition 53. Let also make use of the same notation as described in Remark 4. Then, an EQR dialogue that employs* D-scheme *instantiations and their critical questions (rendered as D-ASPIC$^+$ arguments) unfolds in the following way:*

(54.0) *PRO moves the first set of locutions, i.e., (P_1)*`turn`*:*

- (a) *P[**enter dialogue**], P[**turn start**]*
- (b) *P[**state**]:Arg*
- (c) *P[**turn finished**], P[**change player**]*

(54.1) *OPP moves the second set of locutions, i.e., (O_2)*`turn`*:*

[vii] Although each *D-scheme* will still maintain their specific CQs-informed attacks.

(a) O[***enter dialogue***], O[***turn start***]

(b) O[***deny initial***]:$\mathscr{A}'$ *or* O[***ask***]:$\mathscr{A}'$*, where* $\mathscr{A}' \Rightarrow_{\mathscr{S}_1} Arg$

(c) O[***turn finished***], O[***change player***]

(54.2) *If* (Pl_n)turn *and* $n = 2k+1$ *(for* $k > 0$*), then it is PRO's turn to move, i.e.,* (P_n)turn*. Otherwise, if* (Pl_n)turn *and* $n = 2k$ *(for* $k > 1$*), then it is OPP's turn to move, i.e.,* (O_n)turn*. That is to say, PRO moves on odd turns, while OPP moves on even turns.*

(54.3) *A generic PRO's turn, i.e.,* (P_n)turn*, is such that (in order):*

(a) P[***turn start***]

(b) *PRO chooses one among:*

* P[***concede***]:$\mathscr{A}'$*, if* $\mathscr{A}' \in (O_{n-1})$turn *and corresponds either to* O[***deny***]:$\mathscr{A}'$, O[***ask***]:$\mathscr{A}'$*, or* O[***state***]:$\mathscr{A}'$
* P[***reject***]:$\mathscr{A}'$*, if* $\mathscr{A}' \in (O_{n-1})$turn *and corresponds either to* O[***deny***]: $\mathscr{A}'$ *or* O[***state***]:$\mathscr{A}'$*. Alternatively,* O[***deny initial***]:$\mathscr{A}'$ *if* $\mathscr{A}' \Rightarrow_{\mathscr{S}_{n-2}} Arg$
* P[***state***]:$\mathscr{A}''$*, if* $\mathscr{A}'' \Rightarrow_{\{\mathscr{A}'\}} \mathscr{A}'$, $\mathscr{A}' \in (O_{n-1})$turn *and corresponds to* O[***ask***]:$\mathscr{A}'$

(c) *According to the choice of point (54.3(b)), PRO selects one among:*

◇ P[***deny***]:$\mathscr{A}''$*, or* P[***ask***]:$\mathscr{A}''$*, where* $\mathscr{A}'' \Rightarrow_{\{\mathscr{A}'''\}} \mathscr{A}'''$*, and* $\mathscr{A}''' \in (O_i)$turn *(for* $i < n$*) if uttered after a* P[***concede***]:$\mathscr{A}'$*, such that* $\mathscr{A}''' \neq \mathscr{A}'$

◇ P[***deny***]:$\mathscr{A}''$*, or* P[***ask***]:$\mathscr{A}''$*, where* $\mathscr{A}'' \Rightarrow_{\{\mathscr{A}'\}} \mathscr{A}'$ *if uttered after a* P[***reject***]:$\mathscr{A}'$

(d) $Arg \in \mathscr{S}_n$*, update* $\mathscr{S}_n$

(e) P[***turn finished***], P[***change player***]

(54.4) *A generic OPP's turn, i.e.,* (O_n)turn*, is such that (in order):*

(a) O[***turn start***]

(b) *OPP chooses one among:*

* O[***concede***]:$\mathscr{A}'$*, if* $\mathscr{A}' \in (P_{n-1})$turn*,* $\mathscr{A}' \neq Arg$ *and corresponds either to* P[***deny***]:$\mathscr{A}'$, P[***ask***]:$\mathscr{A}'$*, or* P[***state***]:$\mathscr{A}'$
* O[***reject***]:$\mathscr{A}'$*, if* $\mathscr{A}' \in (P_{n-1})$turn *and corresponds either to* P[***deny***]: $\mathscr{A}'$ *or* P[***state***]:$\mathscr{A}'$
* O[***state***]:$\mathscr{A}''$*, if* $\mathscr{A}'' \Rightarrow_{\mathscr{S}_{n-1}} \mathscr{A}'$, $\mathscr{A}' \in \mathscr{S}_{n-1}$, $\mathscr{A}' \in (P_{n-1})$turn *and corresponds to* P[***ask***]:$\mathscr{A}'$

(c) *According to the choice of point (54.4(b)), OPP selects one among:*

- ◇ $O[\textbf{deny}]{:}\mathscr{A}''$, *or* $O[\textbf{ask}]{:}\mathscr{A}''$, *where* $\mathscr{A}'' \Rightarrow_{\mathscr{S}_{n-1}} \mathscr{A}'''$, *and* $\mathscr{A}''' \in \mathscr{S}_{n-1}$ *if uttered after a* $O[\textbf{concede}]{:}\mathscr{A}'$
- ◇ $O[\textbf{deny}]{:}\mathscr{A}''$, *or* $O[\textbf{ask}]{:}\mathscr{A}''$, *where* $\mathscr{A}'' \Rightarrow_{\mathscr{S}_{n-1}} \mathscr{A}'$, *and* $\mathscr{A}' \in \mathscr{S}_{n-1}$ *if uttered after a* $O[\textbf{reject}]{:}\mathscr{A}'$
- ◇ $O[\textbf{deny initial}]{:}\mathscr{A}''$ *such that* $\mathscr{A}'' \Rightarrow_{\mathscr{S}_{n-1}} Arg$ *if uttered after a* $O[\textbf{concede}]{:}\mathscr{A}'$

(d) *Every argument* $\mathscr{A}''$ *of points (54.4(b)) and (54.4(c)) is such that it has not already been moved and defeated (and not defended) in* (O_i)`turn` *(for* $i < n$*)*

(e) $O[\textbf{turn finished}]$, $O[\textbf{change player}]$

(54.5) *The turn of the first team having no more locutions [**deny**], [**deny initial**], [**ask**] or [**state**] available, i.e.,* (Pl_n)`turn`*, is such that it overwrites any other previous move requirements and (in order):*

(a) $Pl[\textbf{turn start}]$

(b) $Pl[\textbf{concede}]{:}\mathscr{A}'$ *or* $Pl[\textbf{reject}]{:}\mathscr{A}'$, *where* $\mathscr{A}' \in (Pl_{n-1})$`turn` *and corresponds either to:* $Pl[\textbf{deny}]{:}\mathscr{A}'$, $Pl[\textbf{state}]{:}\mathscr{A}'$ *or* $Pl[\textbf{ask}]{:}\mathscr{A}'$ *(alternatively, it corresponds to* $O[\textbf{deny initial}]{:}\mathscr{A}'$, *if* (Pl_{n-1})`turn` $= (P_{n-1})$`turn`*)*

(c) $Pl[\textbf{leave dialogue}]$

Definition 54 formalises the moves available to each team of players during an *EQR dialogue* that employs D-ASPIC$^+$ arguments. PRO starts the game and utters a specific set of locutions [(54.0)], after which it will be OPP's turn to move [(54.1)]. Then, the two teams of participants will alternate, uttering ordered lists of locutions in accordance with the previous moves and phases of the dialogue [(54.3), (54.4)]. Notice that both PRO and OPP will have to abide by the respective *relevance conditions*. These are rules that force the teams to change the (temporary) outcome of the game to their advantage. That is to say, at the end of its turn, PRO must have (directly or indirectly) defended the initial *D-scheme Arg* [(54.3(d))]. On the other hand, at the end of its turn, OPP must have (directly or indirectly) challenged the validity of *Arg* (i.e., the acceptability of *Arg* with respect to $\mathscr{S}_n$)[(54.4(b)), (54.4(c))]. Indeed, it suffices to defeat any argument in $\mathscr{S}_n$ to (directly o indirectly) challenge *Arg*, hence satisfying OPP's relevance condition. Finally, consider also the importance of the *non-repetition rule* [(54.4(d))]: without this constraint, OPP will be allowed to repeat the same attacks that have already been countered by PRO, generating, in this way, an infinite dialogue.

The conclusion of the dialogue will occur whenever one of the two teams will terminate its available move [(54.5)]. At this point, PRO will be proclaimed the winner if OPP leaves the dialogue first after having uttered the **concede** locution. In the opposite circumstance, OPP will win the dialogue instead. Otherwise, the result will be a draw. Recall that a victory means being able to assess the validity of *Arg* (for PRO) or to dialectically prove its invalidity (for OPP).

6.2.4 An example of an EQR dialogue involving a D-claim scheme instantiation

In this section, we are going to devise an implementation of an EQR dialogue that begins with a *D-claim* scheme and employs D-ASPIC$^+$ arguments. Such an example will combine both formal (the introduced locutions) and informal (natural language) elements providing a clear illustration of the protocol presented in Definition 54.

Example 16. *The dialogue we are going to consider is a model that seeks to assess the validity of the following initial (instantiated)* D-claim *scheme:*

(Arg) "Since we are in a pandemic emergency $[R_1]$
the epidemiologist $[F_1]$ Dr.Stone $[E_1]$
suggests that we should all be wearing face masks $[\alpha_1]$.
This will result in a less critical situation $[S_1]$,
where the virus spreads slower $[A_1]$,
and people's health is preserved $[v_1]$.

Therefore, we should assume that (acting upon the suggestion) the virus
will spread slower $[A_1]$ (thus preserving people's health)."

We are also assuming that the participating agents have preemptively agreed on the general argumentation system ASY. *Moreover, a preliminary inquiry dialogue has been conducted, leading to the identification of the strict partial order $\prec$ over the arguments defined by the argumentation theory $\mathcal{T} = \langle \mathsf{ASY}, \mathcal{K} \rangle$. The knowledge base $\mathcal{K}$ is composed of (for simplicity only) two subsets $\mathcal{KB}_{PRO}$ and $\mathcal{KB}_{OPP}$, describing the knowledge of the two teams of players*[viii]. *Let us consider Figure 6.3. Every box represents a turn of the proponent or the opponent team, i.e., (Pl_n)*`turn`*, where n coincides with the number in*

[viii]More precisely, $\mathcal{KB}_{PRO}$ is the union of all the knowledge bases of each agent that sides with the proponent team. Similarly, $\mathcal{KB}_{OPP}$ is the union of all the knowledge bases of each agent that sides with the opponent team.

*the top left corner of such a box. Notice that, although we have not included control locutions as Pl[**turn start**], Pl[**turn finshed**], Pl[**change player**], Pl[**enter dialogue**] or Pl[**leave dialogue**], this will not affect the efficacy of the example, and it will allow for an easier display.*
The dialogue is based on the following elements:

- $\mathscr{R}_s = \vdash_{CL}$
- $\mathscr{R}_d = \{(R_1, \alpha_1, S_1, A_1, v_1) \rightsquigarrow A_1; (E_1, F_1) \rightsquigarrow \alpha_1; (R_1, \alpha_1) \rightsquigarrow S_1; S_1 \rightsquigarrow A_1; A_1 \rightsquigarrow v_1; r_1 \rightsquigarrow \sim F_1; e_1 \rightsquigarrow \neg r_1; (E_1, F_1) \rightsquigarrow \neg\alpha_1; (E_2, F_1) \rightsquigarrow \alpha_1; E_3 \rightsquigarrow \neg d_7\}$
- $n((R_1, \alpha_1, S_1, A_1, v_1) \rightsquigarrow A_1) = d_0; n((E_1, F_1) \rightsquigarrow \alpha_1) = d_1; n((R_1, \alpha_1) \rightsquigarrow S_1) = d_2; n(S_1 \rightsquigarrow A_1) = d_3; n(A_1 \rightsquigarrow v_1) = d_4; n(r_1 \rightsquigarrow \sim F_1) = d_5; n(e_1 \rightsquigarrow \neg r_1) = d_6; n((E_1, F_1) \rightsquigarrow \neg\alpha_1) = d_7; n((E_2, F_1) \rightsquigarrow \alpha_1) = d_8; n(E_3 \rightsquigarrow \neg d_7) = d_9$
- $\mathscr{K}\mathscr{B}_{PRO} = \{R_1, E_1, E_2, E_3, F_1, e_1\}$
- $\mathscr{K}\mathscr{B}_{OPP} = \{F_1, \neg F_1, E_1, \sim E_2, r_1\}$
- $\mathscr{K}_n = \emptyset$
- $\mathscr{K}_p = \mathscr{K}\mathscr{B}_{PRO} \cup \mathscr{K}\mathscr{B}_{OPP}$

- *Initial* D-claim *scheme:* $\text{Arg} = (\{\text{Arg1}, \text{Arg4}, \text{Arg5}, \text{Arg6}, \text{Arg7}, d_0\}, \emptyset, A_1)$

Where $\text{Arg1}, \ldots, \text{Arg7}$ *equals to:*

- $\text{Arg1} = (\{R_1\}, \emptyset, R_1)$; $\text{Arg2} = (\{E_1\}, \emptyset, E_1)$; $\text{Arg3} = (\{F_1\}, \emptyset, F_1)$;
  $\text{Arg4} = (\{\text{Arg2}, \text{Arg3}, d_1\}, \emptyset, \alpha_1)$; $\text{Arg5} = (\{\text{Arg1}, \text{Arg4}, d_2\}, \emptyset, S_1)$;
  $\text{Arg6} = (\{\text{Arg5}, d_3\}, \emptyset, A_1)$; $\text{Arg7} = (\{\text{Arg6}, d_4\}, \emptyset, v_1)$.

Finally, the remaining involved D-ASPIC$^+$ arguments are as follows:

- *PRO's arguments:* $\mathscr{A}_4 = (\{e_1, d_6\}, \{r_1\}, \curlywedge)$; $\mathscr{A}_6 = (\{E_2, F_1, d_8\}, \emptyset, \alpha_1)$; $\mathscr{A}_8 = (\{E_3, d_9\}, \emptyset, \neg d_7)$
- *OPP's arguments:* $\mathscr{A}_1 = (\{\neg F_1\}, \emptyset, \neg F_1)$; $\mathscr{A}_3 = (\{r_1, d_5\}, \emptyset, \sim F_1)$; $\mathscr{A}_5 = (\{E_1, F_1, d_7\}, \{\text{Arg4}\}, \curlywedge)$; $\mathscr{A}_7 = (\{\sim E_2\}, \emptyset, \sim E_2)$

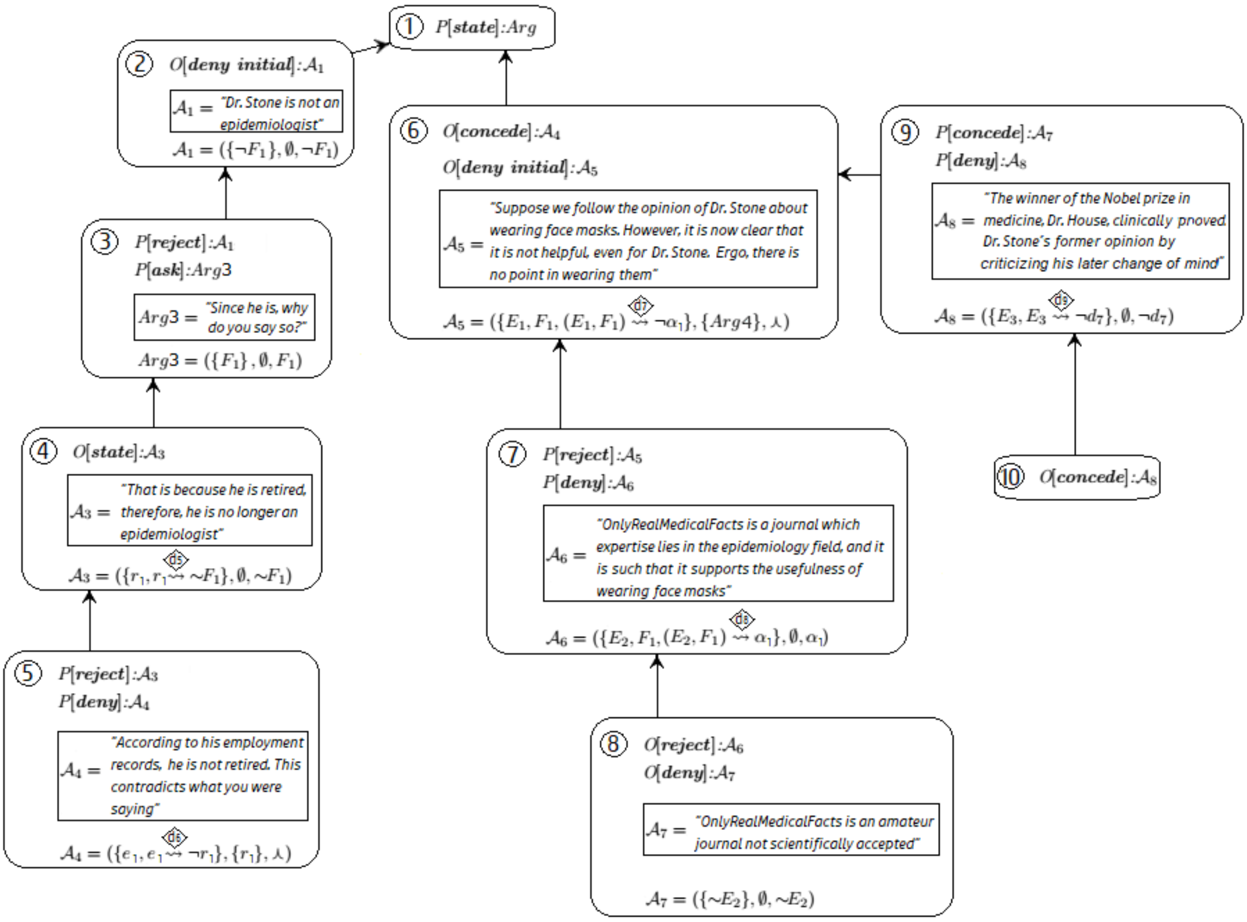

Figure 6.3: Example of a (terminated) *EQR* dialogue involving a D-claim scheme and D-ASPIC$^+$ arguments. The boxes represent PRO and OPP turns, and their order of playing is denoted by the number in their respective top left corners. The arrows serve to highlight dialectical defeats between the connected arguments. As it appears from the figure, since the opponent team concedes PRO's last argument remaining with no further move available, the proponent team wins the dialogue. Therefore, the initial *D-claim* scheme instantiation must be valid.

The dialogue starts with PRO uttering the locution ***state*** *and the conveyed argument Arg. It then unfolds according to the protocol introduced in Definition 54. That is to say, the opponent team challenges Arg by moving O*[***deny initial***]:$\mathscr{A}_1$*, after which the two teams alternate each other in moving locutions to reply to their competitor. We can easily follow the order according to which the dialogue has been developed, thanks to the numbers in the boxes. Consider also that all the argument attacks succeed as defeats* (*denoted by the arrows in Figure 6.3*) *and that such defeats are divided into* undermine ([contradictory undermines]*:* $\mathscr{A}_1 \Rightarrow_{\mathscr{S}_1}$ Arg3*, and so it indirectly defeats also* Arg; Arg3 $\Rightarrow_{\{\mathscr{A}_1\}} \mathscr{A}_1$; [contrary undermines]*:* $\mathscr{A}_3 \Rightarrow_{\mathscr{S}_3}$ Arg3; $\mathscr{A}_7 \Rightarrow_{\mathscr{S}_7} \mathscr{A}_6$), falsum ($\mathscr{A}_4 \Rightarrow_{\{\mathscr{A}_3\}} \mathscr{A}_3$ *where* $Supp(\mathscr{A}_4) = \{r_1\}$; $\mathscr{A}_5 \Rightarrow_{\mathscr{S}_5}$ *Arg where* $Supp(\mathscr{A}_5) = \{$Arg4$\}$), rebut

($\mathscr{A}_6 \Rightarrow_{\{\mathscr{A}_5\}} \mathscr{A}_5$) *and* undercut ($\mathscr{A}_8 \Rightarrow_{\{\mathscr{A}_5\}} \mathscr{A}_5$). *The dialogue terminates with the victory of PRO: since OPP has no more available moves at its disposal, and utters* ***concede****, it must be the first team to leave the dialogue.*

The termination of the dialogue leaves us with a number of arguments (included in a dialectical AF informed by the argumentation theory $\mathscr{T}$) that, along with the existing defeats, can be semantically evaluated. Definition 53 characterized $\mathscr{S}_n$ as a set of committed arguments moved by PRO that might also contain the initial instantiated *D-scheme* (say *Arg*) among its members. We can now show the following formal proof:

Theorem 15 (Soundness and Completeness)**.** *Let* Arg *be an instantiated* D-scheme *included among the arguments of a dialectical AF informed by an argumentation theory* $\mathscr{T}$*. A terminated* EQR *dialogue starting with* Arg *is won or tied by PRO iff* Arg *is a member of an admissible extension* Adm *of the dialectical AF.*

Proof. According to the protocol of Definition 54 (and the winning conditions of chapter 5), if PRO won or tied the dialogue, then *Arg* is a member of the set $\mathscr{S}_n$, where n is the last move that occurs in the (terminated) dialogue. The theorem can be proven if we show the equivalence existing between the two extensions $\mathscr{S}_n$ and Adm.

(⇒) By Definition 53, the set $\mathscr{S}_n$ is such that it defends its arguments, and it is also conflict free (otherwise, OPP would have moved the conflicting argument against $\mathscr{S}_n$, preventing the assumed PRO's victory or draw). However, these are the same features that identify an admissible extension, as stated in Definition 31.

(⇐) Suppose that $Arg \in$ Adm. We can play an EQR dialogue starting with *Arg* and following the protocol of Definition 54. At each challenge moved by OPP, we can respond with a locution that conveys an argument member of Adm. Since Adm is admissible, the dialogue can terminate only with the victory of PRO or a tie between the contenders (depending on the preference of the involved arguments). In either case, *Arg* will be defended and included in a conflict free set $\mathscr{S}_n$ composed of the same arguments of Adm.

Therefore, $\mathscr{S}_n$ and Adm must be equivalent. □

6.2.5 A comparison with dialectical admissible argument games

The formal protocol devised in Definition 54 discloses several similarities between the monological architecture of the dialectical argument games and the structure encompassed by the EQR dialogues employing *D-scheme* and D-ASPIC$^+$ arguments. Indeed, by examining them thoroughly we can detect that (i) although there might be multiple agents participating in an EQR dialogue, they are considered as two teams of players joined together for the same purpose. That is to say, as we have seen, we can deem the dialogue as being played only by a proponent (PRO) and an opponent (OPP), similarly to the dialectical argument games. (ii) The proponent of the dialogue topic (topic represented by the first *D-scheme* instantiation played and meant to be tested) must defend from all the attacks/defeats moved against it (including the critical questions converted into D-ASPIC$^+$ arguments), much like what PRO has to accomplish in a dialectical argument game. Observe also that (iii) the list of possible attacks of an EQR dialogue is such that it comprehends moves that target every existing element of the D-ASPIC$^+$ arguments, enabling also the same types of attacks/defeats of the dialectical argument games. Finally, it is worth taking into account (iv) the presence of *relevance conditions*[ix] in both games and dialogues protocols.

The fact that EQR dialogues contemplate rational disagreement among (the two parties of) agents makes them befitting the dialectical admissible/preferred game protocol. That is to say, the outcome might generate different blocks of justified information which can be formalised as arguments members of conflicting admissible/preferred extensions according to Dung's credulous semantics. From these considerations, it logically follows:

Proposition 14. *Let* Arg *be an instantiated* D-scheme *in a pdAF. Let also* Φ_P-$\mathscr{D}^n$ *identifies a terminated* Φ_P*-dialectical game for* Arg. *Then a terminated* EQR *dialogue starting with* Arg *is won or tied by PRO iff there exists a dialectical winning strategy* Φ_P-$\mathscr{W}^n$ *for* Arg*, such that the set PRO*$(\Phi_P$-$\mathscr{W}^n)$ *of arguments moved by PRO in* Φ_P-$\mathscr{W}^n$ *is conflict free.*

Proof. It suffices to consider the proof of soundness and completeness of the dialectical games (Theorem 10) and Theorem 15. Indeed, there exists a semantical equivalence between EQR dialogues and the dialectical admissible/preferred argument games. □

6.3 Future Work

When dealing with natural language, agents do not always move fully formed arguments. For instance, during a conversation, *enthymemes*, i.e., arguments with an incomplete logical structure, may occur. Enthymemes are very common in everyday dialectical interactions, and they are usually the consequence of leaving implicit some piece of information. Such additional information can be obtained by *backwards* or *forward* expanding the

[ix]Refer to Definition (54.3(d)), (54.4(b)) and (54.4(c)).

given argument. The thorough research conducted on this topic [71, 146, 147] may elicit a new line of investigation that introduces an analysis of enthymemes in the EQR dialogue protocol that involves *D-scheme* and D-ASPIC$^+$ instantiations. This would certainly generate a better account of the everyday exchange of arguments performed by real-world agents.

Future studies may also lead to consideration of modifications to the protocol of Definition 54 such as to perform more (semantically) restrictive games. Theorem 15 showed how the instantiated *D-scheme* in an EQR dialogue, won or tied by PRO, is also a member of an admissible extension of the AF that defines the dialogue arguments. An adaptation of the current rules may enable the protocol to encode a more sceptical type of reasoning such that it does not accept rational disagreement as an outcome. This would naturally lead the dialogue topic's argument, if defended by PRO, to be included in the grounded extension of the AF. An interesting alternative could also be the consideration of a protocol that renders the kind of reasoning encoded by the stable semantics. Indeed, it can be shown that stable extensions enjoy properties that might be deemed desirable: "[...]*there is the possible absence of stable extensions. When applying stable semantics in, for instance, answer set programming, this can in fact be a desirable property. If one encodes a problem such that the possible solutions correspond with the stable extensions, then the absence of stable extensions indicates the absence of solutions to the original problem."* [31].

6.4 Conclusion

Building on the previously introduced EQR scheme and dialogue, the current chapter presents two main findings. The satisfaction of the rationality postulates under minimal resource consumption renders D-ASPIC$^+$ properties enticing for any formal system whose purpose is to model real-world agents argumentation. As such, the first contribution provided concerns embedding these features into the developed EQR scheme. Indeed, an adaptation to its logical structure generates the (1.) *D-scheme*, whose instantiation by logical formulae yields a fully-fledged D-ASPIC$^+$ argument instantiation. As with the former argument scheme from which it derive, also *D-scheme* can be defined as *D-claim*, *D-endorsement* and *D-endorsed-by-whom*, thus representing the specific focus of the distinct variants. Similarly, they can be employed as core elements of the EQR dialogue. The resulting (2.) formal protocol is: (a) unique, independently from the adopted variant; (b) informed by dialectical defeats; and (c) specifically conceived for D-ASPIC$^+$ arguments. That is to say, an EQR dialogue involving *D-scheme* instantiation ensures a better approximation of a real-world dialectical interaction between resource-bounded agents compared to the ‘more general’ version of EQR dialogue protocols introduced in Chapter 5.

Chapter 7

EQR dialogues and Explainable AI

The EQR scheme is an Argument Scheme (AS) specifically designed to work as the core element of an EQR dialogue. Considering that the EQR dialogue has been devised as a model of Explanation-Question-Response [90] interactions between an explainer and an explainee, it is easy to uncover the underlying connection with Explainable AI. This research field aims at providing a more transparent and understandable communication between users and AIs by explaining the rationale behind the operations and decisions of the artificial intelligence algorithms. The recent past has seen the rise of several new AI-driven recommendation systems, especially in the healthcare industry. In the medical sector, it is imperative that the exchange of information occurs in a clear and accurate way, and this has to be reflected in any deployed systems. For these reasons, AS and their critical questions represent suitable formal tools for modelling explanations, as shown by the state-of-the-art literature. Argument schemes provide templates for the explanations to be delivered. These templates will then be instantiated by the required information (which might differ depending on the situation). In this chapter, I am going to deploy the *EQR claim* scheme and dialogue to provide clinical explanations and assess their validity from the patient's viewpoint. An EQR scheme comprise multiple premises that can be investigated, with further inquiries, to disclose additional information. Moreover, such explanations can benefit from the conversion of EQR scheme instantiations into *D-scheme* instantiations, hence succeeding in conveying data more tailored to real-world resource-bounded agents.

7.1 Background

7.1.1 Clinical decision support systems

Artificial Intelligence constitutes a powerful means when deployed for assisting people in making well-informed decisions. Such assistance is rendered as a set of recommendations on which a human, who is interacting with the AI-based system, has the final word. In the healthcare sector, decision support systems (DSS) can be defined as those systems that:

> " [...]*provide clinicians or patients with computer-generated clinical knowledge and patient-related information, intelligently filtered or presented at appropriate times, to enhance patient care"* [101]

DSS prove to be especially useful since they present: 1) time-saving virtual assistance for practitioners; 2) help for patients in self-managing their health conditions, especially when they need routine information regarding their treatments (suggested by human clinicians); 3) better documentation, retrieval and presentation of data[i]; 4) cost saving due to the partial automation and optimization (while preferring cheaper, but still effective, treatment options) of the workflow [120]. Several DSS employ advanced machine learning algorithms as their main AI reasoning mechanism, although they do not seem to provide robust evidence of improved diagnostic performance in clinical environments [129]. Other DSS employ instead computational argumentation as their AI reasoning mechanism. Indeed, the handling of inconsistent and conflicting knowledge is a common feature in medical decision-making processes, when the opinions of several medical experts are solicited with regard to specific cases [82]. That is to say, argumentation-based reasoning enables clinical decision support systems (cDSS) to manage such inconsistencies. Arguments can reflect the opinion of a single practitioner, of a general/local medical guideline or even represent the viewpoint of a patient with respect to a particular treatment. The latter is of particular importance since it includes the patients' preferences in the reasoning process. As an example of argumentation-driven cDSS, the authors of [76] model medical recommendations via meta-level arguments that allows determining the ground on which object-level arguments are justified or preferred. The work of [45] moves towards the creation of a cDSS that employs the structured argumentation formalism of ABA^+ (stemming from the Assumption-Based Argumentation framework originally described in [22]) for automated reasoning with conflicting clinical guidelines, patients' information and preferences. The proposed clinical decision support system makes use

[i]Nevertheless, data must still prove to be reliable and trustworthy by showing that its provenance is non-repudiable [62].

of the Transition-based Medical Recommendation model (TMR) and its integration with electronic health records (EHR). This research is then extended in [46] where TMR representations of guideline recommendations are mapped to $ABA^{+}G$ an enhanced version of ABA^{+} capable of dealing with patient-centric goals and priorities among them.

Several studies have also been conducted in the field of cDSS considering patients suffering from multimorbidities (as [40] and [100]). Although the results thus far achieved have mostly been positive, in [18] the authors emphasize the need for further investigations in regard to: considerations of shared decisions, patients' preferences and social contexts, and a broader range of drug interactions (including food-drug interactions). Argumentation-based cDSS have been developed also in this specific research area: the CONSULT system (promoted in papers such as [74, 61, 11]) introduces a decision support tool to help patients with chronic conditions manage their multimorbidities in collaboration with their carers and the health care professionals who are looking after them.

7.1.2 Explainable AI (XAI)

The urge to overcome ethical issues involving AI-based systems, along with distrust from their users, constitutes the reason for the recent interest in Explainable AI (XAI). The idea is that the trustworthiness of AIs can be improved by building more transparent and interpretable tools capable of: explaining what the system has done, what it is doing now and what it is going to do next while disclosing salient information during these processes [13]. Nevertheless, XAI researchers seem to employ only their intuitions of what they consider to be a 'good' explanation when developing XAI models. Drawing from social sciences studies, Miller identifies four main points that can help deliver better explanations [94]:

(I) Explanations are *contrastive*, i.e., people usually don't ask why an event occurred, but why it has occurred instead of another event;

(II) Explanations are *selected*, i.e., people usually choose (being influenced by cognitive biases in the process) to provide a specific explanation from amongst a (possibly infinite) series of similar explanations;

(III) Referring to causality is *more important* in explanation than a reference to probabilities;

(IV) Explanations are *social* i.e., they are conveyed as part of dialogical interactions between people and account for their beliefs.

These points converge around a single conclusion: explanations are *contextual*.*"While an event may have many causes, often the explainee cares only about a small subset (relevant to the context), the explainer selects a subset of this subset (based on several different criteria), and explainer and explainee may interact and argue about this explanation."* [94].

There still remain many active issues concerning XAI. In [68], the authors present a (non-exhaustive) list of such challenges, that includes topics such as: accuracy versus interpretability, the use of abstractions to simplify explanations, or prioritizing competencies over decisions. Another problem is related to the end-user who is meant to receive the explanation. Indeed, the explainee might be an individual with a specific background as: an analyst, a judge, an IT developer, a policy-maker, a medical practitioner, etc. An effective explanation will take the target user groups into account since they might vary in regards to their knowledge and needs for what should be explained. Finally, it is interesting to notice that the research presented in [3], and more recently in [47], propose an account of explanations that is primarily argumentative. Similarly, the survey of [130] concludes that using argumentation to justify why an event started, or what led to a decision, can enhance explainability. These intuitions are also backed by [90], where it is suggested that AI systems should adopt an argumentation-based approach to explanations. The advocated approach points toward Douglas Walton's Argument schemes (AS), thoroughly discussed in [142].

7.2 Argument schemes and explanations in clinical setting

Argument schemes have been extensively investigated and employed in AI literature as a way to directly convey presumptive reasoning in multi-agent interactions (for example, [125, 107, 8]). Each AS is characterized by sets of critical questions (**CQ**s) whose purpose is to establish the validity of the scheme instantiations generating an argumentation framework that can then be evaluated according to one of Dung's semantics [53]. Such evaluation embeds the rationale for choosing an argument over another, meaning that instantiations of justified schemes can be employed for conveying explanations[ii]. The use of argument schemes for providing explanations is, indeed, not unusual, especially in

[ii]Notice, however, that it might also occur a case in which argumentation and explanation are tightly intertwined and it is difficult to distinguish one from the other. This can easily lead to the commitment of logical fallacies, such as the well known arguing in circle fallacy. A possible solution to determine such distinction involves looking at the context of the unfolding dialogue [17].

the clinical setting. In [128], the authors introduce the *Explain Argument Scheme*, which models explanations based on the reasons, types (of reasons) and levels (of abstraction) and shows a support or counter rationale for giving a particular drug to a patient. The work presented in [75, 111, 110] employs *Explanation templates* that differ according to the reasoning and argument scheme represented and include placeholders for the actual instantiated variables specific to a given application of the scheme. Formally:

Definition 55 (Explanation Template)**.** *Let* AS *be an argument scheme,* $\mathsf{Var}_{\mathrm{AS}} = \{x | x$ is a variable in $AS\}$ *be the set of variables of* AS *and* txt *is a natural language text that includes elements from* $\mathsf{Var}_{\mathrm{AS}}$. *Then, an Explanation template for* AS *can be rendered as the tuple* $\mathsf{Expl}_{\mathrm{AS}} = \langle \mathsf{AS}, \mathsf{Var}_{\mathrm{AS}}, \mathsf{txt} \rangle$.

Definition 56 (Explanation)**.** *An explanation is a tuple* $\langle \mathsf{Expl}_{\mathrm{AS}}, \mathrm{AS}_i \rangle$ *such that* $\mathsf{Expl}_{\mathrm{AS}}$ *is the explanation template introduced in Definition 55,* AS_i *is an acceptable* (*under one of Dung's semantics*) *instantiation of* AS*, and every variable in* txt *of* $\mathsf{Expl}_{\mathrm{AS}}$ *is instantiated by the corresponding element in* AS_i*.*

In the previously cited works, explanation templates have been mostly mapped to the *Argument Scheme for Proposed Treatment* (ASPT) which are particularly suited for providing medical explanations in a clinical scenario, as we are going to appreciate in the following section.

7.2.1 Argument Scheme for Proposed Treatment

Introduced in [77], the *Argument Scheme for Proposed Treatment* derives from the *Argument Scheme for Practical Reasoning* [5]. It instantiates an argument in support of a possible treatment *T*, given the facts *Ft* about the patients and the goal *G* to be achieved. As with each argument scheme, ASPT is accompanied by a series of critical questions that serve to assess the efficacy of the suggested treatment.

ASPT
Premise : Given the patient fact *Ft* *Premise* : In order to realise goal *G* *Premise* : Treatment *T* promotes goal *G*
Conclusion : Treatment *T* should be considered

The following list composes the critical questions specifically designed for ASPT (i.e., **ASPT.CQs**) [110]:

- (**ASPT.CQ1**) Has treatment *T* been unsuccessfully used on the patient in the past?
- (**ASPT.CQ2**) Has treatment *T* caused side effects on the patient?
- (**ASPT.CQ3**) Given patient facts *Ft*, are there any counter-indications to treatment *T* at step *i*?
- (**ASPT.CQ4**) Are there alternative actions to achieve the same goal *G*?

The instantiation of three of the above critical questions is informed by clinical argument schemes (which are medical specializations of argument schemes presented in [142]) that cover particular aspects of the suggested treatment. These schemes are *AS from Patient Medical History*, *AS from Negative Side Effect*, *AS for Contraindications* and they inform, respectively, ASPT.CQ1, ASPT.CQ2 and ASPT.CQ3 [110].

7.3 Explainer agents employing the EQR scheme: providing explanations via *EQR claim* dialogue in medical settings

7.3.1 EQR scheme

As we have already seen in the previous chapters, the EQR scheme is an AS expressly developed to be employed (e.g. as the starting point) in an EQR dialogue. The underlying idea is to merge the knowledge elicited by the ASOPA and ASEO formal patterns in a single scheme that would then yield the advantage of concentrating and synthesizing the same amount of information in a unique data structure that may be queried more conveniently. That is to say, the purpose of the EQR scheme is to formalise the consequences arising (and the presumptive reasoning leading to them) by acting upon a specific expert opinion focusing on either the: assertion (*EQR claim* scheme), the endorsement (*EQR endorsement* scheme) or the specific endorsement (*EQR endorsed-by-whom* scheme). In particular, the *EQR claim* scheme is accompanied by specifically designed critical questions[iii]:

- (**EQR_claim.CQ1**) How knowledgeable is E as an expert source?
- (**EQR_claim.CQ2**) Is E an expert in the field F that α is in?

[iii] These are the same CQs presented in Chapter 5 but displayed in a single list with a slightly different order (given that the former ASEO CQ7 refers to the Self-interest Question, which is not included among the *EQR claim* critical questions).

- (**EQR_claim.CQ3**) Did E's assertion imply A?
- (**EQR_claim.CQ4**) Is E's assertion entailing contradictory propositions?
- (**EQR_claim.CQ5**) Is A consistent with what other experts assert?
- (**EQR_claim.CQ6**) Is E's assertion based on the (facts expressed by) state R?
- (**EQR_claim.CQ7**) Is F a relevant disciplinary field?
- (**EQR_claim.CQ8**) Is E's assertion promoting a negative value?

Although not completely unrelated, considerations about the endorsement of the expert's assertion are less relevant in a medical setting. The focus is on the efficacy of the treatment ensuing from the expert's assertion and the conveyed value rather than the illocutionary endorsing act (or the identity of the endorser\s). These elements might instead be more significant when the EQR scheme is employed in different settings, such as a political debate or a trial. For such reasons, only the *claim* scheme will be taken into account in the remainder of the chapter.

7.3.2 EQR claim scheme and ASPT

Intuitively, the *EQR claim* scheme can display a large number of information bits to an *explainee* when looking for clarifications about a proposed treatment. Notice indeed that the *EQR claim* scheme can encompass ASPT such that it renders: (i) the treatment T as expert's opinion α (from an expert E in a field F), (ii) the patient fact Ft as part of the current state R and (iii) the goal to be realised G as proposition A. That is to say, by embedding ASPT into the *EQR claim* scheme, it will be possible to give more opportunities for inquiry to an agent seeking medical explanations. Certainly, in this way, there are more aspects that can be interrogated and that can help in finding a satisfactory (and more complete) explanation. For example, the additional data comprised in the current state R, the field of expertise F of the expert E, the immediate consequence S entailed by the proposed treatment, or the value v conveyed by the truth-value of A, all of these are elements that can be questioned by the patients. Especially, knowing the source of the recommendation (E) may boost the agent's trust in the explainer and the patient compliance with the advised medical care plan. Moreover, the rationale behind the provided explanation can be further investigated (resulting in additional, more detailed, explanations) thanks to the extra information supplied by the answers to each critical question and argument that informs the *EQR claim* scheme. An *EQR claim* Explanation Template is then determined as in Definition 55, although it employs the *EQR claim* scheme rather than a generic AS. Similarly, we can formalise an instance of such a template as:

Definition 57 (*EQR claim* Explanation)**.** *An* EQR claim *explanation is a tuple* $\langle \mathsf{Expl}_{\text{EQR\_claim}}, \text{EQR\_claim}_i \rangle$ *such that* $\mathsf{Expl}_{\text{EQR\_claim}}$ *is the explanation template for the* EQR claim *scheme,* EQR_claim_i *is an acceptable (under one of Dung's semantics) instantiation of the* EQR claim *scheme, and every variable in* txt *of* $\mathsf{Expl}_{\text{EQR\_claim}}$ *is instantiated by the corresponding element in* EQR_claim_i.

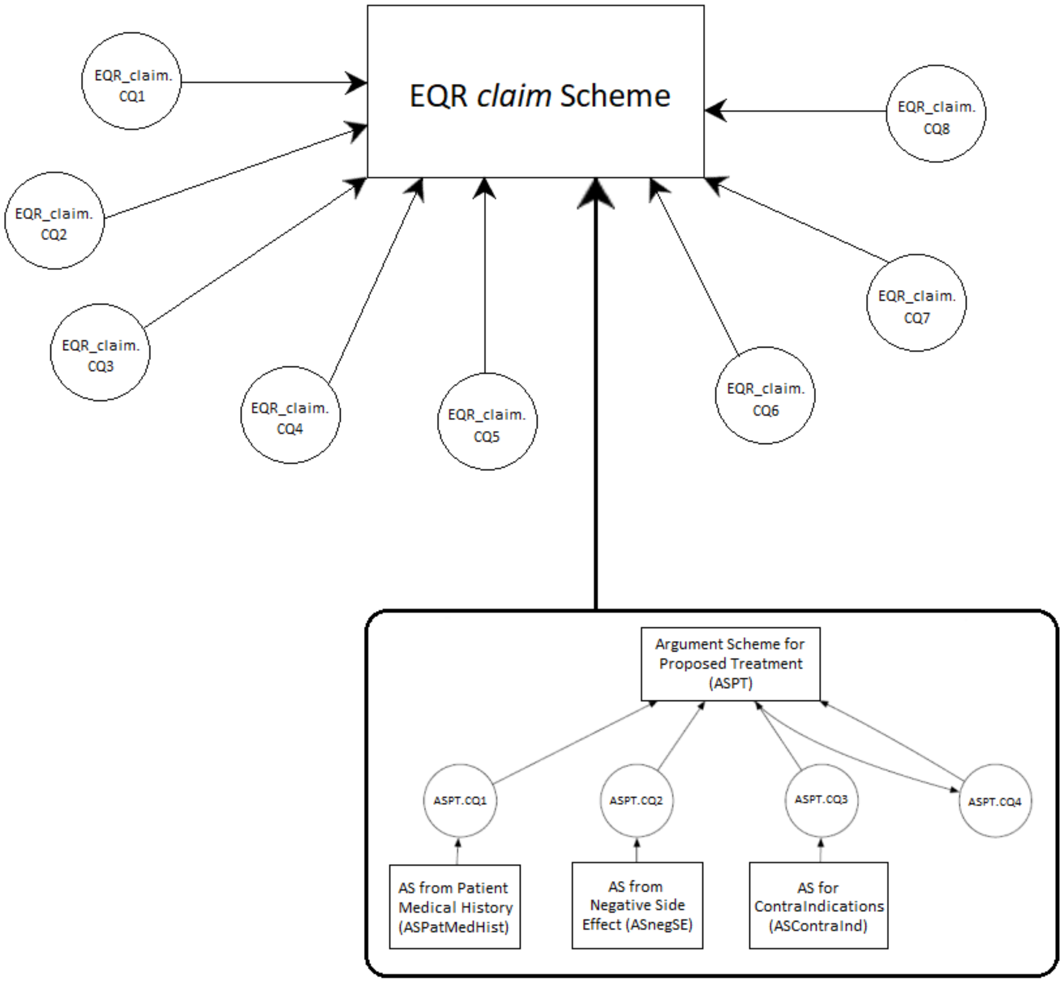

Figure 7.1: Graphical representation of the relationship between argument schemes and critical questions. The arrows identify which elements inform which arguments. Indeed, CQs help to inform their corresponding argument scheme since they allow to assess its validity. Consider also that ASPT and ASPT.CQ4 instantiate each other (as specified in [110]).

Example 17. *Suppose that we have an acceptable* (*under one of Dung's semantics*) *instantiation of the* EQR claim *scheme, informed by its critical questions, ASPT and a knowledge base* $\mathscr{K}$. *Assume also that the scheme variables* $\mathsf{Var}_{\mathrm{EQR\_claim}} = \{\mathrm{R}, \mathrm{E}, \mathrm{F}, \alpha, \mathrm{S}, \mathrm{A}, \mathrm{v}\}$ *are equivalent to the following:*

$\boxed{R}$: the patient's previous health record and the current fever and headache (due to COVID-19)

$\boxed{E}$: the NICE guidelines[iv]

$\boxed{F}$: medical management of COVID-19

$\boxed{\alpha}$: the administering of paracetamol

$\boxed{S}$: the reduction of fever and headache

$\boxed{A}$: controlling the virus negative effects

$\boxed{v}$: the patient's wellbeing

Finally, let txt *be the natural language text:* "Given {R}, the expertise of {E} in the field of {F} indicates {α} as an effective treatment. This should lead to {S} which will bolster the goal of {A} and promote {v}". *Then, the actual* EQR claim *Explanation would be*[v]*:*

> "Given **the patient's previous health record and the current fever and headache (due to COVID-19)**, the expertise of **the NICE guidelines** in the field of **medical management of COVID-19** indicates **the administering of paracetamol** as an effective treatment. This should lead to **the reduction of fever and headache** which will bolster the goal of **controlling the virus negative effects** and promote **the patient's wellbeing**".

7.3.3 EQR claim Dialogue and explanations

Explanations provided to real-world agents might involve different types of dialogues, e.g. mostly *persuasion* [105], *information-giving*, *information-seeking* [141] and *query* [43]. Such explanations should then account for simultaneous occurrences of multiple (possibly intertwined) dialogues. The research presented in [87] (which results have also been

[iv] `https://www.nice.org.uk/guidance`

[v] Consider that the conclusion of the *EQR claim* scheme has not been included in the text txt. That is because, in the presented example, making the conclusion explicit would add redundancy to the explanation, which might undermine its clarity and overload the patient with information.

employed in [116, 111]) makes use of the *Control Layer* construct, i.e., specific rules that determine the commencement and termination of a protocol, to allow for these simultaneous occurrences. Nevertheless, in Chapter 5 we have presented each Explanation-Question-Response dialogue (EQR) as already incorporating locutions for handling persuasion, information-giving/seeking and query tasks without the need for adopting a Control Layer. As such, an *EQR claim* dialogue will start from the *EQR claim* explanation of Definition 57 and will then try to assess its validity from the explainee point of view eschewing, in the process, any formalisms that would unnecessarily complicate the protocol (such as the mentioned Control Layer).

7.3.4 Explanations via D-Scheme instantiations

In the previous chapters, we have examined D-ASPIC$^+$ arguments[vi] along with their features and the advantages they entail over standard ASPIC$^+$. Indeed, D-ASPIC$^+$ is a recently introduced version of ASPIC$^+$ that provides a fully rational account for resource-bounded real-world agents [51]. As already proved in [75], ASPIC$^+$ arguments can be easily mapped to argument schemes. Similarly, we can show an analogous representation and render the *EQR claim* scheme as a D-ASPIC$^+$ argument. This will generalise the Definitions of chapter 6 through a more comprehensive Proposition:

Proposition 15. *Every* EQR *scheme can be represented as a D-ASPIC$^+$ argument.*

Proof. Let *Arg* be an *EQR* scheme, hence composed of Premises and Conclusion. Consider now that $\mathsf{prem}(Arg) = \mathsf{prem}(\text{Premise}_1) \cup \ldots \cup \mathsf{prem}(Premise_n)$, for $5 \leq n \leq 6$, and $\mathsf{conc}(Arg) = \text{Conclusion}$. Assume also that no strict rules are involved, i.e., $\mathscr{R}_s = \emptyset$, and that $\mathsf{sub}(Arg) = \mathsf{sub}(\text{Premise}_1) \cup \ldots \cup \mathsf{sub}(Premise_n) \cup \{Arg\}$, for $5 \leq n \leq 6$. Finally, let *Arg* be exclusively labelled with either Ⓒ or Ⓢ (meaning that the speaker is committing to the argument or it is just supposing it).
Then, we can trivially identify a direct mapping between *Arg* and the D-ASPIC$^+$ $Arg' = (\{\text{Premise}_1, \ldots, \text{Premise}_n, d\}, \emptyset, \text{Conclusion})$ or $Arg'' = (\emptyset, \{\text{Premise}_1, \ldots, \text{Premise}_n, d\}, \text{C-onclusion})$, depending upon the Ⓒ or Ⓢ labelling of the argument, where $n(\mathsf{conc}(\text{Premise}_1), \ldots, \mathsf{conc}(\text{Premise}_n) \rightsquigarrow \text{Conclusion}) = d$, such that $d \in \mathscr{R}_d$ is the defeasible TopRule. □

Following the conclusion yielded by Proposition 15, the *EQR claim* dialogue can then be played employing D-ASPIC$^+$ arguments, meaning that we will be able to reproduce a better approximation of a real-world interaction in a more human-friendly dialogue. As

[vi] Refer especially to D-ASPIC$^+$ section in chapter 2.

such, in a medical scenario, this could mean being capable of delivering explanations better tailored and understandable for the patients and their needs.

7.4 Practical implementation

A possible implementation of the proposed EQR explanation (Definition 57) may be rendered through a virtual chatbot that interacts with a user via an *EQR claim* dialogue. In this sense, the chatbot can be regarded as the last operative element of a cDSS such as, for example, the CONSULT system. CONSULT [74, 61, 11] is a cDSS designed to help patients self-manage their condition and adhere to agreed-upon treatment plans in collaboration with healthcare professionals. Its main components are outlined in Figure 7.2.

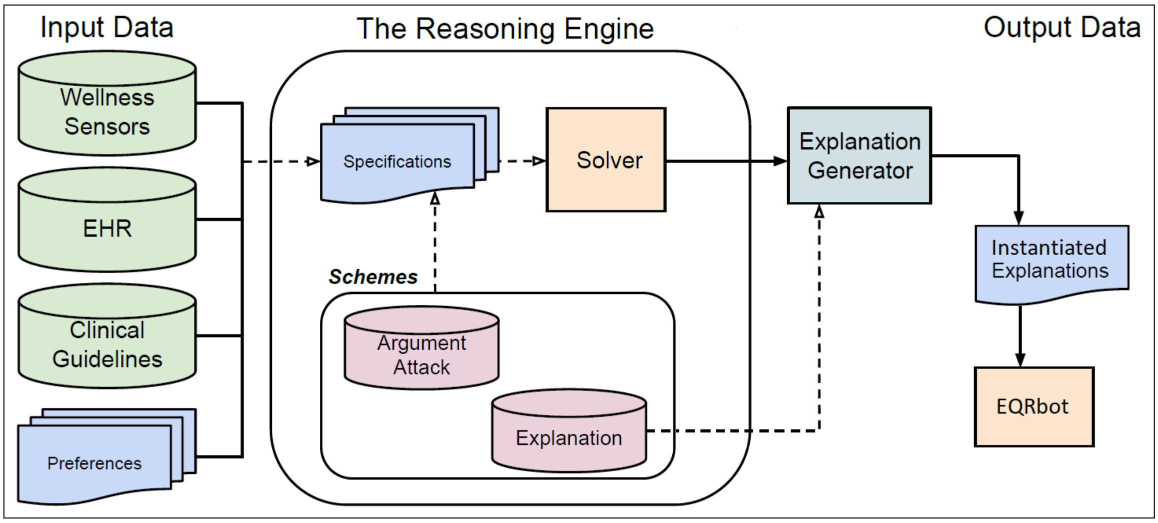

Figure 7.2: Flow chart describing the internal structure of the CONSULT cDSS. The input data is provided by different sources. The *Schemes* are templates for structuring and representing arguments, attacks and explanations. A formal language (i.e., first-order logic) is used to encode the knowledge, retrieved from the input data, in terms of *Specifications* that will then instantiate the attack and argument schemes subsequently computed in the resulting AF by ASPARTIX [60]. ASPARTIX is an answer-set *Solver* that takes argumentation frameworks as input and calculates admissible extensions (according to one of Dung's semantics) as specified by the user. The *Explanation Generator* (based on the sound and complete algorithm developed in [75]) constructs textual explanations for the recommendations according to explanation templates and the acceptable arguments produced by the *Solver*. The output will be stored in the *Instantiated Explanations* repository which elements will feed the *EQRbot*, the chatbot responsible for interacting with the patient.

7.4.1 EQRbot

The agent that will handle the interaction with the patient is a retrieval type chatbot, i.e., a kind of bot that focuses on retrieving contexts and keywords from the user's prompts in order to select the best response to give[vii]. The explanation process will occur as delineated in Figure 7.3. After having provided the initial explanation (i.e., the EQR explanation informed by an acceptable instantiation of the *EQR claim* scheme), the patient will be asked to express their opinion. If the user is satisfied with the explanation, then the conversation will immediately end. Alternatively, the *EQR claim* dialogue will initiate, and the chatbot will demand: a brief context (e.g., *"Would you please specify the context of your explanation request?"*) along with the actual request from the patient. Consider that the interaction is not limited by a specific set of options to which the explainee needs to comply: the choice of words to use for formulating the inquiries is completely unrestricted. By matching stored explanations (all of which account for the stakeholders' preferences), context and user input, the bot will output the additional solicited information (Figure 7.4). Observe that the double query prompted by the bot ensures a significant reduction of misunderstandings when providing answers to the patient. That is because the matching occurs via a double-layer word similarity counter function based on a BoW (Bag of Words) model. The *EQR claim* dialogue will then unfold according to its protocol. The explainer (chatbot) can be considered successful in its clarification attempt (which will bring the dialogue to an end) if the proposed explanation is deemed satisfactory by the user. Consider that the patient is aware of the EQRbot's inability to address questions regarding information not stored within (or not accessible by) the CONSULT system. As such, a satisfactory explanation may also be depicted as the realization that the user has to contact a healthcare professional should they have further queries. This will stop the loop of answers/questions and will end the conversation. It will continue otherwise.

It should be noted that the presence of multiple initial acceptable EQR explanations will not affect the chatbot operations. Since all of the explanations are acceptable, there is no need to further invoke the reasoning engine. The explanations are all considered equally good, seeing that our criteria for presenting an explanation is its acceptability (in turn influenced by the stakeholders' preferences), and so the EQRbot will randomly choose one of the available options and will then begin its interaction with the user. To this end, observe also that the bot is designed to avoid any unnecessary prolongation of the interaction to focus only on the required explanations. For this reason, the EQRbot

[vii] `https://github.com/FCast07/EQRbot`

will not start a conversation (nor even send a message) without the user prompt, but will react to each received text.

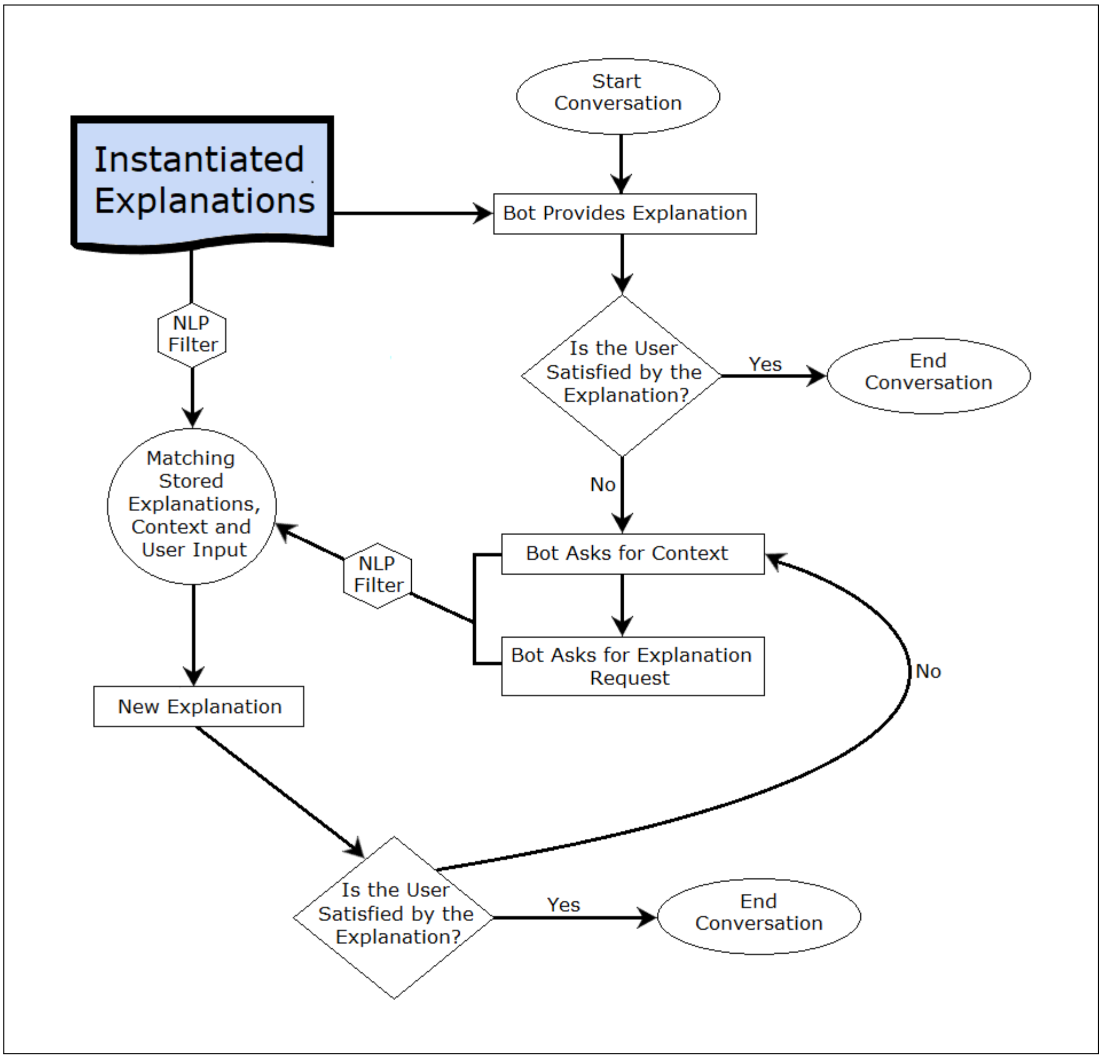

Figure 7.3: Flow chart describing the high-level operations performed by the chatbot (EQRbot).

7.4.1.1 NLP filter

The chatbot employs a Natural Language Processing (NLP) filter we developed in order to refine the input it receives from the patient and the stored instantiated explanations (Figure 7.3). The filtering process comprises: (a) the separation of the considered data into lists of single words (tokenization); (b) the elimination of the most common English words, including conjunctions and prepositions (stop-word removal); (c) the transformation of each verb into its infinitive form (lemmatization). The purpose of this refinement procedure is to ease the word matching between a patient's request and the system stored

Algorithm 5 Matching Queries/Explanations

Input: EXP, an EQR_explanation, and the (finite) set of the possible user's queries Q
Output: all the requested explanation

```
print(EQR_explanation)
for each q ∈ Q:
    q == (c, r)        ## q is a pair composed by a context (c) and specific request (r) ##
    find_specific_explanation(q)
end for each
·
·
Function find_specific_explanation(q)
    NLP_filter(c)
    NLP_filter(r)
    specific_explanation = " "
    similarity_counter = 0
    provisional_explanation = " "
    for each EX ∈ EXP \ {EQR_explanation}
        NLP_filter(EX)
        if double_layer_matcher(c, r, EX) > similarity_counter then
            similarity_counter = double_layer_matcher(c, r, EX)
            provisional_explanation = EX
        endif
    end for each
    specific_explanation = provisional_explanation
    print(specific_explanation)
end Function
```

information. Notice that NLP does not influence the reasoning engine nor its outcome (i.e., the resulting arguments and their status), it only facilitates the matching operation.

7.4.1.2 The algorithm

The EQRbot's inner operations can be described by an algorithm, Algorithm 5, that takes as input the *Instantiated Explanations* repository (EXP), along with the set of all the possible user queries (Q) related to the data conveyed by the initially provided EQR_explanation (which is also an element of EXP). The procedure continues until the depletion of all the possible queries of Q, that is to say, until the user is satisfied with the received information. Intuitively, NLP_filter corresponds to the function that performs a series of Natural Language Process operations as outlined in 7.4.1.1. double_layer_matcher, instead, represents the BoW similarity procedure in charge of identifying the appropriate response to be delivered. double_layer_matcher takes advantage of the context designation, the frequency of key terms occurrence and multiple cross-counts of the input words and the system stored data. Each resulting explanation will then be printed and displayed in the chatbot graphical user interface (GUI).

Proposition 16. *Given the interacting user collaboration (i.e., no out-of-context, nonsense or out-of-the-system-capability input), Algorithm 5 is both sound and complete.*

Indeed, the procedure can provide the requested information that is correct according to the user's input (soundness), and all such answers can be conveyed by the algorithm (completeness). Obviously, this is limited by the data held by the system at the time of the explanation delivery. That is to say, the procedure can only generate explanations determined by the information saved in the system's knowledge base.

Proof.

- [Soundness] The chatbot retrieves the patient's prompt (q) as a pair of context (c) and request (r). Then, the function `find_specific_explanation` (lines 8-23) matches the input with one of the explanations stored in the system (EX) according to a BoW similarity procedure denoted double_layer_matcher (lines 16-18). The result of this operation will then consist of the information requested by the user. In case of a mismatch, the process can be repeated until the user's satisfaction (lines 2-5).
- [Completeness] All the requested information can be conveyed by the algorithm. Indeed, each additional explanation the patient might require (associated with the initial EQR explanation) is already saved in the system. They can all be retrieved with the corresponding query (lines 2-5).

□

Since no machine learning operation is involved, hence no time is consumed in training a model, the algorithm will take polynomial time to run. That is because the function `find_specific_explanation` will be called a maximum of $|\mathsf{Q}|$ times, i.e., up to the number of elements of Q.

7.4.2 Current Implementation

Let us consider the EQR explanation of Example 17. We implemented it via a Telegram GUI. We chose to deploy the EQRbot via Telegram due to (i) its reputation as one of the most well-known and utilized instant messenger applications, and (ii) its programmer-friendly BOT API. To clarify the interaction depicted in Figure 7.4, let us suppose that the user monitored by the CONSULT system is a woman named Frida. The electronic health record supplies the cDSS with two pieces of information: the patient is pregnant, and she is currently suffering from fever and headache caused by the Covid-19 virus. To ease Frida from the pain, when prompted, the CONSULT reasoning engine computes an acceptable (as per Definition 3) piece of advice in the form of an EQR explanation. The EQRbot will display such a recommendation while encouraging also to ask for more details. Supplying the context and the specific request, the patient will demand the rationale behind the choice of the expert that provides the received clinical advice (similarly to **EQR_claim.CQ1**). The chatbot reply involves a natural language explanation (Figure 7.4, left frame). In the example, the system considers NICE guidelines as the most reliable source and provides an explanation accordingly. Notice, however, that CONSULT is engineered as a cDSS that supplies recommendations attained from general health guidelines (e.g., NICE). As explicitly stated before its usage, since the system is not supposed to handle conflicts that require professional medical knowledge to be solved, the users should seek advice from their general practitioners would such a circumstance occur. Indeed, this may cause significant harm to the patient if not handled correctly, as emphasised in [119]. For the same reason, the cDSS (hence the EQRbot) is also updated by the patient's latest wellness sensor readings, the data in their EHR (so, for example, it will not recommend a therapy that has caused negative side effects in the past) and their preferences regarding treatments. The conversation continues (Figure 7.4, right frame) and Frida interrogates the chatbot for additional information: she desires to know if alternative treatments are available[viii] (because, for instance, the drug indicated by CONSULT is not currently accessible to her). However, the cDSS confirms its previous recommendation informing Frida that, due to her pregnancy, paracetamol is the most appropriate remedy to assume.

7.4.2.1 Evaluating the EQRbot against the previous baseline

A seven day within-subjects mixed-methods run in-the-wild ([143]) study has been conducted to assess the usability and acceptability of the CONSULT system with two differ-

[viii] similarly to a variation of **EQR_claim.CQ3** that could be phrased as: 'Are there alternative E's assertions that can be acted upon to (consistently) imply A?'.

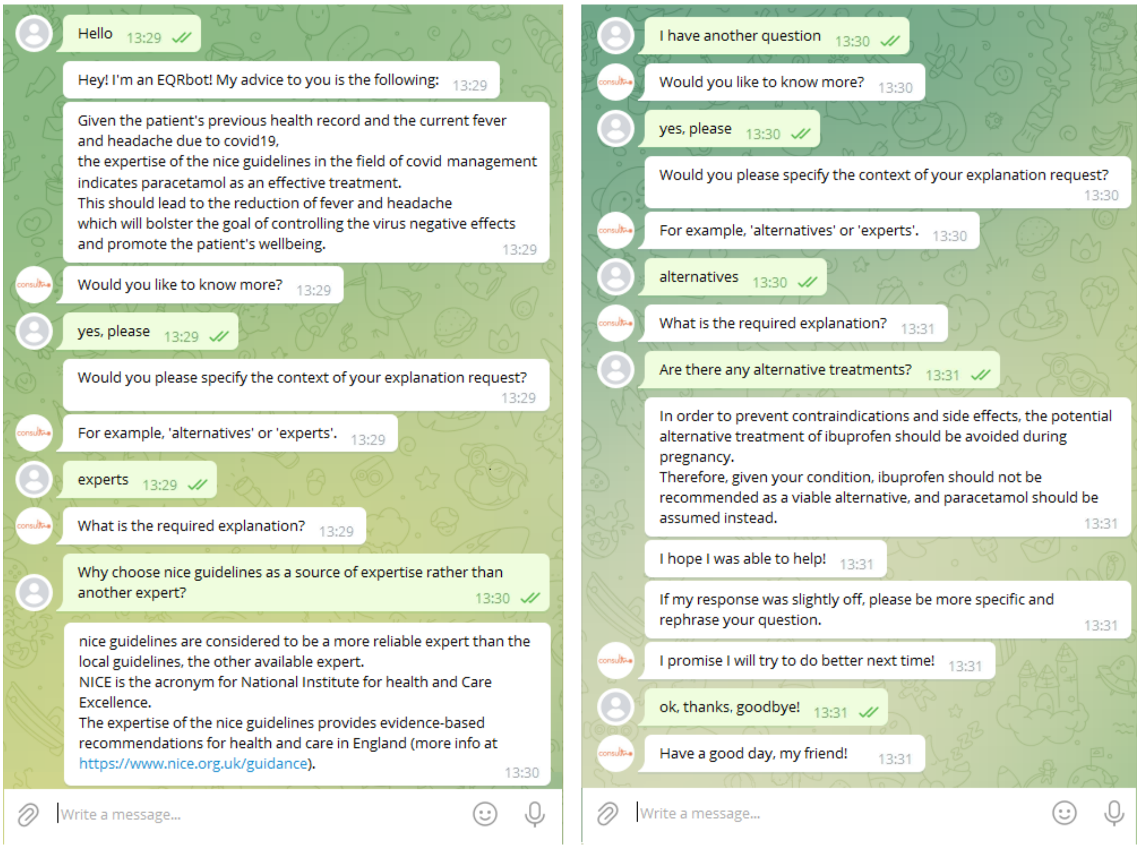

Figure 7.4: Instance of an *EQR claim* dialogue starting from the explanation of Example 17.

ent versions: with and without a chatbot. Such a pilot study demonstrated that real users could employ the application over an extended period [10]. Connie, the conversational agent previously equipped with the cDSS at the time of the experiment, accommodates the patients willing to seek immediate evidence-based advice about a specific health problem. Informed by the user's vital data, preferences, EHR and clinical guidelines retrieved by the CONSULT system, the chatbot provides any additional explanation regarding the proposed recommendation. The main aspects that characterise Connie can be outlined as:

- *User's Input.* No free interaction occurs since the user's prompt is restricted to hard-coded multiple options.
- *Interface.* The chat, and related conversation log, are graphically displayed via Mattermost[ix].
- *Chatbot Type.* Connie is a rule-based chatbot[x], i.e., an agent capable of responding only by following predetermined (scripted) replies according to the user's input.

[ix] `https://mattermost.com/`
[x] `https://www.codecademy.com/article/what-are-chatbots`

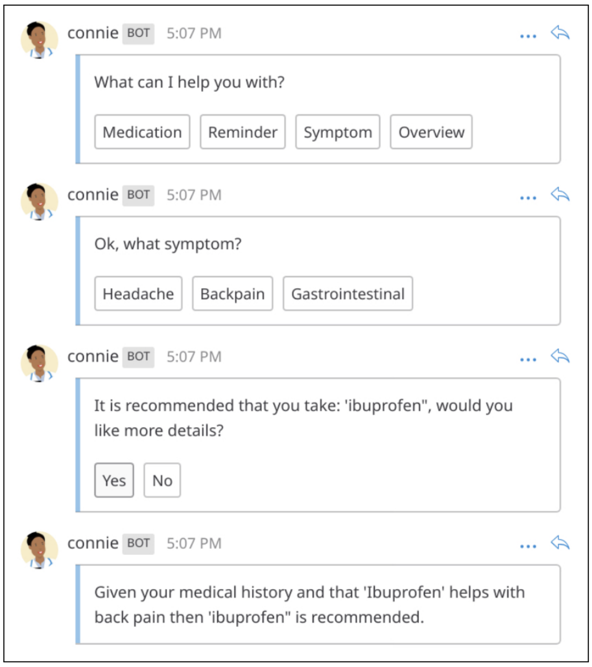

Figure 7.5: Example of interaction with the CONSULT's chatbot.

- *Reasoning Engine.* The bot leverages the results of the operations performed by the CONSULT system by means of the argumentation solver ASPARTIX.
- *Explanation Delivery.* No particular strategy is deployed. The explanations are triggered via the options selected by the user.

An example of a conversation with Connie is illustrated in Figure 7.5. Here the interacting patient is given the choice of selecting among four different options in response to the question *"What can I help you with?"*. The user then decides to report a symptom concerning backpain, asking also for more details once a reply is given. This option triggers one last response from the chatbot, thus providing the explanation behind the rationale of the proposed recommendation. Nonetheless, Connie presents some limitations, as summarised by the result of the pilot study: "[...] *the lack of a more natural conversation flow when interacting with the chatbot (e.g. close to the one that they [the patients] would have with their GP)"* [10]. Against Connie, considered as the previous baseline, the EQRbot yields several advantages, as highlighted by the following bullet points:

- *User's Input.* Free textual interaction. Each user's prompt will be parsed by the chatbot NLP filter and matched with the most appropriate reply. Any nonsense or out-of-context input will be addressed by a random response from the bot.

- *Interface.* The chat, and related conversation log, are graphically displayed via Telegram[xi].

- *Chatbot Type.* EQRbot is a retrieval-based chatbot, i.e., an agent that mostly retrieves its replies from a database of potential responses according to the most relevant match with the user's input.

- *Reasoning Engine.* The bot leverages the results of the operations performed by the CONSULT system by means of the argumentation solver ASPARTIX.

- *Explanation Delivery.* The strategy follows the (partial) deployment of an *EQR claim* dialogue. The aim is to reduce the number of potential user queries (including possible follow-on questions) and concerns by concentrating the most relevant information about a specific recommendation within a single explanation, i.e., the one elicited by an acceptable instantiation of the EQR scheme.

The EQRbot represent an improvement over Connie since it addresses (in four out of the five listed main features) the shortcomings ensuing from the pilot study outcome. Indeed, it allows for *(i)* better approximations of natural conversations without textual restriction, by employing *(ii)* Telegram GUI, i.e., a more user-friendly, and popular messaging application than Mattermost. In general, *(iii)* retrieval-based chatbots are more versatile and flexible than rule-based ones, hence more suited for real-world exchange of arguments. Finally, despite its simplicity, *(iv)* having an explanation strategy brings the EQRbot closer to an authentic question-answer dialogue.

7.5 Future Work

The envisaged practical implementation presents some limitations. For example, the user must always utter the **ask** locution to challenge the received explanation, while the chatbot plays 'on the defence' since it is only answering questions (employing the **state** locution) and uttering the same **ask** locutions repeatedly. As such, moves as **deny** and **deny initial** aren't available to the players. Furthermore, the system does not dynamically generate additional arguments while the dialogue unfolds[xii]. This reduces the number of potential questions that the chatbot can address. That is to say, if the user inquiries regard the initial

[xi] `https://telegram.org/`

[xii] Notice that this does not exclude the generation of additional arguments, once the patient's conditions change, before the starting of a new dialogue.

explanation or its specifics, then no problem will occur. Indeed, the *EQR claim* explanation has been informed by several CQs that should comprehend all the possible challenges moved to it. However, if the inquiries concern a reference that involves a modification to the knowledge base held by the system, then the virtual agent will fail in providing a coherent response. For these reasons, future implementations might include a thorough deployment of all the available locutions and an additional adjustment to the chatbot's capability of producing new arguments (given the potential extra information acquired while the dialogue unfolds). Further improvements could also arise by combining the recent developments in the field of *Argument Mining* [24] with additional chatbot code-based instructions. Indeed, the mining and detection of the user's arguments[xiii] can assist the bot in minimizing its misunderstanding of the sentences occurring in the dialogue. Finally, the clarity of the explanation might benefit from a graphical visualization of the arguments deployed during the dialogue. This could be achieved by Monkeypuzzle [52] a web-based tool designed for argument analysis and visualisation (which is also part of OAPL, an open-source suite of argumentation software [144]).

To fully evaluate the EQRbot performances after the proposed software updates, we will then set up a user study specifically designed to test its new and enhanced explanatory capabilities.

7.6 Conclusion

In this chapter, an approach that integrates *EQR claim* dialogue and scheme in the current research landscape involving cDSS and argument schemes-based explanations has been proposed. Focusing, in particular, on studies regarding the Argument Scheme for Proposed Treatment, a possible way for enhancing the related explanation template has been presented. Indeed, one of the main advantages entailed by the provided contributions is (1.) the incorporation of ASPT into the *EQR claim* scheme. This will give more chances of inquiry to an agent seeking explanations since there are more aspects that can be interrogated and that can help in finding a satisfactory (and more complete) explanation. For example, which specific expert is informing the treatment is a piece of information that might increase the users' trust in the medical recommendation system. In addition, as already seen in chapter 6 and proven in Proposition 15, the *EQR claim* scheme can be rendered as a D-ASPIC$^+$ argument, hence providing instantiations that better account

[xiii] Argument(ation) mining has been defined as "*the general task of analyzing discourse on the pragmatics level and applying a certain argumentation theory to model and automatically analyze the data at hand*" [69].

for real-world resource-bounded agents. Another contribution concerns the use of a virtual chatbot to deliver explanations. The proposed practical implementation envisages equipping the CONSULT cDSS with a (2.) chatbot that employs the *EQR claim* dialogue and scheme to convey explanations. This is a fair contribution to the research field of argumentation-based dialogues. Indeed, such a bot is guided exclusively by an argumentation reasoning engine in its decision-making process while it follows the *EQR claim* protocol when interacting with the user. NLP is resorted to only as a means for enhancing the words matching between the user input and the stored explanations accessible to the chatbot. Such choice does not affect the status of arguments, the AF, or the validity of the operations of the reasoning engine (including the soundness and completeness of its core algorithm).

Chapter 8

Conclusion

8.1 Major contributions

In this chapter, the most important findings and the envisaged possible future research are summarized. Overall, we can distinguish eleven main contributions stemming from the presented work.

I. Design of specific argument game proof theories for Dialectical Cl-Arg admissible/preferred semantics.

Incorporating dialectical defeats in the standard structure of the argument games proved to be a non-trivial process which yielded the discovery of interesting properties that differentiate dialectical games from the standard ones. That is to say, dialectical game proof theories enjoy (a) specific relevance conditions that identify their protocols and yield (b) the uniqueness of their winning strategies, whilst property F1 ensures (c) the conflict-freeness of the set of arguments moved by the proponent in the winning strategy. The latter is of particular importance since it provides the games with various possible efficiency improvements. Aside from these common features, dialectical admissible/preferred game protocols enjoy also the soundness and completeness properties while restricting the moves played by OPP (i.e., the opponent of the player PRO, that is the proponent and starter of the game). Intuitively, if the arguments moved by PRO have already shown their acceptability with respect to the current winning strategy, there is no need to endlessly extend the game by allowing OPP to repeat its arguments in the same disputes.

II. Design of specific argument game proof theories for Dialectical Cl-Arg grounded semantics.

The design of the dialectical grounded game protocol can benefit from the same features enjoyed by the dialectical admissible/preferred game. However, to prove the properties of soundness and completeness of the procedure, another condition must be satisfied. That is to say, if the game terminates and PRO is the winner, the set of arguments the proponent moved in the winning strategy must be epistemically maximal (*em*). Such requirement ensures that the resulting admissible set is a subset of the pdAF grounded extension (which can be constructed by iterating the framework characteristic function *em* sets). Another prerequisite to securing the validity of the dialectical grounded game concerns the intuitive idea that the defence of an argument member to the grounded extension must resort to some other argument than itself. This is reflected by an additional constraint in the protocol that prevents PRO from repeating the arguments it has already moved in the same disputes.

III. Development of 3-value labelling for Dialectical Cl-Arg and entailed properties.

The characterisation of this new method relies on an adaptation of the dialectical defeats (which parametrize the set of justified arguments labelled *IN* without affecting *OUT* or *UNDEC* arguments) to the standard labelling approach illustrated in [97]. As in the standard case, it is possible to show the correspondence between dialectical extensions and dialectical labellings. In addition, the latter also enjoys several properties thanks to its connection with Dialectical Cl-Arg. These properties mostly involve the presence and relations existing among dialectically *IN* arguments, thus providing innovative insights for potential algorithms that employ dialectical labellings. It should also be noted that dialectical labellings and dialectical defeats affect each other. Indeed, if the label of an argument changes, this can influence the dialectical defeats related to that argument which, in turn, can cause other changes in the labelling.

IV. Creation of an algorithmic procedure for the generation of the dialectical preferred labellings.

Inspired by algorithms and procedures already existing in the literature [25], the enumeration of the dialectical preferred labellings introduces original strategies which take advantage of the properties enjoyed by the dialectical labellings. The result comprises a (sound and complete) naive algorithm and its optimized version. The general idea for both procedures is to reduce the number of dialectical transition sequences (which yields dialectical admissible labellings as an outcome) that need to be investigated in order to get dialectical preferred labellings. In particular, the optimized algorithm prunes the search space of the procedure by preventing further examinations of dialectical transition sequences where arguments are dialectically labelled *IN* but neither their elementary nor their logically equivalent arguments (whose premises are subsets of the considered argument premises) are *IN* as well. This strategy provides a significant efficiency improvement over the naive algorithm.

V. Creation of an algorithmic procedure for the generation of the dialectical grounded labelling.

In Dialectical Cl-Arg, the iteration of the characteristic function, which yields the constructive definition of the grounded extension when starting from the empty set, requires the epistemic maximality of the admissible sets generated at each iteration by the function. Since the dialectical algorithm simulates this operation, it is essential to include an epistemically maximal check (and, if needed, a transformation) of the set in its procedure. Nevertheless, thanks to the notion of *epistemically preserved* admissible sets, it has been possible to prove that, against expectations, the procedure did not require an epistemically maximal closure at the end of each of its iterations. As such, there is little difference between the standard [97] and the dialectical (sound and complete) algorithms in charge of handling the grounded semantics.

VI. Introduction of the Explanation-Question-Response (EQR) argument scheme and its variants.

The EQR scheme is an argument scheme whose logical structure can be seen as halfway between the *Argument Scheme Over Proposals for Action* [7] and the *Argument Scheme from Expert Opinion* [139]. Its purpose is to formalise the consequences arising (and the presumptive reasoning leading to them) by acting upon a specific expert opinion and presents three variants depending on the particular focus of the provided scheme: assertion (*EQR claim* scheme); endorsement (*EQR endorsement* scheme); or precise endorsement (*EQR endorsed-by-whom* scheme). The scheme proves to be especially useful for explanation tasks that involve the retrieval of information from experts.

VII. Fully-fledged introduction of the Explanation-Question-Response dialogue protocols.

The Explanation-Question-Response dialogue (EQR), first sketched in [90], is a novel type of dialogue halfway between persuasion, information-giving/seeking and query that already incorporates locutions for handling each of these tasks without the need for adopting a Control Layer. For this reason, the EQR dialogue (declined in its three alternatives *claim, endorsement,* and *endorsed-by-whom*) provides a model more suited to supply explanations to agents eschewing formalisms that would unnecessarily complicate the protocol. Specifically, the dialogue protocols present a series of locutions along with an axiomatic semantics which describes the pre-conditions necessary for each locution, and any post-conditions arising from them. Such pre- and post-conditions influence the *commitment store* of every agent participating in the dialogues and also affect the legal moves (where the term 'move' is used to denote the utterance of a single locution) available to the players.

VIII. Integration of EQR scheme and D-ASPIC$^+$ arguments to generate D-Scheme instantiations.

The full rational account conveyed by D-ASPIC$^+$, the general (dialectical) framework for structured arguments, can be encapsulated into the EQR scheme as well. That is to say, it is possible to translate the properties of D-ASPIC$^+$ (including the satisfaction of practical desiderata and rationality postulates) into EQR scheme instantiations. The result of such an operation produces the so-called *D-Scheme* instantiations, which leverage the EQR template whose instantiations generate real-world characterizations of resource-bounded uses of arguments rendered as D-ASPIC$^+$ arguments.

IX. Specification of a formal protocol for an EQR dialogue that involves D-scheme instantiations.

Following the introduction of the *D-scheme* instantiations, it is possible to accommodate an EQR dialogue protocol to formalise a dialogue that handles them. This requires introducing a way of dealing with dialectical defeats and interpreting the locutions relations as standard AF attacks/defeats. The first condition has been solved by parametrizing the set of locutions moved by the proponent to defend the argument that started the dialogue. The latter, instead, has been addressed by a comprehensive analysis of the locutions available to the players. As a consequence, the resulting dialogue can be easily evaluated according to one of Dung's semantics. Indeed, the devised formal protocol discloses several similarities with the monological architecture of the dialectical preferred/admissible argument games which, as expected, proves to be equivalent.

X. Implementation of the EQR claim *scheme as an explanation template for a clinical decision support system.*

Clinical decision support systems (cDSS) provide treatment recommendations to patients /practitioners following specific medical guidelines. The *EQR claim* scheme constitutes an explanation template particularly suited for medical scenarios where a patient self-manages his/her health conditions. Indeed, by embedding ASPT, i.e., a clinical specialized AS, the *EQR claim* scheme is capable of supplying additional information that enhances the given explanation and may also increase the reliability towards the overall system. That is to say, the possibility of retrieving data, as the identity of the recommender expert and its field of expertise, may ensure patient compliance with the suggested treatment.

XI. Engineering of a software chatbot capable of delivering explanations (partially) following the Explanation-Question-Response dialogue protocol.

In order to provide an application of the *EQR claim* dialogue, a software chatbot has been engineered. After having disclosed the initial explanation (i.e., the EQR explanation informed by an acceptable instantiation of the EQR scheme stored in a database), the bot starts a (partial) *EQR claim* dialogue, if the user is still not satisfied. The software will then retrieve the context and the requested information. Matching stored explanations, context and user's input (with the assistance of an NLP filter), the chatbot will output the additional solicited data. This operation will be repeated until no more information is demanded (or until the user realizes she has to contact a healthcare professional should she have further queries). Notice that the bot deploys only a partial *EQR claim* dialogue because it does not make use of locutions other than **state** and **ask**.

8.2 Future work

There are several future theoretical research directions and potential practical applications established by the current thesis findings.

8.2.1 Future theoretical work

The introduction of the dialectical argument game proof theories provides a fruitful ground for additional inquiries. Indeed, the study commenced in [32, 33] and extensively presented in Chapter 3 may be further expanded by increasing the range of protocols under consideration. This would include possible investigations of stable [31], semi-stable [25] and ideal semantics [55, 26]. An analysis of the different semantics can also inspire the design of new dialectical labelling algorithms, especially considering that the procedures for the generation of dialectical stable and semi-stable labellings can benefit from the ones already developed in Chapter 4. Moreover, a supplementary optimization of all the proposed algorithms may arise from a consideration of a 4 or 5-values labellings procedure [41, 99] (after having provided a suited adaptation to dialectical labellings).

The EQR scheme variants presented in Chapter 5 provide another interesting research direction that may be pursued. That is to say, investigating the logical nature of the *endorsement* of an expert opinion may result in a formalism that would produce a new kind of labelling or propose an update of the existing bipolar argumentation frameworks (BAFs) [39, 131]. Notice that the former may involve a consideration of the illocutionary force conveyed by the EQR scheme instantiations, thus listing 'endorse' among the other pragmatical meanings (such as 'command' and 'promise') associated with speech acts. The latter, instead, suggests a connection with BAFs, i.e., AFs characterised by the presence of a *support* relation among the arguments of the framework. An exploration of the potential link between support and endorsement may provide additional insight regarding the formal role of the EQR scheme variants focussing on them.

Another possible line of investigation may be elicited by the formal EQR dialogue protocol of Chapter 6. This protocol, informed by dialectical defeats and specifically conceived for D-ASPIC$^+$ arguments resulting from the instantiations of the *D-Scheme* template, is particularly suited for resource-bounded real-world agents. In real-world dialogues, agents do not always move fully formed arguments. Rather, it often occurs that they move incomplete arguments instead, called *enthymemes* [71, 146, 147]. Introducing a logical analysis of enthymemes in the dialogue protocol would generate a better approximation of the everyday exchange of arguments performed by resource-bounded agents.

8.2.2 Possible applications of current research

8.2.2.1 Fake news filter: detecting a 'false expert'

Let us now consider a potential application of an *EQR claim* dialogue in a social media scenario. As a special case of such a dialogue involving *D-claim* scheme instantiations, we can consider a *fake news filter*. Notice indeed that if the expert (role fulfilled by PRO[i]) is not capable of positively answering each one of the proffered critical questions, then it can be identified as a 'false expert'. That is to say, the information claimed by such agent(s) should not be considered true, but false. The *EQR claim* dialogue protocol that handles *D-claim* scheme instantiations can then be considered as a suitable dialectical tool to (probabilistically) prevent fake news. Such a tool can be administered by an automated software (the filter) responsible for leading the dialogue and assessing its outcome. Finding a reliable dataset from which to derive the agents' knowledge bases (agents composing the two teams of players) constitutes the first step. The instantiation of the initial *D-claim* scheme with the news we want to test represents the subsequent phase of the procedure. The following steps will comprehend the unfolding of the dialogue according to the formal protocol defined in Chapter 6. Then, if the opponent team succeeds in challenging the news and winning the dialogue, we can assume (with a high degree of probability) that the information embedded in the initial *D-claim* scheme is not acceptable according to the current knowledge base, hence, it is false. Similar reasoning would occur in case the outcome of the dialogue results in a draw, i.e., when a rational disagreement between the contender parties is reached. Otherwise, the victory of the proponent team allows establishing (with the same degree of probability) that the information conveyed by the initial *D-claim* scheme is justifiable according to the current knowledge base, hence, it is true.

8.2.2.2 EQR dialogue as a general explanations provider

Humans may interact with chatbots to request explanations in several different circumstances, for example: in a medical setting, where a patient could interrogate the bot about its health conditions or ask about its prescribed medications; in a teaching or training environment, where the student can benefit from an additional review on the subject of study; during a trial, when the involved lawyers can acquire data from the bot regarding

[i]Consider that the proponent's team may not always correspond with the expert that asserts the expert opinion of the *D-scheme*. Indeed, it can also be identified with the agent(s) that simply support(s) or deliver(s) the expert opinion.

past cases; or even a situation where tech support is needed, or, more generally, any scenario where information retrievals may be profitable. In all of these situations, the chatbot acts as a virtual assistant that, not only helps the human retrieve the desired information, but also permits the assessment of the validity of the attained explanation. The involved dialogue also ensures the acquisition of further data by questioning (if necessary) the provided answers. To successfully reach such a broader versatility, the bot introduced in Chapter 7 would need to improve its functionalities by employing the entire spectrum of locutions available in the EQR protocol. Resorting to Argument Mining [24] may result in more accurate users' arguments detection which translates into higher understanding capability and smoother conversations. Although the chatbot would still work on an argumentation-based reasoning engine (as in Chapter 7), it will not count only on a single database and will, indeed, be able to extend its knowledge base. Equipped with a web-scraping algorithm, the bot could independently retrieve the required data from reliable web pages (such as university-related encyclopedias) and construct its arguments from these newly established knowledge bases. This last additional procedure will largely increase such a chatbot's ability to provide explanations in multiple contexts and according to different topics.

On the other hand, the EQR dialogue protocol could also be employed to structure explanations in large language models, such as the recent Meta's LLama [127] and the (multimodal) GPT-4 [1]. Indeed, although they have great potential as AI-generative mechanisms, they still lack full explanatory capabilities concerning their underlying procedures and black-box algorithms. As we have seen, argumentation and EQR dialogue can be leveraged as powerful tools to enhance AI-systems explanation delivery.

8.3 Outcome of Research Objectives

This project stemmed from a philosophical curiosity towards the contrast existing between real-world agents' limitations and the seeming omniscient requirements of classical logic formalism. Being the powerful theoretical tool that it is, logic has always served the purpose of modelling the surrounding world and human reasoning processes, often paired with argumentation. Yet many of the elaborated systems lack a proper consideration of the scarce availability of resources that characterise real-world entities and their interactions. Comprehending how to formalise the everyday argumentative exchange of information between agents may entail significant contributions in multiple disciplinary fields, including a more accurate approximation of our brain functions. However, before being able to improve our logical mapping of the world, we should try to reduce the gap that spans reality and our argumentative models. Developing approaches that account for resource constraints would be a move in the right direction, but first, we have to answer the following:

- *Can we envisage and design logically structured argumentative proof-theoretical models of dialectical interactions that approximate real-world resource-bounded agent reasoning?*

Guided by such a question, I have attempted to address the smaller related issues:

1. (a) Does a proof theoretical account of real-world exchanges of arguments differ from the standard argument games?
   (b) If so, can we still devise a procedure that enjoys soundness and completeness properties?
2. (a) Can the dialectical inconsistencies of arguments be efficiently represented by the standard 3-value labelling method and algorithms?
   (b) Otherwise, can we adapt such a labelling approach to provide sound and complete algorithms more suited to dialectical characterisations of arguments?
3. (a) Can we envisage an argument scheme capable of enhancing explanations specifically addressed to real-world resource-bounded agents?
   (b) If so, can we design a dialogue protocol informed by this template and the enhanced explanation it conveys?
   (c) Finally, can we outline a practical implementation for such an argument scheme?

The response to 1(a) is positive, whereas the results emerging from answering 1(b) can be found in points *I* and *II* of Section 8.1. On the contrary, the outcome of 2(a) is negative, and a reply to 2b) produced the findings listed in *III*, *IV* and *V* of Section 8.1. The remaining contributions *VI-XI* of Section 8.1 arose from the positive answer to 3(a) and the responses to 3(b) and 3(c).

Of course, this is only the tip of the iceberg. A minuscule step, let us say. There is still a lot to learn and discover if we hope to deepen our understanding of the rationale underpinnings human thoughts and mental processes. Nevertheless, there is one thing I believe: regardless of all of our restrictions, shortcomings, and limitations, we will keep moving forward. Being real-world resource-bounded agents is what drives us towards the pursuit of science.

Appendix A

List of locutions

Table A.1

Table 1. **Locutions to control the dialogue**		
Locution	**Pre-conditions**	**Post-conditions**
Enter dialogue	- Speaker has not already uttered enter dialogue	- Speaker has entered dialogue
Leave dialogue	- Speaker has uttered enter dialogue	- Speaker has left dialogue
Turn start	- Speaker has not already made its move	- Speaker has started its turn
Turn finished	- Speaker has started its turn - Speaker has finished making its move	- Speaker has finished its turn
Concede	- (1) Hearer has made an attack **or** - (2) Hearer has asked a question on an element of speaker's position **or** - (3) Hearer has answered a question asked by the speaker	- (1) Speaker committed to the negation of the element that was denied by the hearer **or** - (2) Speaker does not know the answer **or** - (3) Speaker committed to the statement given as a response by the hearer
Reject	- Hearer has made an attack **or** - Hearer has answered a question asked by the speaker	- Disagreement reached
Change player	- Speaker has uttered turn finished	- Speaker and hearer switch roles so new speaker can now make a move

Table A.2

Table 2a. **Locutions to state a proposition**		
Locution	**Pre-conditions**	**Post-conditions**
State circumstances(R)	- Speaker uttered enter dialogue	- Speaker committed to R - Speaker committed to $R \in$ *States*
State opinion(α)	- Speaker uttered enter dialogue	- Speaker committed to $\alpha \in$ *Opinions*
State expert(E)	- Speaker uttered enter dialogue	- Speaker committed to $E \in$ *Experts*
State field(F)	- Speaker uttered enter dialogue	- Speaker committed to $F \in$ *Fields*
State competences($\alpha_{\langle E,F \rangle}$)	- Speaker uttered enter dialogue - Speaker committed to $\alpha \in$ *Opinions* - Speaker committed to $E \in$ *Experts* - Speaker committed to $F \in$ *Fields*	- Speaker committed to $\alpha_{\langle E,F \rangle} \in$ *Competences*
State consequences($\alpha_{\langle E,F \rangle}, R, S$)	- Speaker uttered enter dialogue - Speaker committed to R - Speaker committed to $R \in$ *States* - Speaker committed to $\alpha \in$ *Opinions* - Speaker committed to $E \in$ *Experts* - Speaker committed to $F \in$ *Fields* - Speaker committed to $\alpha_{\langle E,F \rangle} \in$ *Competences*	- Speaker committed to $\texttt{assert}(\alpha_{\langle E,F \rangle}, R, S) \in \texttt{assert}$ - Speaker committed to $S \in$ *States*
State logical consequences(S, A)	- Speaker uttered enter dialogue - Speaker committed to R - Speaker committed to $R \in$ *States* - Speaker committed to $\alpha \in$ *Opinions* - Speaker committed to $E \in$ *Experts* - Speaker committed to $F \in$ *Fields* - Speaker committed to $\alpha_{\langle E,F \rangle} \in$ *Competences* - Speaker committed to $\texttt{assert}(\alpha_{\langle E,F \rangle}, R, S) \in \texttt{assert}$ - Speaker committed to $S \in$ *States*	- Speaker committed to $S \models A$ - Speaker committed to $A \in$ *Prop*
State purpose(A, v)	- Speaker uttered enter dialogue - Speaker committed to R - Speaker committed to $R \in$ *States* - Speaker committed to $\alpha \in$ *Opinions* - Speaker committed to $E \in$ *Experts* - Speaker committed to $F \in$ *Fields* - Speaker committed to $\alpha_{\langle E,F \rangle} \in$ *Competences* - Speaker committed to $\texttt{assert}(\alpha_{\langle E,F \rangle}, R, S) \in \texttt{assert}$ - Speaker committed to $S \in$ *States* - Speaker committed to $S \models A$ - Speaker committed to $A \in$ *Prop*	- Speaker committed to $v \in$ *Values*

Table A.3

Table 3a. **Locutions to ask about an agent's position**		
Locution	**Pre-conditions**	**Post-conditions**
Ask circumstances(R)	- Hearer uttered enter dialogue - Speaker uttered enter dialogue - Speaker not committed to circumstances(R)	- Hearer must reply with state circumstances(R) or don't know(R)
Ask opinion(α)	- Hearer uttered enter dialogue - Speaker uttered enter dialogue - Speaker not committed to opinion(α)	- Hearer must reply with state opinion(α) or don't know(α)
Ask expert(E)	- Hearer uttered enter dialogue - Speaker uttered enter dialogue - Speaker not committed to expert(E)	- Hearer must reply with state expert(E) or don't know(E)
Ask field(F)	- Hearer uttered enter dialogue - Speaker uttered enter dialogue - Speaker not committed to field(F)	- Hearer must reply with state field(F) or don't know(F)
Ask competences($\alpha_{\langle E,F \rangle}$)	- Hearer uttered enter dialogue - Speaker uttered enter dialogue - Speaker not committed to competences($\alpha_{\langle E,F \rangle}$)	- Hearer must reply with state competences($\alpha_{\langle E,F \rangle}$) or don't know($\alpha_{\langle E,F \rangle}$)
Ask consequences($\alpha_{\langle E,F \rangle}, R, S$)	- Hearer uttered enter dialogue - Speaker uttered enter dialogue - Speaker not committed to consequences($\alpha_{\langle E,F \rangle}, R, S$)	- Hearer must reply with state consequences($\alpha_{\langle E,F \rangle}, R, S$) or don't know($\alpha_{\langle E,F \rangle}, R, S$)
Ask logical consequences(S, A)	- Hearer uttered enter dialogue - Speaker uttered enter dialogue - Speaker not committed to logical consequences(S, A)	- Hearer must reply with state logical consequences(S, A) or don't know(S, A)
Ask purpose(A, v)	- Hearer uttered enter dialogue - Speaker uttered enter dialogue - Speaker not committed to purpose(A, v)	- Hearer must reply with state purpose(A, v) or don't know(A, v)

Table A.4

Table 4a. **Locutions to attack elements of a position**		
Locution	**Pre-conditions**	**Post-conditions**
Deny circumstances(R)	- Speaker uttered enter dialogue - Hearer uttered enter dialogue - Hearer committed to R - Hearer committed to $R \in States$	- Speaker committed to deny circumstances(R)
Deny opinion(α)	- Speaker uttered enter dialogue - Hearer uttered enter dialogue - Hearer committed to $\alpha \in Opinions$	- Speaker committed to deny opinion(α)
Deny expert(E)	- Speaker uttered enter dialogue - Hearer uttered enter dialogue - Hearer committed to $E \in Experts$	- Speaker committed to deny expert(E)
Deny field(F)	- Speaker uttered enter dialogue - Hearer uttered enter dialogue - Hearer committed to $F \in Fields$	- Speaker committed to deny field(F)
Deny consequences($\alpha_{\langle E,F \rangle}, R, S$)	- Speaker uttered enter dialogue - Hearer uttered enter dialogue - Hearer committed to R - Hearer committed to $R \in States$ - Hearer committed to $\alpha \in Opinions$ - Hearer committed to $E \in Experts$ - Hearer committed to $F \in Fields$ - Hearer committed to $\alpha_{\langle E,F \rangle} \in Competences$ - Hearer committed to $\texttt{assert}(\alpha_{\langle E,F \rangle}, R, S) \in \texttt{assert}$ - Hearer committed to $S \in States$	- Speaker committed to deny consequences($\alpha_{\langle E,F \rangle}, R, S$) $\in$ $\texttt{assert}$
Deny logical consequences(S, A)	- Speaker uttered enter dialogue - Hearer uttered enter dialogue - Hearer committed to R - Hearer committed to $R \in States$ - Hearer committed to $\alpha \in Opinions$ - Hearer committed to $E \in Experts$ - Hearer committed to $F \in Fields$ - Hearer committed to $\alpha_{\langle E,F \rangle} \in Competences$ - Hearer committed to $\texttt{assert}(\alpha_{\langle E,F \rangle}, R, S) \in \texttt{assert}$ - Hearer committed to $S \in States$ - Hearer committed to $S \models A$ - Hearer committed to $A \in Prop$	- Speaker committed to deny logical consequences(S, A) $S \models A$
Deny purpose(A, v)	- Speaker uttered enter dialogue - Hearer uttered enter dialogue - Hearer committed to R - Hearer committed to $R \in States$ - Hearer committed to $\alpha \in Opinions$ - Hearer committed to $E \in Experts$ - Hearer committed to $F \in Fields$ - Hearer committed to $\alpha_{\langle E,F \rangle} \in Competences$ - Hearer committed to $\texttt{assert}(\alpha_{\langle E,F \rangle}, R, S) \in \texttt{assert}$ - Hearer committed to $S \in States$ - Hearer committed to $S \models A$ - Hearer committed to $A \in Prop$ - Hearer committed to $v \in Values$	- Speaker committed to deny purpose(A, v)

Table A.5

Table 5a. **Locutions to attack validity of elements**		
Locution	**Pre-conditions**	**Post-conditions**
Deny circumstances exist(R)	- Speaker uttered enter dialogue - Hearer uttered enter dialogue - Hearer committed to R	- Speaker committed to deny circumstances exist(R)
Deny competences exist($\alpha_{\langle E,F \rangle}$)	- Speaker uttered enter dialogue - Hearer uttered enter dialogue - Hearer committed to R - Hearer committed to $R \in States$ - Hearer committed to $\alpha \in Opinions$ - Hearer committed to $E \in Experts$ - Hearer committed to $F \in Fields$ - Hearer committed to $\alpha_{\langle E,F \rangle} \in Competences$	- Speaker committed to deny competences exist($\alpha_{\langle E,F \rangle}$)
Deny resultant state exists(S)	- Speaker uttered enter dialogue - Hearer uttered enter dialogue - Hearer committed to R - Hearer committed to $R \in States$ - Hearer committed to $\alpha \in Opinions$ - Hearer committed to $E \in Experts$ - Hearer committed to $F \in Fields$ - Hearer committed to $\alpha_{\langle E,F \rangle} \in Competences$ - Hearer committed to $S \in States$	- Speaker committed to deny resultant state exists(S)
Deny proposition exists(A)	- Speaker uttered enter dialogue - Hearer uttered enter dialogue - Hearer committed to R - Hearer committed to $R \in States$ - Hearer committed to $\alpha \in Opinions$ - Hearer committed to $E \in Experts$ - Hearer committed to $F \in Fields$ - Hearer committed to $\alpha_{\langle E,F \rangle} \in Competences$ - Hearer committed to $S \in States$ - Hearer committed to $A \in Prop$	- Speaker committed to deny proposition exists(A)
Deny value exists(v)	- Speaker uttered enter dialogue - Hearer uttered enter dialogue - Hearer committed to R - Hearer committed to $R \in States$ - Hearer committed to $\alpha \in Opinions$ - Hearer committed to $E \in Experts$ - Hearer committed to $F \in Fields$ - Hearer committed to $\alpha_{\langle E,F \rangle} \in Competences$ - Hearer committed to $S \in States$ - Hearer committed to $A \in Prop$ - Hearer committed to $v \in Values$	- Speaker committed to deny value exists(v)

Table A.6

Table 2b. **Locutions to state a proposition**		
Locution	**Pre-conditions**	**Post-conditions**
State expert's assertion($\alpha_{\langle E,F\rangle}$)	- Speaker uttered enter dialogue - Speaker committed to R - Speaker committed to $R \in$ *States* - Speaker committed to $\alpha \in$ *Opinions* - Speaker committed to $E \in$ *Experts* - Speaker committed to $F \in$ *Fields* - Speaker committed to $\alpha_{\langle E,F\rangle} \in$ *Competences*	- Speaker committed to $\mathtt{assert}(\alpha_{\langle E,F\rangle}) \in \mathtt{assert}$
State consequences ($\mathtt{assert}(\alpha_{\langle E,F\rangle}), R, S$)	- Speaker uttered enter dialogue - Speaker committed to R - Speaker committed to $R \in$ *States* - Speaker committed to $\alpha \in$ *Opinions* - Speaker committed to $E \in$ *Experts* - Speaker committed to $F \in$ *Fields* - Speaker committed to $\alpha_{\langle E,F\rangle} \in$ *Competences* - Speaker committed to $\mathtt{assert}(\alpha_{\langle E,F\rangle}) \in \mathtt{assert}$	- Speaker committed to $\mathsf{endorsement}(\mathtt{assert}(\alpha_{\langle E,F\rangle}), R, S)$ $\in \mathsf{endorsement}$ - Speaker committed to $S \in$ *States*
State logical consequences(S, A)	- Speaker uttered enter dialogue - Speaker committed to R - Speaker committed to $R \in$ *States* - Speaker committed to $\alpha \in$ *Opinions* - Speaker committed to $E \in$ *Experts* - Speaker committed to $F \in$ *Fields* - Speaker committed to $\alpha_{\langle E,F\rangle} \in$ *Competences* - Speaker committed to $\mathtt{assert}(\alpha_{\langle E,F\rangle}) \in \mathtt{assert}$ - Speaker committed to $\mathsf{endorsement}(\mathtt{assert}(\alpha_{\langle E,F\rangle}), R, S)$ $\in \mathsf{endorsement}$ - Speaker committed to $S \in$ *States*	- Speaker committed to $S \models A$ - Speaker committed to $A \in$ *Prop*
State purpose(A, v)	- Speaker uttered enter dialogue - Speaker committed to R - Speaker committed to $R \in$ *States* - Speaker committed to $\alpha \in$ *Opinions* - Speaker committed to $E \in$ *Experts* - Speaker committed to $F \in$ *Fields* - Speaker committed to $\alpha_{\langle E,F\rangle} \in$ *Competences* - Speaker committed to $\mathtt{assert}(\alpha_{\langle E,F\rangle}) \in \mathtt{assert}$ - Speaker committed to $\mathsf{endorsement}(\mathtt{assert}(\alpha_{\langle E,F\rangle}), R, S)$ $\in \mathsf{endorsement}$ - Speaker committed to $S \in$ *States* - Speaker committed to $S \models A$ - Speaker committed to $A \in$ *Prop*	- Speaker committed to $v \in$ *Values*

Table A.7

Table 3b. **Locutions to ask about an agent's position**		
Locution	**Pre-conditions**	**Post-conditions**
Ask expert's assertion($\alpha_{\langle E,F\rangle}$)	- Hearer uttered enter dialogue - Speaker uttered enter dialogue - Speaker not committed to expert's assertion($\alpha_{\langle E,F\rangle}$)	- Hearer must reply with state expert's assertion($\alpha_{\langle E,F\rangle}$) or don't know expert's assertion($\alpha_{\langle E,F\rangle}$)
Ask consequences ($\texttt{assert}(\alpha_{\langle E,F\rangle}), R, S$)	- Hearer uttered enter dialogue - Speaker uttered enter dialogue - Speaker not committed to consequences($\texttt{assert}(\alpha_{\langle E,F\rangle}), R, S$)	- Hearer must reply with state consequences($\texttt{assert}(\alpha_{\langle E,F\rangle}), R, S$) or don't know($\texttt{assert}(\alpha_{\langle E,F\rangle}), R, S$)

Table A.8

Table 4b. **Locutions to attack elements of a position**		
Locution	**Pre-conditions**	**Post-conditions**
Deny consequences $(\texttt{assert}(\alpha_{\langle E,F\rangle}), R, S)$	- Speaker uttered enter dialogue - Hearer uttered enter dialogue - Hearer committed to R - Hearer committed to $R \in \mathit{States}$ - Hearer committed to $\alpha \in \mathit{Opinions}$ - Hearer committed to $E \in \mathit{Experts}$ - Hearer committed to $F \in \mathit{Fields}$ - Hearer committed to $\alpha_{\langle E,F\rangle} \in \mathit{Competences}$ - Hearer committed to $\texttt{assert}(\alpha_{\langle E,F\rangle}) \in \texttt{assert}$ - Hearer committed to $\mathsf{endorsement}(\texttt{assert}(\alpha_{\langle E,F\rangle}), R, S)$ $\in \mathsf{endorsement}$ - Hearer committed to $S \in \mathit{States}$	- Speaker committed to deny consequences $(\texttt{assert}(\alpha_{\langle E,F\rangle}), R, S) \in \mathsf{endorsement}$
Deny logical consequences(S,A)	- Speaker uttered enter dialogue - Hearer uttered enter dialogue - Hearer committed to R - Hearer committed to $R \in \mathit{States}$ - Hearer committed to $\alpha \in \mathit{Opinions}$ - Hearer committed to $E \in \mathit{Experts}$ - Hearer committed to $F \in \mathit{Fields}$ - Hearer committed to $\alpha_{\langle E,F\rangle} \in \mathit{Competences}$ - Hearer committed to $\texttt{assert}(\alpha_{\langle E,F\rangle}) \in \texttt{assert}$ - Hearer committed to $\mathsf{endorsement}(\texttt{assert}(\alpha_{\langle E,F\rangle}), R, S)$ $\in \mathsf{endorsement}$ - Hearer committed to $S \in \mathit{States}$ - Hearer committed to $S \models A$ - Hearer committed to $A \in \mathit{Prop}$	- Speaker committed to deny logical consequences(S,A) $S \models A$
Deny purpose(A,v)	- Speaker uttered enter dialogue - Hearer uttered enter dialogue - Hearer committed to R - Hearer committed to $R \in \mathit{States}$ - Hearer committed to $\alpha \in \mathit{Opinions}$ - Hearer committed to $E \in \mathit{Experts}$ - Hearer committed to $F \in \mathit{Fields}$ - Hearer committed to $\alpha_{\langle E,F\rangle} \in \mathit{Competences}$ - Hearer committed to $\texttt{assert}(\alpha_{\langle E,F\rangle}) \in \texttt{assert}$ - Hearer committed to $\mathsf{endorsement}(\texttt{assert}(\alpha_{\langle E,F\rangle}), R, S)$ $\in \mathsf{endorsement}$ - Hearer committed to $S \in \mathit{States}$ - Hearer committed to $S \models A$ - Hearer committed to $A \in \mathit{Prop}$ - Hearer committed to $v \in \mathit{Values}$	- Speaker committed to deny purpose(A,v)

Table A.9

Table 5b. **Locutions to attack validity of elements**		
Locution	**Pre-conditions**	**Post-conditions**
Deny expert's assertion($\alpha_{\langle E,F\rangle}$)	- Speaker uttered enter dialogue - Hearer uttered enter dialogue - Hearer committed to *R* - Hearer committed to $R \in States$ - Hearer committed to $\alpha \in Opinions$ - Hearer committed to $E \in Experts$ - Hearer committed to $F \in Fields$ - Hearer committed to $\alpha_{\langle E,F\rangle} \in Competences$ - Hearer committed to $\texttt{assert}(\alpha_{\langle E,F\rangle}) \in \texttt{assert}$	- Speaker committed to deny expert's assertion exist($\alpha_{\langle E,F\rangle}$)
Deny resultant state exists(*S*)	- Speaker uttered enter dialogue - Hearer uttered enter dialogue - Hearer committed to *R* - Hearer committed to $R \in States$ - Hearer committed to $\alpha \in Opinions$ - Hearer committed to $E \in Experts$ - Hearer committed to $F \in Fields$ - Hearer committed to $\alpha_{\langle E,F\rangle} \in Competences$ - Hearer committed to $\texttt{assert}(\alpha_{\langle E,F\rangle}) \in \texttt{assert}$ - Hearer committed to $S \in States$	- Speaker committed to deny resultant state exists(*S*)
Deny proposition exists(*A*)	- Speaker uttered enter dialogue - Hearer uttered enter dialogue - Hearer committed to *R* - Hearer committed to $R \in States$ - Hearer committed to $\alpha \in Opinions$ - Hearer committed to $E \in Experts$ - Hearer committed to $F \in Fields$ - Hearer committed to $\alpha_{\langle E,F\rangle} \in Competences$ - Hearer committed to $\texttt{assert}(\alpha_{\langle E,F\rangle}) \in \texttt{assert}$ - Hearer committed to $S \in States$ - Hearer committed to $A \in Prop$	- Speaker committed to deny proposition exists(*A*)
Deny value exists(*v*)	- Speaker uttered enter dialogue - Hearer uttered enter dialogue - Hearer committed to *R* - Hearer committed to $R \in States$ - Hearer committed to $\alpha \in Opinions$ - Hearer committed to $E \in Experts$ - Hearer committed to $F \in Fields$ - Hearer committed to $\alpha_{\langle E,F\rangle} \in Competences$ - Hearer committed to $\texttt{assert}(\alpha_{\langle E,F\rangle}) \in \texttt{assert}$ - Hearer committed to $S \in States$ - Hearer committed to $A \in Prop$ - Hearer committed to $v \in Values$	- Speaker committed to deny value exists(*v*)

Table A.10

Table 2c. **Locutions to state a proposition**		
Locution	**Pre-conditions**	**Post-conditions**
State endorser(X)	- Speaker uttered enter dialogue - Speaker committed to R - Speaker committed to $R \in States$ - Speaker committed to $\alpha \in Opinions$ - Speaker committed to $E \in Experts$ - Speaker committed to $F \in Fields$ - Speaker committed to $\alpha_{\langle E,F \rangle} \in Competences$ - Speaker committed to $\texttt{assert}(\alpha_{\langle E,F \rangle}) \in \texttt{assert}$	- Speaker committed to X - Speaker committed to $X \subseteq Endorsers$
State consequences $(\texttt{assert}(\alpha_{\langle E,F \rangle}), X, R, S)$	- Speaker uttered enter dialogue - Speaker committed to R - Speaker committed to $R \in States$ - Speaker committed to $\alpha \in Opinions$ - Speaker committed to $E \in Experts$ - Speaker committed to $F \in Fields$ - Speaker committed to $\alpha_{\langle E,F \rangle} \in Competences$ - Speaker committed to $\texttt{assert}(\alpha_{\langle E,F \rangle}) \in \texttt{assert}$ - Speaker committed to X - Speaker committed to $X \subseteq Endorsers$	- Speaker committed to $\mathsf{endorsement}(\texttt{assert}(\alpha_{\langle E,F \rangle}), X, R, S) \in \mathsf{endorsement}$ - Speaker committed to $S \in States$
State logical consequences(S,A)	- Speaker uttered enter dialogue - Speaker committed to R - Speaker committed to $R \in States$ - Speaker committed to $\alpha \in Opinions$ - Speaker committed to $E \in Experts$ - Speaker committed to $F \in Fields$ - Speaker committed to $\alpha_{\langle E,F \rangle} \in Competences$ - Speaker committed to $\texttt{assert}(\alpha_{\langle E,F \rangle}) \in \texttt{assert}$ - Speaker committed to X - Speaker committed to $X \subseteq Endorsers$ - Speaker committed to $\mathsf{endorsement}(\texttt{assert}(\alpha_{\langle E,F \rangle}), X, R, S) \in \mathsf{endorsement}$ - Speaker committed to $S \in States$	- Speaker committed to $S \models A$ - Speaker committed to $A \in Prop$
State purpose(A, v)	- Speaker uttered enter dialogue - Speaker committed to R - Speaker committed to $R \in States$ - Speaker committed to $\alpha \in Opinions$ - Speaker committed to $E \in Experts$ - Speaker committed to $F \in Fields$ - Speaker committed to $\alpha_{\langle E,F \rangle} \in Competences$ - Speaker committed to $\texttt{assert}(\alpha_{\langle E,F \rangle}) \in \texttt{assert}$ - Speaker committed to X - Speaker committed to $X \subseteq Endorsers$ - Speaker committed to $\mathsf{endorsement}(\texttt{assert}(\alpha_{\langle E,F \rangle}), X, R, S) \in \mathsf{endorsement}$ - Speaker committed to $S \in States$ - Speaker committed to $S \models A$ - Speaker committed to $A \in Prop$	- Speaker committed to $v \in Values$

Table A.11

Table 3c. **Locutions to ask about an agent's position**		
Locution	**Pre-conditions**	**Post-conditions**
Ask endorser(X)	- Hearer uttered enter dialogue - Speaker uttered enter dialogue - Speaker not committed to endorser(X)	- Hearer must reply with state endorser(X) or don't know(X)
Ask consequences ($\texttt{assert}(\alpha_{\langle E,F\rangle}),X,R,S$)	- Hearer uttered enter dialogue - Speaker uttered enter dialogue - Speaker not committed to consequences($\texttt{assert}(\alpha_{\langle E,F\rangle}),X,R,S$)	- Hearer must reply with state consequences($\texttt{assert}(\alpha_{\langle E,F\rangle}),X,R,S$) or don't know($\texttt{assert}(\alpha_{\langle E,F\rangle}),X,R,S$)

Table A.12

Table 4c. **Locutions to attack elements of a position**		
Locution	**Pre-conditions**	**Post-conditions**
Deny endorser(X)	- Speaker uttered enter dialogue - Hearer uttered enter dialogue - Hearer committed to X - Hearer committed to $X \subseteq Endorsers$	- Speaker committed to deny endorser(X)
Deny consequences (assert($\alpha_{\langle E,F\rangle}$),$X,R,S$)	- Speaker uttered enter dialogue - Hearer uttered enter dialogue - Hearer committed to R - Hearer committed to $R \in States$ - Hearer committed to $\alpha \in Opinions$ - Hearer committed to $E \in Experts$ - Hearer committed to $F \in Fields$ - Hearer committed to $\alpha_{\langle E,F\rangle} \in Competences$ - Hearer committed to assert($\alpha_{\langle E,F\rangle}$) $\in$ assert - Hearer committed to X - Hearer committed to $X \subseteq Endorsers$ - Hearer committed to endorsement(assert($\alpha_{\langle E,F\rangle}$),$X,R,S$) $\in$ endorsement - Hearer committed to $S \in States$	- Speaker committed to deny consequences (assert($\alpha_{\langle E,F\rangle}$),$X,R,S$) $\in$ endorsement
Deny logical consequences(S,A)	- Speaker uttered enter dialogue - Hearer uttered enter dialogue - Hearer committed to R - Hearer committed to $R \in States$ - Hearer committed to $\alpha \in Opinions$ - Hearer committed to $E \in Experts$ - Hearer committed to $F \in Fields$ - Hearer committed to $\alpha_{\langle E,F\rangle} \in Competences$ - Hearer committed to assert($\alpha_{\langle E,F\rangle}$) $\in$ assert - Hearer committed to X - Hearer committed to $X \subseteq Endorsers$ - Hearer committed to endorsement(assert($\alpha_{\langle E,F\rangle}$),$X,R,S$) $\in$ endorsement - Hearer committed to $S \in States$ - Hearer committed to $S \models A$ - Hearer committed to $A \in Prop$	- Speaker committed to deny logical consequences(S,A) $S \models A$
Deny purpose(A,v)	- Speaker uttered enter dialogue - Hearer uttered enter dialogue - Hearer committed to R - Hearer committed to $R \in States$ - Hearer committed to $\alpha \in Opinions$ - Hearer committed to $E \in Experts$ - Hearer committed to $F \in Fields$ - Hearer committed to $\alpha_{\langle E,F\rangle} \in Competences$ - Hearer committed to assert($\alpha_{\langle E,F\rangle}$) $\in$ assert - Hearer committed to X - Hearer committed to $X \subseteq Endorsers$ - Hearer committed to endorsement(assert($\alpha_{\langle E,F\rangle}$),$X,R,S$) $\in$ endorsement - Hearer committed to $S \in States$ - Hearer committed to $S \models A$ - Hearer committed to $A \in Prop$ - Hearer committed to $v \in Values$	- Speaker committed to deny purpose(A,v)

Table A.13

Table 5c. **Locutions to attack validity of elements**		
Locution	**Pre-conditions**	**Post-conditions**
Deny endorser exists(X)	- Speaker uttered enter dialogue - Hearer uttered enter dialogue - Hearer committed to X	- Speaker committed to deny endorser exists(X)
Deny resultant state exists(S)	- Speaker uttered enter dialogue - Hearer uttered enter dialogue - Hearer committed to R - Hearer committed to $R \in States$ - Hearer committed to $\alpha \in Opinions$ - Hearer committed to $E \in Experts$ - Hearer committed to $F \in Fields$ - Hearer committed to $\alpha_{\langle E,F \rangle} \in Competences$ - Hearer committed to $\texttt{assert}(\alpha_{\langle E,F \rangle}) \in \texttt{assert}$ - Hearer committed to X - Hearer committed to $X \subseteq Endorsers$ - Hearer committed to $S \in States$	- Speaker committed to deny resultant state exists(S)
Deny proposition exists(A)	- Speaker uttered enter dialogue - Hearer uttered enter dialogue - Hearer committed to R - Hearer committed to $R \in States$ - Hearer committed to $\alpha \in Opinions$ - Hearer committed to $E \in Experts$ - Hearer committed to $F \in Fields$ - Hearer committed to $\alpha_{\langle E,F \rangle} \in Competences$ - Hearer committed to $\texttt{assert}(\alpha_{\langle E,F \rangle}) \in \texttt{assert}$ - Hearer committed to X - Hearer committed to $X \subseteq Endorsers$ - Hearer committed to $S \in States$ - Hearer committed to $A \in Prop$	- Speaker committed to deny proposition exists(A)
Deny value exists(v)	- Speaker uttered enter dialogue - Hearer uttered enter dialogue - Hearer committed to R - Hearer committed to $R \in States$ - Hearer committed to $\alpha \in Opinions$ - Hearer committed to $E \in Experts$ - Hearer committed to $F \in Fields$ - Hearer committed to $\alpha_{\langle E,F \rangle} \in Competences$ - Hearer committed to $\texttt{assert}(\alpha_{\langle E,F \rangle}) \in \texttt{assert}$ - Hearer committed to X - Hearer committed to $X \subseteq Endorsers$ - Hearer committed to $S \in States$ - Hearer committed to $A \in Prop$ - Hearer committed to $v \in Values$	- Speaker committed to deny value exists(v)

List of Figures

List of Tables

Bibliography

[1] Open AI. Gpt-4 technical report. arXiv, 2023.

[2] Leila Amgoud and Claudette Cayrol. A reasoning model based on the production of acceptable arguments. In *Annals of Mathematics and Artificial Intelligence*, volume 34, pages 197–215. Springer, 2002.

[3] Charles Antaki and Ivan Leudar. Explaining in conversation: Towards an argument model. In *European Journal of Social Psychology*, volume 22, pages 181–194. Wiley Online Library, 1992.

[4] Aristotle. The Organon. Jazzybee Verlag, 2015.

[5] Katie Atkinson and Trevor Bench-Capon. Practical reasoning as presumptive argumentation using action based alternating transition systems. In *Artificial Intelligence*, volume 171, pages 855–874. Elsevier, 2007.

[6] Katie Atkinson, Trevor Bench-Capon, and Peter McBurney. A Dialogue Game Protocol for Multi-Agent Argument over Proposals for Action. In *Autonomous Agents and Multi-Agent Systems*, volume 11, pages 153–171. Springer, 2005.

[7] Katie Atkinson, Trevor Bench-Capon, and Peter McBurney. Computational Representation of Practical Argument. In *Synthese*, volume 152, pages 157–206. Springer, 2006.

[8] Katie Atkinson, Trevor Bench-Capon, and Sanjay Modgil. Value added: Processing information with argumentation. In *Second International Workshop on Argumentation in Multi-Agent Systems (ArgMAS 2005)*, 2005.

[9] John L Austin. How to do things with words. Oxford University Press, 1962.

[10] Panagiotis Balatsoukas, Isabel Sassoon, Martin Chapman, Nadin Kokciyan, Archie Drake, Sanjay Modgil, Mark Ashworth, Vasa Curcin, Elizabeth I Sklar, and Simon

Parsons. In the wild pilot usability assessment of a connected health system for stroke self management. In *2020 IEEE International Conference on Healthcare Informatics (ICHI)*, pages 1–3. IEEE, 2020.

[11] Panos Balatsoukas, Talya Porat, Isabel Sassoon, Kai Essers, Nadin Kokciyan, Martin Chapman, Archie Drake, Sanjay Modgil, Mark Ashworth, Elizabeth I Sklar, et al. User involvement in the design of a data-driven self-management decision support tool for stroke survivors. In *IEEE EUROCON 2019-18th International Conference on Smart Technologies*, pages 1–6. IEEE, 2019.

[12] Ernest Barker and Richard Stalley. Aristotle: Politics. Oxford: Oxford University Press, 1995.

[13] Victoria Bellotti and Keith Edwards. Intelligibility and accountability: human considerations in context-aware systems. In *Human–Computer Interaction*, volume 16, pages 193–212. Taylor & Francis, 2001.

[14] Trevor JM Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. In *Journal of Logic and Computation*, volume 13, pages 429–448. Oxford University Press, 2003.

[15] Trevor JM Bench-Capon and Paul E Dunne. Argumentation in artificial intelligence. In *Artificial intelligence*, volume 171, pages 619–641. Elsevier, 2007.

[16] Philippe Besnard and Anthony Hunter. Elements of argumentation. volume 47. MIT press Cambridge, 2008.

[17] Floris Bex and Douglas N Walton. Combining explanation and argumentation in dialogue. In *Argument & Computation*, volume 7, pages 55–68. IOS Press, 2016.

[18] Eda Bilici, George Despotou, and Theodoros N Arvanitis. The use of computer-interpretable clinical guidelines to manage care complexities of patients with multimorbid conditions: A review. In *Digital health*, volume 4. SAGE Publications Sage UK: London, England, 2018.

[19] Stefano Bistarelli, Lars Kotthoff, Francesco Santini, and Carlo Taticchi. Summary report for the third international competition on computational models of argumentation. In *AI Magazine*, volume 42, pages 70–73, 2021.

[20] Elizabeth Black and Anthony Hunter. A generative inquiry dialogue system. In *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*, pages 1–8. Association for Computing Machinery, 2007.

[21] Elizabeth Black, Nicolas Maudet, and Simon Parsons. Argumentation-based dialogue. In Dov Gabbay, Massimiliano Giacomin, Guillermo Simari , and Matthias Thimm, editors, *Handbook of Formal Argumentation, Volume 2*, page 511. College Publications, 2021.

[22] Andrei Bondarenko, Phan Minh Dung, Robert A Kowalski, and Francesca Toni. An abstract, argumentation-theoretic approach to default reasoning. In *Artificial intelligence*, volume 93, pages 63–101. Elsevier, 1997.

[23] Gerhard Brewka. Preferred subtheories: An extended logical framework for default reasoning. In *IJCAI*, volume 89, pages 1043–1048. International Joint Conferences on Artificial Intelligence Organization, 1989.

[24] Elena Cabrio and Serena Villata. Five years of argument mining: a data-driven analysis. In *IJCAI*, volume 18, pages 5427–5433. International Joint Conferences on Artificial Intelligence Organization, 2018.

[25] Martin Caminada. An algorithm for computing semi-stable semantics. In *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 222–234. Springer, 2007.

[26] Martin Caminada. A labelling approach for ideal and stage semantics. In *Argument and Computation*, volume 2, pages 1–21. Taylor & Francis, 2011.

[27] Martin Caminada and Leila Amgoud. On the evaluation of argumentation formalisms. In *Artificial Intelligence*, volume 171, pages 286–310. Elsevier, 2007.

[28] Martin Caminada, Walter Carnielli, and Paul Dunne. Semi-stable semantics. In *Journal of Logic and Computation*, volume 22, pages 1207–1254. OUP, 2012.

[29] Martin Caminada, Sanjay Modgil, and Nir Oren. Preferences and unrestricted rebut. In *Computational Models of Argument*, pages 209–220. IOS Press, 2014.

[30] Martin Caminada and Gabriella Pigozzi. On judgment aggregation in abstract argumentation. In *Autonomous Agents and Multi-Agent Systems*, volume 22, pages 64–102. Springer, 2011.

[31] Martin Caminada and Yining Wu. An argument game for stable semantics. In *Logic Journal of IGPL*, volume 17, pages 77–90. Oxford University Press, 2009.

[32] Federico Castagna. Argument games for dialectical classical logic argumentation. In *Online Handbook of Argumentation for AI*, pages 2–6, 2020.

[33] Federico Castagna. A dialectical characterisation of argument game proof theories for classical logic argumentation. In *International Workshop on Advances in Argumentation in Artificial Intelligence (AI^3)*, 2021.

[34] Federico Castagna. Labelling procedures for dialectical classical logic argumentation. In *Online Handbook of Argumentation for AI*, pages 7–11, 2021.

[35] Federico Castagna. Towards a fully-fledged formal protocol for the Explanation-Question-Response dialogue. In *Online Handbook of Argumentation for AI*, pages 17–21, 2022.

[36] Federico Castagna. Dialectical argument game proof theories for classical logic [Accepted]. In *Journal of Applied Logic (IfCoLog)*, 2023.

[37] Federico Castagna, Alexandra Garton, Peter Mcburney, Simon Parsons, Isabel Sassoon, and Elizabeth I Sklar. EQRbot: a chatbot delivering EQR argument-based explanations. In *Frontiers in Artificial Intelligence*, volume 6, page 39. Frontiers, 2023.

[38] Federico Castagna, Simon Parsons, Isabel Sassoon, and Elizabeth I Sklar. Providing explanations via the EQR argument scheme. In *Computational Models of Argument: Proceedings of COMMA 2022*, pages 351–352. IOS Press, 2022.

[39] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. On the acceptability of arguments in bipolar argumentation frameworks. In *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 378–389. Springer, 2005.

[40] Martin Chapman, Panagiotis Balatsoukas, Nadin Kökciyan, Kai Essers, Isabel Sassoon, Mark Ashworth, Vasa Curcin, Sanjay Modgil, Simon Parsons, and Elizabeth I Sklar. Computational argumentation-based clinical decision support. In *18th International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2019*, pages 2345–2347. International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS), Association for Computing Machinery, 2019.

[41] Günther Charwat, Wolfgang Dvořák, Sarah A Gaggl, Johannes P Wallner, and Stefan Woltran. Methods for solving reasoning problems in abstract argumentation–a survey. In *Artificial intelligence*, volume 220, pages 28–63. Elsevier, 2015.

[42] Carlos Iván Chesnevar, Ana Gabriela Maguitman, and Ronald Prescott Loui. Logical models of argument. In *ACM Computing Surveys (CSUR)*, volume 32, pages 337–383. ACM New York, NY, USA, 2000.

[43] Eva Cogan, Simon Parsons, and Peter McBurney. New types of inter-agent dialogues. In *International Workshop on Argumentation in Multi-Agent Systems*, pages 154–168. Springer, 2005.

[44] Marco Colombetti and Mario Verdicchio. An analysis of agent speech acts as institutional actions. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 3*, pages 1157–1164. Association for Computing Machinery, 2002.

[45] Kristijonas Cyras, Brendan Delaney, Denys Prociuk, Francesca Toni, Martin Chapman, Jesús Domínguez, and Vasa Curcin. Argumentation for explainable reasoning with conflicting medical recommendations. 2018.

[46] Kristijonas Čyras, Tiago Oliveira, Amin Karamlou, and Francesca Toni. Assumption-based argumentation with preferences and goals for patient-centric reasoning with interacting clinical guidelines. In *Argument & Computation*, volume 12, pages 149–189. IOS Press, 2021.

[47] Kristijonas Čyras, Antonio Rago, Emanuele Albini, Pietro Baroni, and Francesca Toni. Argumentative xai: a survey. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI-21) Survey Track*. International Joint Conferences on Artificial Intelligence, 2021.

[48] Marcello D'Agostino and Sanjay Modgil. A rational account of classical logic argumentation for real-world agents. In *ECAI*, pages 141–149. IOS Press, 2016.

[49] Marcello D'Agostino and Sanjay Modgil. Classical logic, argument and dialectic. In *Artificial Intelligence*, volume 262, pages 15–51. Elsevier, 2018.

[50] Marcello D'Agostino and Sanjay Modgil. A study of argumentative characterisations of preferred subtheories. In *IJCAI*, pages 1788–1794. AAAI Press, 2018.

[51] Marcello D'Agostino and Sanjay Modgil. A fully rational account of structured argumentation under resource bounds. In *International Joint Conference on Artificial Intelligence (IJCAI-20)*, pages 1841–1847. IJCAI, 2020.

[52] John Douglas and Simon Wells. Towards next generation, free & open-source, argument analysis tools. In *The 17th International Workshop on Computation Models of Natural Argument (CMNA17)*, 2017.

[53] Phan Minh Dung. On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and N-person Games. In *Artificial Intelligence*, volume 77, pages 321–357. Elsevier, 1995.

[54] Phan Minh Dung, Robert A Kowalski, and Francesca Toni. Assumption-based argumentation. In *Argumentation in artificial intelligence*, pages 199–218. Springer, 2009.

[55] Phan Minh Dung, Paolo Mancarella, and Francesca Toni. Computing ideal sceptical argumentation. In *Artificial Intelligence*, volume 171, pages 642–674. Elsevier, 2007.

[56] Phan Minh Dung, Francesca Toni, and Paolo Mancarella. Some design guidelines for practical argumentation systems. In *Computational Models of Argument*, pages 183–194. IOS Press, 2010.

[57] Catarina Dutilh Novaes. Argument and Argumentation. In *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Fall 2022 edition, 2022.

[58] Frans H van Eemeren and Rob Grootendorst. Argumentation, communication, and fallacies: A pragma-dialectical perspective. Lawrence Erlbaum Associates, Inc., 1992.

[59] Frans H van Eemeren, Rob Grootendorst, and Tjark Kruiger. Handbook of argumentation theory: A critical survey of classical backgrounds and modern studies. De Gruyter, 1987.

[60] Uwe Egly, Sarah Alice Gaggl, and Stefan Woltran. Aspartix: Implementing argumentation frameworks using answer-set programming. In *International Conference on Logic Programming*, pages 734–738. Springer, 2008.

[61] Kai Essers, Martin Chapman, Nadin Kokciyan, Isabel Sassoon, Talya Porat, Panagiotis Balatsoukas, Peter Young, Mark Ashworth, Vasa Curcin, Sanjay Modgil, et al. The consult system. In *Proceedings of the 6th International Conference on Human-Agent Interaction*, pages 385–386, 2018.

[62] Elliot Fairweather, Rudolf Wittner, Martin Chapman, Petr Holub, and Vasa Curcin. Non-repudiable provenance for clinical decision support systems. In *Provenance and Annotation of Data and Processes*, pages 165–182. Springer, 2020.

[63] Sarah A Gaggl, Thomas Linsbichler, Marco Maratea, and Stefan Woltran. Design and results of the second international competition on computational models of argumentation. In *Artificial Intelligence*, volume 279. Elsevier, 2020.

[64] Sarah A. Gaggl, Matthias Thimm, and Mauro Vallati. Proceedings of the third international workshop on systems and algorithms for formal argumentation (SAFA20). volume 2672. CEUR-WS.org, September 2020.

[65] Roderic A. Girle. Commands in dialogue logic. In Dov M. Gabbay and Hans Jürgen Ohlbach, editors, *Practical Reasoning*, pages 246–260. Springer Berlin Heidelberg, 1996.

[66] Nikos Gorogiannis and Anthony Hunter. Instantiating abstract argumentation with classical logic arguments: Postulates and properties. In *Artificial Intelligence*, volume 175, pages 1479–1497. Elsevier, 2011.

[67] Mitchell Green. Speech Acts. In *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, 2021.

[68] David Gunning, Mark Stefik, Jaesik Choi, Timothy Miller, Simone Stumpf, and Guang-Zhong Yang. Xai—explainable artificial intelligence. In *Science Robotics*, volume 4, page eaay7120. American Association for the Advancement of Science, 2019.

[69] Ivan Habernal and Iryna Gurevych. Argumentation mining in user-generated web discourse. In *Computational Linguistics*, volume 43, pages 125–179. MIT Press, 2017.

[70] Charles L Hamblin. Fallacies. Methuen, London, 1970.

[71] Seyed A Hosseini, Sanjay Modgil, and Odinaldo Rodrigues. Dialogues incorporating enthymemes and modelling of other agents' beliefs. In *PhD thesis*. King's College London, 2017.

[72] Joris Hulstijn. Dialogue models for inquiry and transaction. PhD thesis, Universiteit Twente, Enschede, The Netherlands, 2000.

[73] Amin Karamlou, Francesca Toni, Kristijonas Cyras, and Erisa Karafili. Complexity results, algorithms, and an implementation for bipolar argumentation. 2018.

[74] Nadin Kokciyan, Martin Chapman, Panagiotis Balatsoukas, Isabel Sassoon, Kai Essers, Mark Ashworth, Vasa Curcin, Sanjay Modgil, Simon Parsons, and Elizabeth I Sklar. A collaborative decision support tool for managing chronic conditions. In *The 17th World Congress of Medical and Health Informatics*, 2019.

[75] Nadin Kökciyan, Simon Parsons, Isabel Sassoon, Elizabeth I Sklar, and Sanjay Modgil. An argumentation-based approach to generate domain-specific explanations. In *Multi-Agent Systems and Agreement Technologies*, pages 319–337. Springer, 2020.

[76] Nadin Kökciyan, Isabel Sassoon, Elizabeth I Sklar, Sanjay Modgil, and Simon Parsons. Applying metalevel argumentation frameworks to support medical decision making. In *IEEE Intelligent Systems*, volume 36, pages 64–71. IEEE, 2021.

[77] Nadin Kokciyan, Isabel Sassoon, Anthony P Young, Martin Chapman, Talya Porat, Mark Ashworth, Vasa Curcin, Sanjay Modgil, Simon Parsons, and Elizabeth I Sklar. Towards an argumentation system for supporting patients in self-managing their chronic conditions. In *Workshops at the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[78] Erik CW Krabbe. Dialogue logic revisited. In *Aristotelian Society Supplementary Volume*, volume 75, pages 33–49. Oxford University Press Oxford, UK, 2001.

[79] Saul A Kripke. A completeness theorem in modal logic1. In *The journal of symbolic logic*, volume 24, pages 1–14. Cambridge University Press, 1959.

[80] Yannis Labrou, Tim Finin, and Yun Peng. Agent Communication Languages: The Current Landscape. In *IEEE Intelligent Systems and Their Applications*, volume 14, pages 45–52. IEEE, 1999.

[81] Jean-Marie Lagniez, Emmanuel Lonca, Jean-Guy Mailly, and Julien Rossit. Design and results of iccma 2021. In *arXiv preprint arXiv:2109.08884*, 2021.

[82] Helena Lindgren, Timotheus Kampik, Esteban Guerrero Rosero, Madeleine Blusi, and Juan Carlos Nieves. Argumentation-based health information systems: A design methodology. In *IEEE Intelligent Systems*, volume 36, pages 72–80. IEEE, 2020.

[83] Michael Luck, Peter McBurney, Onn Shehory, and Steve Willmott. Agent technology, computing as interaction: A roadmap for agent based computing. University of Southampton on behalf of AgentLink III, 2005.

[84] Christoph Lumer. An epistemological appraisal of walton's argument schemes. In *Informal Logic*, volume 42, pages 203–290. Érudit, 2022.

[85] Peter McBurney, David Hitchcock, and Simon Parsons. The eightfold way of deliberation dialogue. In *International Journal of Intelligent Systems*, volume 22, pages 95–132. Wiley Online Library, 2007.

[86] Peter McBurney and Simon Parsons. Chance discovery using dialectical argumentation. In *Annual Conference of the Japanese Society for Artificial Intelligence*, pages 414–424. Springer, 2001.

[87] Peter McBurney and Simon Parsons. Games that agents play: A formal framework for dialogues between autonomous agents. In *Journal of logic, language and information*, volume 11, pages 315–334. Springer, 2002.

[88] Peter McBurney and Simon Parsons. Retraction and revocation in agent deliberation dialogs. In *Argumentation*, volume 21, pages 269–289. Springer, 2007.

[89] Peter McBurney and Simon Parsons. Dialogue games for agent argumentation. In *Argumentation in artificial intelligence*, pages 261–280. Springer, 2009.

[90] Peter McBurney and Simon Parsons. Argument Schemes and Dialogue Protocols: Doug Walton's Legacy in Artificial Intelligence. In *Journal of Applied Logics*, volume 8, pages 263–286. College Publications, 2021.

[91] Peter McBurney, Simon Parsons, K Atkinson, H Prakken, and A Wyner. Talking about doing. In *From Knowledge Representation to Argumentation in AI, Law and Policy Making*, pages 151–166. College Publications, 2013.

[92] Peter McBurney, Rogier M Van Eijk, Simon Parsons, and Leila Amgoud. A dialogue game protocol for agent purchase negotiations. In *Autonomous Agents and Multi-Agent Systems*, volume 7, pages 235–273. Springer, 2003.

[93] Hugo Mercier and Dan Sperber. Why do humans reason? Arguments for an argumentative theory. In *Behavioral and brain sciences*, volume 34, pages 57–74. Cambridge University Press, 2011.

[94] Tim Miller. Explanation in artificial intelligence: Insights from the social sciences. In *Artificial intelligence*, volume 267, pages 1–38. Elsevier, 2019.

[95] Sanjay Modgil. Dialogical scaffolding for human and artificial agent reasoning. In *AIC*, pages 58–71. CEUR-WS, 2017.

[96] Sanjay Modgil and Trevor JM Bench-Capon. Metalevel argumentation. In *Journal of Logic and Computation*, volume 21, pages 959–1003. OUP, 2011.

[97] Sanjay Modgil and Martin Caminada. Proof theories and algorithms for abstract argumentation frameworks. In *Argumentation in artificial intelligence*. Springer, 2009.

[98] Sanjay Modgil and Henry Prakken. A general account of argumentation with preferences. In *Artificial Intelligence*, volume 195, pages 361–397. Elsevier, 2013.

[99] Samer Nofal, Katie Atkinson, and Paul E Dunne. Algorithms for decision problems in argument systems under preferred semantics. In *Artificial Intelligence*, volume 207, pages 23–51. Elsevier, 2014.

[100] Tiago Oliveira, Jérémie Dauphin, Ken Satoh, Shusaku Tsumoto, and Paulo Novais. Argumentation with goals for clinical decision support in multimorbidity. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, 2018.

[101] Jerome A Osheroff, Jonathan M Teich, Blackford Middleton, Elaine B Steen, Adam Wright, and Don E Detmer. A roadmap for national action on clinical decision support. In *Journal of the American medical informatics association*, volume 14, pages 141–145. BMJ Group BMA House, Tavistock Square, London, WC1H 9JR, 2007.

[102] Simon Parsons, Peter McBurney, Elizabeth I Sklar, and Michael Wooldridge. On the relevance of utterances in formal inter-agent dialogues. In *International Workshop on Argumentation in Multi-Agent Systems*, pages 47–62. Springer, 2007.

[103] John L Pollock. Defeasible reasoning. In *Cognitive science*, volume 11, pages 481–518. Elsevier, 1987.

[104] Henry Prakken. Coherence and flexibility in dialogue games for argumentation. In *Journal of logic and computation*, volume 15, pages 1009–1040. Oxford University Press, 2005.

[105] Henry Prakken. Formal systems for persuasion dialogue. In *The knowledge engineering review*, volume 21, pages 163–188. Cambridge University Press, 2006.

[106] Henry Prakken. An abstract framework for argumentation with structured arguments. In *Argument and Computation*, volume 1, pages 93–124. Taylor & Francis, 2010.

[107] Henry Prakken. On the nature of argument schemes. In *Dialectics, dialogue and argumentation. An examination of Douglas Walton's theories of reasoning and argument*, pages 167–185. College Publications, London, 2010.

[108] Raymond Reiter. A logic for default reasoning. In *Artificial intelligence*, volume 13, pages 81–132. Elsevier, 1980.

[109] Stuart J Russell and Peter Norvig. Artificial intelligence: A modern approach 3rd edition. Saddle River, NJ: Prentice Hall., 2009.

[110] Isabel Sassoon, Nadin Kökciyan, Sanjay Modgil, and Simon Parsons. Argumentation schemes for clinical decision support. In *Argument & Computation*, number Preprint, pages 1–27. IOS Press, 2021.

[111] Isabel Sassoon, Nadin Kökciyan, Elizabeth I Sklar, and Simon Parsons. Explainable argumentation for wellness consultation. In *International Workshop on Explainable, Transparent Autonomous Agents and Multi-Agent Systems*, pages 186–202. Springer, 2019.

[112] John Searle. Speech acts: An essay in the philosophy of language. Cambridge University Press, 1969.

[113] John Searle. A taxonomy of illocutionary acts. In *Language, mind, and knowledge*, pages 1–29. University of Minnesota Press, 1975.

[114] John Searle. Rationality in action. MIT press, 2001.

[115] Munindar P Singh. An ontology for commitments in multiagent systems. In *Artificial intelligence and law*, volume 7, pages 97–113. Springer, 1999.

[116] Elizabeth I Sklar and Mohammad Q Azhar. Argumentation-based dialogue games for shared control in human-robot systems. In *Journal of Human-Robot Interaction*, volume 4, pages 120–148, 2015.

[117] Elizabeth I Sklar and Simon Parsons. Towards the application of argumentation-based dialogues for education. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems*, volume 4, pages 1420–1421. IEEE Computer Society, 2004.

[118] Robin Smith. Aristotle's Logic. In Edward N. Zalta and Uri Nodelman, editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2022 edition, 2022.

[119] Mark Snaith, Rasmus Ø Nielsen, Sita R Kotnis, and Alison Pease. Ethical challenges in argumentation and dialogue in a healthcare context. In *Argument & Computation*, volume 12, pages 249–264. IOS Press, 2021.

[120] Reed T Sutton, David Pincock, Daniel C Baumgart, Daniel C Sadowski, Richard N Fedorak, and Karen I Kroeker. An overview of clinical decision support systems: benefits, risks, and strategies for success. In *NPJ digital medicine*, volume 3, pages 1–10. Nature Publishing Group, 2020.

[121] Matthias Thimm, Federico Cerutti, Hannes Strass, and Mauro Vallati. Proceedings of the first international workshop on systems and algorithms for formal argumentation (SAFA16). volume 1672. CEUR-WS.org, September 2016.

[122] Matthias Thimm, Federico Cerutti, and Mauro Vallati. Proceedings of the second international workshop on systems and algorithms for formal argumentation (SAFA18). volume 2171. CEUR-WS.org, September 2018.

[123] Matthias Thimm and Tjitze Rienstra. Approximate reasoning with ASPIC+ by argument sampling. In *SAFA@ COMMA*, pages 22–33, 2020.

[124] Matthias Thimm and Serena Villata. The first international competition on computational models of argumentation: Results and analysis. In *Artificial Intelligence*, volume 252, pages 267–294. Elsevier, 2017.

[125] Pancho Tolchinsky, Sanjay Modgil, Katie Atkinson, Peter McBurney, and Ulises Cortés. Deliberation dialogues for reasoning about safety critical actions. In *Autonomous Agents and Multi-Agent Systems*, volume 25, pages 209–259. Springer, 2012.

[126] Stephen E Toulmin. The uses of argument. Cambridge university press, 2003.

[127] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothee Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. Llama: Open and efficient foundation language models. arXiv, 2023.

[128] Qurat ul-ain Shaheen, Alice Toniolo, and Juliana KF Bowles. Argumentation-based explanations of multimorbidity treatment plans. In *International Conference on Principles and Practice of Multi-Agent Systems*, pages 394–402. Springer, 2020.

[129] Baptiste Vasey, Stephan Ursprung, Benjamin Beddoe, Elliott H Taylor, Neale Marlow, Nicole Bilbro, Peter Watkinson, and Peter McCulloch. Association of clinician diagnostic performance with machine learning–based decision support systems: a systematic review. In *JAMA network open*, volume 4, pages e211276–e211276. American Medical Association, 2021.

[130] Alexandros Vassiliades, Nick Bassiliades, and Theodore Patkos. Argumentation and explainable artificial intelligence: a survey. In *The Knowledge Engineering Review*, volume 36. Cambridge University Press, 2021.

[131] Serena Villata, Guido Boella, Dov M Gabbay, and Leendert Van Der Torre. Modelling defeasible and prioritized support in bipolar argumentation. In *Annals of Mathematics and Artificial Intelligence*, volume 66, pages 163–197. Springer, 2012.

[132] Gerard AW Vreeswik and Henry Prakken. Credulous and sceptical argument games for preferred semantics. In *European Workshop on Logics in Artificial Intelligence*, pages 239–253. Springer, 2000.

[133] Douglas Walton. Types of dialogue and burdens of proof. In *COMMA*, pages 13–24. IOS Press, 2010.

[134] Douglas N Walton. Burden of proof. In *Argumentation*, volume 2, pages 233–254. Kluwer Academic Publishers, 1988.

[135] Douglas N Walton. What is reasoning? What is an argument? In *The journal of Philosophy*, volume 87, pages 399–419. Journal of Philosophy Inc, 1990.

[136] Douglas N Walton. Plausible argument in everyday conversation. SUNY Press, 1992.

[137] Douglas N Walton. Argument structure: A pragmatic theory. University of Toronto Press Toronto, 1996.

[138] Douglas N Walton. Argumentation Schemes for Presumptive Reasoning. In *Lawrence Erlbaum Associates*, 1996.

[139] Douglas N Walton. Appeal to Expert Opinion: Arguments from Authority. In *Pennsylvania State University Press*, 1997.

[140] Douglas N Walton and Thomas F Gordon. Modeling Critical Questions as Additional Premises. In *Proceedings of the 8th International OSSA Conference*, 2011.

[141] Douglas N Walton and Erik CW Krabbe. Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning. SUNY Press, 1995.

[142] Douglas N Walton, Christopher Reed, and Fabrizio Macagno. Argumentation schemes. Cambridge University Press, 2008.

[143] Sarah Waterson, James A Landay, and Tara Matthews. In the lab and out in the wild: remote web usability testing for mobile devices. In *CHI'02 extended abstracts on Human factors in computing systems*, pages 796–797. Association for Computing Machinery, 2002.

[144] Simon Wells. The open argumentation platform (OAPL). In *Computational Models of Argument*, pages 475–476. IOS Press, 2020.

[145] Charles Willard. Authority. In *Informal Logic*, volume 12, pages 11–22, 1990.

[146] Andreas Xydis, Christopher Hampson, Sanjay Modgil, and Elizabeth Black. Enthymemes in dialogues. In *COMMA*, pages 395–402. IOS Press, 2020.

[147] Andreas Xydis, Christopher Hampson, Sanjay Modgil, and Elizabeth Black. Towards a sound and complete dialogue system for handling enthymemes. In *International Conference on Logic and Argumentation*, pages 437–456. Springer, 2021.