1

Thirteen Challenges in Formal and Computational Argumentation

Liuwen Yu, Leendert van der Torre, and Réka Markovich

ABSTRACT. In this chapter, we present thirteen challenges in formal and computational argumentation. They are organized around Dung's attack-defense paradigm shift. First, we describe four challenges pertaining to the diversity of argumentation. Then we discuss five challenges regarding the attack-defense paradigm shift. Finally, we discuss four challenges for computational AI argumentation arising after the paradigm shift. We illustrate these challenges using examples from machine ethics, AI & Law, decision-making, linguistics, philosophy, and other disciplines to illustrate the breadth of argumentation research. We end each challenge by presenting several open questions for further research.

1 Introduction

Argumentation means different things to different people. Even in the two volumes of the Handbook of Formal Argumentation, one can find a range of definitions, some focusing more on the formal aspects, others focusing more on the computational aspects. We believe that the thirteen challenges discussed in this chapter pertain to all these definitions or can be rephrased to adhere to all these definitions. Nevertheless, to clarify some of the issues we discuss, we present in Table 1 the definitions we will use in this chapter.

Moreover, as its title suggests, this chapter is particularly concerned with formal and computational argumentation as discussed in the Handbook of Formal Argumentation, the proceedings of the Computational Models of Argument (COMMA) conferences, the Argument and Computation journal, and the wider artificial intelligence (AI) literature on argumentation. In particular, whereas formal argumentation has developed as a branch of knowledge representation and reasoning, an essential part of AI, it now intersects with numerous disciplines, including natural language processing (NLP), and multiagent systems.

Therefore, when we refer to *argumentation* without further clarification, we mean *formal and computational argumentation*. When we specifically discuss formal argumentation only or computational argumentation only, we will make this explicit. Similarly, when referring to natural argumentation, we will do so explicitly.

To structure our discussion of these challenges, we use the attack-defense paradigm shift brought about by Dung [1995] as a pivotal point. In particular, Table 2 distinguishes between three groups of challenges. The first group is concerned with the background to this paradigm shift, the second group is

Term	Definition
Natural argumentation	Refers to the way humans naturally reason and communicate in everyday language, combining elements of linguistics, philosophy, and rhetoric.
Formal argumentation	A process of representing, managing and (sometimes) resolving conflicts.
Algorithmic argumentation	A step-by-step procedure or set of rules designed to perform a specific task or solve a particular argumentation problem.
Computational argumentation	Refers to the study and implementation of argumentation processes using computational methods. Involves theoretical and practical aspects of how argumentation can be modeled and executed by computers.
Argumentation technology	A computational approach incorporating argumentation reasoning mechanisms with other technologies, e.g., NLP, large language models (LLMs), distributed ledger technologies, etc.

Table 1: Five types of argumentation discussed in this chapter

concerned with the paradigm shift itself, and the third group is concerned with the consequences of this paradigm shift for computational argumentation.

For each challenge, we begin by presenting an "observation". Here, we mean an observation about the above-mentioned literature on argumentation, i.e., the Handbook on Formal Argumentation, the COMMA proceedings, the Argument and Computation journal, and the wider AI literature on argumentation.

Given the wide range of topics discussed in this literature, and the changes taking place due to technological developments such as LLMs, the observations we chose to focus on have had a big influence on the contents of this chapter. Other researchers in argumentation might make different observations and, as a result, would approach this chapter differently. Thus, this chapter reflects our personal interpretation of the literature on argumentation.

We use a diverse array of examples for illustrative purposes from areas such as decision-making, ethical and legal reasoning, and practical reasoning, and these are listed in Table 3. The selected examples cover a wide range of disciplines and issues, illustrating also the breadth of potential application domains for the techniques discussed in this chapter. We reuse examples across different challenges so that we can look at these examples from different angles. We end each challenge by presenting several open questions for further research.

We selected the topics judiciously, leaving out many topics we would have liked to cover but would have made this chapter too long. To provide the reader with some additional information, we complement these challenges with open research questions.

This chapter follows the structure of Table 2 and is organized as follows. Section 2 discusses natural argumentation approaches that were prevalent before the paradigm shift, identifying four key challenges. Section 3 focuses on the paradigm shift itself, acknowledging its contributions while also highlighting

Sections	Observations	Challenges
Section 2. Context of the attack-defense paradigm shift	1. Diversity in reasoning across disciplines	1.Connecting individual and collective reasoning
	2. Diversity of arguments in natural language	2. Understanding and generating arguments
	3. Diversity in modeling the process of argumentation	3. Conceptualizing argumentation
	4. Diversity of formal methods used in formal argumentation	4. Formalizing argumentation
Section 3. Attack-defense paradigm shift	5. Universality of attack	5. Creating argumentation frameworks and semantics
	6. Variety of nonmonotonic logics and game theory solution concepts	6. Representing nonmonotonic logics and solution concepts
	7. Inconsistent knowledge bases	7. Rationality postulates for defining a new logic
	8. Dialogue is based on agents, strategies, and games	8. Generalizing Dung's attack-defense paradigm for dialogue
	9. Balancing is based on support	9. Generalizing Dung's attack-defense paradigm for balancing
Section 4. Computational turn after the paradigm shift	10. Diversity of argumentation	10. Conducting a principle-based analysis of argumentation
	11. Compositional nature of argumentation	11. Designing efficient algorithms for argumentation semantics
	12. Human-level and human-centered argumentation	12. Explaining argumentation
	13. Argumentation technology	13. Integrating argumentation with technologies

Table 2: Thirteen observations and challenges discussed in this chapter

the challenges it presents to our community. Section 4 examines the consequences of the paradigm shift, including the transition towards computational AI argumentation, and provides a broader view of the community.

2 Context of the attack-defense paradigm shift

In this section, we describe the diversity of argumentation in the literature to position and appreciate the attack-defense paradigm shift. In Section 2.1, we introduce the diversity of reasoning across disciplines, in Section 2.2 we discuss the diversity in natural argumentation, in Section 2.3 we discuss three argumentation conceptualizations, and in Section 2.4, we discuss the diversity of formal methods used in this area. We discuss the challenges of: connecting individual and collective reasoning; understanding, analyzing, evaluating, and generating arguments; and conceptualizing and formalizing the argumentation

Example	**Discipline**	**Challenge Number(s)**
Jiminy ethical governor	Machine ethics	1, 2
Fatio dialogue protocol	Speech act theory	2
Dialogue between autonomous robot NS-4 and Spooner	Speech act theory, argumentation schemes	2
Child custody	AI & law	2, 3, 9
Scottish fitness lover, snoring professor	Knowledge representation and reasoning	4, 6
Untidy room	Neuro-symbolic AI	4
Bachelor vs. married	Knowledge representation and reasoning	7
Dialogue between accuser and suspect	AI & law	8
Murder at Facility C	AI & law	12
Intelligent Human-Input-Based Blockchain Oracle (IHiBO)	Computer science, financial markets, AI & law	13

Table 3: Examples, disciplines, and relevant challenges (see Table 2)

process.

2.1 Individual and collective reasoning

In this section, we discuss reasoning, its philosophical and mathematical foundations, and its use across many disciplines. From this perspective, we illustrate in Table 1 our definition of formal argumentation — representing, managing and (sometimes) resolving conflicts —using as an example the six layers of conflict addressed by the Jiminy ethical governor.

Russell and Norvig [2010] identify four schools of thought on AI — machines that: think like humans, act like humans, think rationally, and act rationally. We can interpret these four schools of thought from various perspectives:

Practical reasoning vs. theoretical reasoning Practical reasoning is oriented towards choosing a course of action on the basis of goals and knowledge of one's own situation, while theoretical reasoning is oriented towards finding reasons for determining that a proposition about the world is true or false [Walton, 1990].

Descriptive reasoning vs. prescriptive reasoning Descriptive reasoning aims to replicate human intelligence and behavior, while prescriptive reasoning aims to simulate decision-making and prescribe actions that align with ethics and laws.

In all these different kinds of reasoning, there could be individual reasoning and collective reasoning. This brings us to the distinction between intelli-

gent systems and multiagent systems across various disciplines. In the social sciences, the distinction between individual and collective reasoning is called the micro-macro dichotomy [Coleman, 1984]. Another prototypical example is the distinction between classical decision theory based on the expected utility paradigm and classical game theory [Savage, 1972]. Whereas classical decision theory is a kind of optimization problem (concerned with maximizing the agent's expected utility), classical game theory is a kind of equilibrium analysis [Nash Jr, 1950]. This leads to the following challenge.

Challenge 1. Connecting individual and collective reasoning.

One of the central topics in reasoning is how to handle conflict, whether it arises among beliefs (logical inconsistency) or choices of actions (practical conflicts). This is relevant both to the reasoning process of an individual agent and to interactions among multiple agents.

Example 2.1 illustrates a conflict from the perspective of a single agent.

Example 2.1 (Car accident dilemma in *I, Robot*). In the film *I, Robot*, Detective Del Spooner is driving when he has an accident, plunging his own car and another vehicle carrying a child into a river. An autonomous general-purpose humanoid robot, NS-4, is passing by and is faced with a conflict because it cannot save all the humans involved in the accident, i.e., the drivers and the child. NS-4 must make a descriptive analysis of the situation and follow prescriptive actions guided by ethical and legal considerations.

Example 2.2 illustrates a conflict from a multiagent perspective.

Example 2.2 (Continued from Example 2.1). Instead of seeing the conflict in terms of saving Spooner or saving the girl, a conflict that is faced by NS-4 only, we can consider it as a disagreement between two stakeholders: NS-4 and Spooner. NS-4 wants to save Spooner, while Spooner wants to save the girl.

NS-4 and Spooner might reach a consensus through a process known as judgment aggregation where they combine their individual judgments to arrive at a collective decision [Caminada and Pigozzi, 2011]. However, in the context of game theory, the goal is not always to resolve all conflicts but rather to understand the dynamics at play and sometimes *agree to disagree* [Aumann, 2016]. This concept, known as equilibrium analysis, allows for a situation where Spooner and NS-4 recognize their differing priorities and accept the disagreement without necessarily resolving the conflict.

In a multi-stakeholder setting, conflicts can be conceptualized and managed at various layers. The Jiminy architecture [Liao *et al.*, 2023] is an ethical governor that uses techniques from normative systems and formal argumentation to get moral agreements from multiple stakeholders. Each stakeholder has their own normative system. The Jiminy architecture combines norms into arguments, identifies their conflicts as moral dilemmas, and evaluates the arguments to resolve each dilemma whenever possible. In particular cases, Jiminy

decides which of the stakeholders' normative systems takes preference over the others.

Example 2.3 (The six layers of conflict of *I, Robot* in the Jiminy architecture).

Layer 1: conclusions only The conflict is based on the conclusions drawn by each stakeholder. NS-4 concludes it should save Spooner, while Spooner concludes that the girl should be saved. This layer represents a straightforward conflict of decisions without going into the underlying reasoning, possibly making it difficult to resolve.

Layer 2: assumptions and reasons Agents present their conclusions along with their assumptions and reasons. We refer to these conclusions together with the assumptions and reasons as arguments. Conflict resolution may involve formal argumentation techniques such as assigning attacks among arguments. The reason Spooner wants to save the girl is that she has a longer potential lifespan. That reason could be attacked by an argument from NS-4 that the girl has incurable cancer and therefore has a short lifetime.

Layer 3: combining normative systems This layer involves combining multiple normative systems into a single normative system. As a consequence, there may be new arguments built from the norms of distinct stakeholders, and the combined knowledge may be sufficient to reach a moral agreement. For example, NS-4 has information unknown to Spooner — the child's illness. By aligning their knowledge bases, they may reach an agreement to save Spooner instead.

Layer 4: context-sensitive meta-reasoning as ethical governors Jiminy considers the agents' norm preferences. These meta-norms are context-dependent norms that select the one stakeholder who has the most relevant expertise. Jiminy may decide that NS-4's preference takes precedence over Spooner's because NS-4 can get a more accurate evaluation of the imminent accident, leading to a more reasonable decision. This mechanism is comparable to those used in private international law [Markovich, 2019].

Layer 5: suspend decisions and observation Traditional conflict resolution often assumes that dilemmas must be addressed immediately. However, suspending a decision to allow for additional information to emerge can be beneficial in certain situations.

Layer 6: dialogue In this layer, stakeholders engage in a dialogue, attempting to persuade one other. Through structured communication, NS-4 and Spooner present their arguments, counterarguments, and justifications. The dialogue helps them explore each other's perspectives and can lead to a more informed and mutually agreeable resolution.

There are many other examples similar to the one in *I, Robot*. For instance, the *tunnel dilemma* and the *trolley dilemma* [Awad *et al.*, 2018] involve ethical decisions by autonomous vehicles and the question of who should decide how they respond in life-and-death situations. Another example is a smart speaker that passively *listens in* and stores voice recordings, acting like a surveillance device [Liao *et al.*, 2023]. Should it assist in the prevention of or investigation into crimes? This presents a moral dilemma involving household members, law enforcement agencies, and the manufacturer of the smart speaker. Which stakeholder should be alerted in such cases?

Jiminy explains the general problem of connecting individual and collective reasoning, and its relation to practical reasoning. Different mechanisms could be implemented in the Jiminy architecture to connect the two kinds of reasoning. For example, philosophical concepts such as the *veil of ignorance* [Rawls, 2001] and *Kantian imperative* [Kant, 1993] offer valuable perspectives. The veil of ignorance principle requires individuals to make decisions without knowledge of their own personal characteristics or societal position, thus promoting impartiality and fairness in collective decision-making. This aligns closely with the AI challenge of designing a system that makes unbiased decisions. Similarly, Kant's categorical imperative suggests that one should act only according to maxims that can be universally applied to build universally ethical and rational guidelines for behavior. Both principles emphasize the importance of considering the broader implications of actions, and they encourage integrating individual actions into collective norms and ethics.

In this section, we discussed the general challenge of connecting individual and collective behavior from the perspective of argumentation. We end this section by raising a number questions for further research.

1. In this section, we considered argumentation as a kind of reasoning, which raises the question: what kinds of reasoning count as argumentation, and what kinds of argumentation do not count as reasoning? More generally, what is the scope of argumentation?

2. Which kinds of reasoning cannot be handled by argumentation? For example, can causal or case-based reasoning be cast as a kind of argumentation [Bengel *et al.*, 2024; Roth and Verheij, 2004]? Can decision-making be regarded as a kind of argumentation [Amgoud, 2009]?

3. How does argumentation relate to other kinds of reasoning? For example, What are the distinctions and connections between reasoning as a cognitive activity and argumentation as a communicative practice [Walton, 1990]?

4. How should argumentation be used in a general theory of reasoning? For example, some articles refer to argumentation and negotiation [Sycara, 1990; Rahwan *et al.*, 2003; Van Laar and Krabbe, 2018], examining how

argumentation can be used as part of a negotiation, or how a negotiation can be seen as a kind of argumentation process.

5. How should argumentation be applied to legal and ethical reasoning? For example, how can argumentation facilitate structured discourse among agents to negotiate conflicting norms, particularly in multiagent legal proceedings where stakeholders argue for or against specific outcomes [Prakken and Sartor, 2015]?

2.2 Natural arguments

In this section, we discuss the diversity of natural argumentation, psychological analysis of natural (and formal) arguments, transitioning from NLP to foundation models and chatbots, understanding and generating arguments using foundation models, and the central role of questions in natural argumentation. We illustrate the latter using critical questions in argument schemes to find weaknesses and by asking *why* questions to obtain justifications according go the so-called Fatio dialogue protocol.

Natural argumentation refers to the way humans reason and communicate naturally in everyday language. It combines elements of linguistics, philosophy, and rhetoric. It is characterised by considerable diversity, with arguments taking various forms, styles, and contexts. Our aim is to avoid developing a separate technical understanding and generating arguments and argumentation with only a weak connection to how these concepts are understood in the humanities and related fields by both scholars and practitioners [Gordon, 2018].

In linguistics, researchers evaluate the diversity of natural language arguments and their role in human interaction with experimental methodologies [Gillioz and Zufferey, 2020]. These methodologies often include human-based techniques such as psychological experiments to verify linguistic intuitions [Schindler *et al.*, 2020] and determine sound arguments and standards of justification [Weinstock, 2006]. In AI, the focus shifts to modeling, formalizing, and automating the argumentation process, generating arguments in both natural and formal languages. However, evaluation in AI also relies heavily on human assessment. For instance, empirical cognitive experiments have been conducted by Cramer and Guillaume [2018a; 2018b; 2019] and Cerutti et al. [2021] to evaluate the connection between natural and abstract argumentation. Two main research questions often guide this evaluation: *Do the features shared by major argumentation semantics (e.g., simple reinstatement) correspond to genuine cognitive aspects of human reasoning? Which argumentation semantics are the best predictors of human evaluation of arguments?* Human evaluations of automatically generated text are also conducted to assess their performance [van der Lee *et al.*, 2021].

In recent years, the transition from NLP [Budzynska *et al.*, 2018] to argumentation-based chatbots [Black *et al.*, 2021] has been accelerated by advancements in foundation models, such as OpenAI's Generative Pre-trained Transformer (GPT) series. This leads us to the following challenge.

Challenge 2. Understanding and generating natural arguments.

Chatbots are conversational software that seek to understand input from human users and generate human-like responses [Black *et al.*, 2021]. In chatbot development, *questions* play a crucial role in enhancing the effectiveness of argumentation-based chatbots and building engaging conversations [McBurney *et al.*, 2021]. The use of *questions* allows chatbots to guide dialogues, challenge assertions, support critical thinking, and provide justifications. Below, we discuss two *question* mechanisms that could be embedded into chatbots — argumentation schemes with their associated critical questions, and justification-seeking questions defined by speech act theory.

Argumentation schemes are investigated by an approach developed in philosophy and rhetoric, representing stereotypical patterns of reasoning that are often informal or semi-formal, rarely fully formalized. Initially developed for teaching critical thinking, these schemes were systematized by Walton *et al.* [2008], who identified sixty-five basic schemes. Argumentation schemes involve the activity of critically evaluating viewpoints and the reasons given to support them. For this reason, every scheme has a corresponding set of critical questions to identify possible weak points, challenge the arguments, and evaluate their strengths. The mechanism of argumentation schemes and critical questions could potentially improve chatbots. Chatbots, or conversational agents like ChatGPT, are good at crafting human-like sentences [Alkaissi and McFarlane, 2023]. But they often present falsehoods as facts and exhibit inconsistent logic, and these can be difficult to detect. Users tend to follow the chatbot's logic when given ready-made answers. However, when chatbots pose questions, they prompt users to engage in deeper critical thinking and question their responses [Danry *et al.*, 2023], which fosters more realistic and reliable interactions.

For example, consider below the argumentation scheme for practical reasoning.

Major premise:	I have a goal G.
Minor premise:	Carrying out action A is a means to realize G.
Conclusion:	Therefore, I ought (practically speaking) to carry out this action A.

There are five critical questions:

CQ1 What other goals do I have that should also be considered even though they might conflict with G?

CQ2 Other than me bringing about A, what alternative actions should be considered that would also bring about G?

CQ3 From the solutions of me bringing about A and these alternative actions, which can be argued to be the most efficient?

CQ4 What grounds are there for arguing that it is possible for me to bring about A in practice?

CQ5 What other consequences of me bringing about A should be taken into account?

Example 2.4 illustrates the dialogue between NS-4 and Spooner based on the argumentation schemes of practical reasoning.

Example 2.4 (Dialogue between NS-4 and Spooner).

Spooner: Save the girl! That is the moral and ethical choice. She deserves the chance to live her life fully.

NS-4: What other goals do you have that might conflict with this one?

Spooner: My goal is to save the most vulnerable lives. There is no conflict.

NS-4: What alternative actions should be considered?

Spooner: Saving the girl should be the only course of action. It should have the highest priority.

NS-4: What is the most efficient choice?

Spooner: Saving the girl. She is lighter, so this course of action is more likely to succeed.

NS-4: What grounds are there for arguing that it is practically possible to save the girl?

Spooner: The girl's lighter weight makes her rescue quicker and less risky.

NS-4: What consequences should be considered?

Spooner: Saving the girl aligns with the duty to protect the vulnerable.

NS-4: Your argument is sound and aligns with ethical and practical considerations. I will save the girl.

Speech act theory, a subfield of pragmatics, studies how words are used not only to present information but also to carry out actions [Austin, 1975]. This theory has been formalized in the Foundation for Intelligent Physical Agents (FIPA) standards, which are widely used in computer science to facilitate communication between autonomous agents [FIPA, 2002]. The scheme allows multiple labels to be applied to one utterance since a single utterance can perform multiple actions in a dialogue [Kissine, 2013]. Such labels range from a few basic types such as assertions, questions and commands to more complex ones like promises and declarations [Searle, 1979]. One of the key features of speech acts, as opposed to physical actions, is that their main effects are on the mental

and interactional states of agents [Traum, 1999]. The attitudes of belief, desire and intention are familiar to agency theories [Georgeff *et al.*, 1999]. In the context of human-like chatbots, speech acts can be used to design interactions between the chatbot and the user [Hakim *et al.*, 2019]. Specifically, questions that seek justification are crucial as they prompt the chatbot to provide reasons and explanations, which not only enriches the interaction but also drives the conversation towards deeper engagement and understanding.

McBurney and Parsons [2004] proposed an interaction protocol called Fatio comprising of five locutions for argumentation which can be considered as a set of speech acts.

F1: assert (P_i, ϕ)**:** A speaker P_i asserts a statement ϕ. In doing so, P_i creates a dialectical obligation within the dialogue to provide a justification for ϕ if required subsequently by another participant.

F2: question (P_j, P_i, ϕ)**:** A speaker P_j questions a prior utterance of **assert**(P_i, ϕ) by another participant P_i and seeks a justification for ϕ. The questioner P_j creates no dialectical obligations.

F3: challenge (P_j, P_i, ϕ)**:** A speaker P_j challenges a prior utterance of **assert**(P_i, ϕ) by another participant P_i and seeks a justification for ϕ. P_j not only asks a question but also creates for himself a dialectical obligation to provide a justification for not asserting ϕ. For example, he must provide an argument against ϕ if questioned or challenged. Thus, challenge (P_j, P_i, ϕ) is stronger than question (P_j, P_i, ϕ).

F4: justify $(P_i, \Phi \vdash \phi)$**:** A speaker P_i, who had uttered **assert**(P_i, ϕ) and was then questioned or challenged by another speaker, is able to provide a justification $\Phi \in A$ for the initial statement ϕ by means of this locution.

F5: retract (P_i, ϕ)**:** A speaker P_i, who had uttered **assert**(P_i, ϕ) or **justify**$(P_i, \Phi \vdash \phi)$, can withdraw this statement with the utterance of **retract**(P_i, ϕ) or the utterance of **retract**$(P_i, \Phi \vdash \phi)$ respectively. This removes the earlier dialectical obligation on P_i to justify ϕ or $\Phi \vdash \phi$ if questioned or challenged.

Example 2.5 illustrates the dialogue between NS-4 and Spooner following the speech act.

Example 2.5 (A dialogue between NS-4 and Del Spooner).

Spooner: Saving the girl is the right choice. *(assert)*

NS-4: Why? *(question)*

Spooner: Because the girl is young and has a much longer lifespan. *(justify)*

NS-4: I disagree that she has a much longer lifespan. *(challenge)*

Spooner: Why? *(question)*

NS-4: I conducted a health evaluation and found that she has a terminal disease. *(justify)*

In this section, we discussed the common understanding of natural arguments and how they are generated. We end with some open research questions.

1. In natural argumentation, we often encounter fallacies [Hamblin, 1970], and on the internet we encounter fake news [Visser *et al.*, 2020]. How should fallacies and fake news be handled in argumentation-based chatbots? For example, how can argumentation schemes be used in a chatbot to evaluate if an argument is fallacious [Walton, 2013]?

2. Programming has been replaced by prompt engineering for interacting with LLMs [Ross *et al.*, 2023]. How can argument schemes and speech act theory be used in prompt engineering? For example, argument schemes can potentially help structure for the generation of well-formed arguments [Musi and Palmieri, 2022].

3. An increasing number of arguments on the internet are generated by AI [Hinton and Wagemans, 2023]. How should AI-generated arguments be evaluated? The evaluation could focus on criteria like logical coherence, the relevance and sufficiency of the evidence, adherence to ethical principles, and the impact of the argument on the intended audience. Additionally, metrics could be developed to assess how well AI arguments handle counterarguments and whether they respect the norms of constructive and respectful discourse.

4. Argumentation has been discussed as a key component of chatbots [Castagna *et al.*, 2024a]. Which domain applications are argumentation-based chatbots suitable for? For example, how can arguments be used in AI therapy?

5. In the previous section, we introduced the Jiminy architecture. How can foundation models, argument schemes, and speech act theory be used to improve or enrich the Jiminy architecture?

2.3 Models of argument

In this section, we discuss three conceptualizations of argumentation — argumentation as inference, argumentation as dialogue, and argumentation as balancing. Each conceptualization embodies: a unique perspective on the construction and purpose of argumentation, a set of formal methods, and application across different disciplines and contexts. As mentioned in the introduction, we view argumentation as representing, managing and sometimes resolving conflicts. We explain how this key idea becomes more concrete with the three conceptualizations.

For argumentation as inference [Prakken, 2018], we consider: coherent positions in cases of conflict, what follows from each coherent position (or what we can infer from all coherent positions), and what can be agreed upon in cases of disagreement. For argumentation as dialogue [Prakken, 2018], we also consider the stakeholders that may hold such coherent positions and how they might interact, for example as proponents and opponents in a debate. This can clarify the conflict that is being managed and sometimes even help to resolve it. In such dialogues, the concerns or goals of the stakeholders can also be made explicit. As in dispute resolution, the process becomes very important. Finally, in argumentation as balancing [Gordon, 2018], we consider conflicts as trade-offs involving taking into account various pros and cons. Here, the central metaphor, referred to frequently in the law and ethical decision-making, is a pair of scales. For fine-grained comparisons, it is not uncommon to use quantitative metrics. In this section, we discuss the following challenge.

Challenge 3. Conceptualizing argumentation.

Table 4 provides a detailed comparison of the three conceptualizations. Notably, the applications of these conceptualizations are neither mutually exclusive nor incompatible. The formal methods are discussed in Section 2.4.

Conceptualization	**Process**	**Theories and Formal Approaches**	**Application**
Argumentation as inference	Logical structure and reasoning to derive conclusions from incomplete and inconsistent premises	Graph theory, nonmonotonic logic, computational logic, causal reasoning, Bayesian reasoning	Automated reasoning systems, knowledge representation, expert systems
Argumentation as dialogue	Dynamic verbal interaction between stakeholders to exchange information or resolve conflicts of opinion	Speech act theory, game theory, axiomatic semantics, operational semantics	Debating technologies, chatbots, recommender systems
Argumentation as balancing	Balancing pros and cons to reach a justified decision	Multi-criteria decision theory, machine ethics, computational law, case-based reasoning	Deliberative decision-making in law, ethics, and economics

Table 4: Conceptualizations of argumentation

Argumentation as inference focuses on determining the conclusions that can be derived from a given body of information, which may be incomplete, inconsistent, or uncertain. Relevant systems ultimately define a nonmonotonic notion of logical consequence in terms of the intermediate notions of argument construction, argument attack, and argument evaluation, and the arguments are seen as constellations of premises, con-

clusions, and inferences [Prakken, 2018]. These systems employ formal methods like nonmonotonic logic for commonsense reasoning, graph theory, computational logic, causal reasoning, and Bayesian reasoning. Argumentation as inference is primarily applied in knowledge representation and reasoning.

Argumentation as dialogue conceptualizes argumentation as a form of verbal interaction aimed at resolving conflicts of opinion [Prakken, 2018]. Relevant systems define argumentation protocols, which serve as the rules of the argumentation game, and they address strategic aspects that guide effective engagement in the game. The exploration of strategies involves understanding how to engage in productive discourse and present arguments effectively. Argumentation as dialogue utilizes speech act theory, game theory, axiomatic semantics, and operational semantics. It is most suitable for debates, chatbots, persuasion systems, negotiation systems, etc.

Argumentation as balancing involves weighing the pros and cons of an issue in order to reach a balanced decision or judgment. It is applicable not only when resolving conflicts of opinion in persuasion dialogues but also, e.g., when deciding courses of action in deliberation dialogues [Gordon, 2018]. In such a system, pro and con arguments for alternative resolutions of the issues (options or positions) are put forward, evaluated, resolved, and balanced. The formal methods used are multi-criteria decision theory, machine ethics, computational law, and case-based reasoning, and they are applied in the realms of law, ethics, and economics.

Table 4 might give the impression that the three approaches are distinct and that they have distinct application areas. We would like to point out that this is not the case. The approaches (or types) of argumentation are not mutually exclusive or even incompatible. You can switch from one to another if you want to look at the same problem or situation from different angles, highlight different aspects, or select a modeling approach that is more suitable for a particular purpose. This also means that complex application areas like the legal domain can make very good use of each approach. Indeed, legal reasoning often engages with each of the three conceptualizations — argumentation as inference, dialogue, and balancing — across different contexts and legal roles. Judges and attorneys may rely on one form of argumentation more than another, depending on the nature of the case and their specific role in the legal process. For instance, inference is commonly used by judges, attorneys, actually any type of lawyer, when applying legal rules to facts or deriving conclusions from incomplete or inconsistent premises. Dialogue plays a central role in courtroom exchanges between opposing parties. The structure of a trial often resembles a dialogue: each party presents their arguments and responds to those of the other while the judge oversees the process to ensure it follows legal procedures.

Balancing is typically the domain of judges as they weigh multiple factors, conflicting interests, or values, to determine the most appropriate outcome. This is particularly important in discretionary decision-making where the law, instead of trying to provide detailed rules, assigns special power to judges so that they can make decisions based on their own evaluations. In such cases, judges exercise their judicial discretion by carefully balancing competing considerations within the framework of legal principles to reach a fair and just decision.

Hence, these different modes of reasoning can correspond to and interact with one another, creating a comprehensive tool set for legal reasoning and decision-making. Below, we shall illustrate each of the three conceptualizations using the legal example of child custody in a divorce case.

Research on argumentation-based dialogue (see the overview of Black *et al.* [2021]) is often carried out against the background of the six types of dialogue and in accordance with their respective goals [Walton and Krabbe, 1995], as shown in Table 5. When argumentation is viewed as a kind of dialogue between multiple agents (whether human or artificial), new issues arise. One issue is the distributed nature of information (among the agents). Another issue is the dynamic nature of information — agents do not reveal everything they believe initially, and they can learn from one other. There are also strategic issues — agents will have their own internal preferences, desires and goals [Prakken, 2018]. In Section 2.2, we described the speech act theory on dialogue formation [McBurney and Parsons, 2002]. For better comparison, we use a legal child custody case [Yu *et al.*, 2020] to illustrate the three conceptualizations.

Type of dialogue	**Initial situation**	**Participant goal**	**Dialogue goal**
Persuasion	Conflict of opinions	Persuade other party	Resolve or clarify issue
Inquiry	Need to have proof	Find and verify evidence	Prove (or disprove) hypothesis
Negotiation	Conflict of interests	Get what they most want	Reasonable settlement they can both live with
Information-seeking	Need information	Acquire or give information	Exchange information
Deliberation	Dilemma or practical choice	Co-ordinate goals and actions	Decide best available course of action
Eristic	Personal conflict	Verbally hit out at opponent	Reveal deeper basis of conflict

Table 5: Types of dialogue [Walton and Krabbe, 1995]

Example 2.6 (Child custody dialogue). Alice and Lucy are talking about a divorce case, specifically whether it is in the child's best interest to live with her mother or with her father. They have the following dialogue.

Alice: It is in the ten-year-old child's best interest that she lives with her mother. (*assert*)

Lucy: Why? (*question*)

Alice: Because the child wants to live with her mother and the civil code states that the judge must take the child's opinion into account. (*justify*)

Lucy: A ten-year-old child does not know what she wants. (*challenge*)

Alice: Why? (*question*)

Lucy: Public opinion says that ten-year-old children do not know what they want. (*justify*)

Alice: Most ten-year-old children do know what they want. (*assert*)

Lucy: Why do you say that? (*question*)

Alice: Peter is a child psychologist, and Peter says that most ten-year-old children know what they want. (*justify*)

Most of the literature in this area is concerned with argumentation as inference. Some formal work had already had been carried out on argumentation-based inference before the publication of Dung's 1995 paper, notably the extensive research by Pollock [1987; 1992; 1994; 1995; 2001; 2009; 2010] on argument structure, the nature of defeasible reasons, the interplay between deductive and defeasible reasons, rebutting versus undercutting defeat, argument strength, argument labeling, self-defeating arguments, etc. Pollock identified reasoning as a process of constructing arguments where reasons provide the atomic links in arguments [Pollock, 1992]. He distinguished between two kinds of reasons: defeasible (prima facie) reasons and nondefeasible (conclusive) reasons [Pollock, 1987]. Nondefeasible reasons are those reasons that logically entail their conclusions while defeasible reasons may be destroyed with additional information. There are two kinds of defeaters that can defeat defeasible reasons. Rebutting defeaters deny the conclusion. Undercutting defeaters attack the connection between the reason and the conclusion. Pollock [1992; 1994; 1995] used so-called inference graphs to represent arguments and the nodes represented the steps of inference. There are three kinds of arrows in the inference graph, and they represent defeasible inferences, deductive inferences, and defeat links [Pollock, 1994].

Example 2.7 (Child custody in an inference graph). The dialogue between Alice and Lucy can also be illustrated in the format of Pollock's inference graph, as shown in Figure 1. Figure 1 illustrates two arguments rebutting the two opposite conclusions *"It is in the child's best interest that she lives with her father"*, and *"It is not in the child's best interest that she lives with her father"*. An undercutting argument, *"Public opinion is not reliable"*, defeats the argument *"Most ten-year-old children do not know what they want"*. In this figure, nondefeasible and defeasible inferences are visualized respectively with solid and dotted lines (without arrowheads). The arrows are defeat relations.

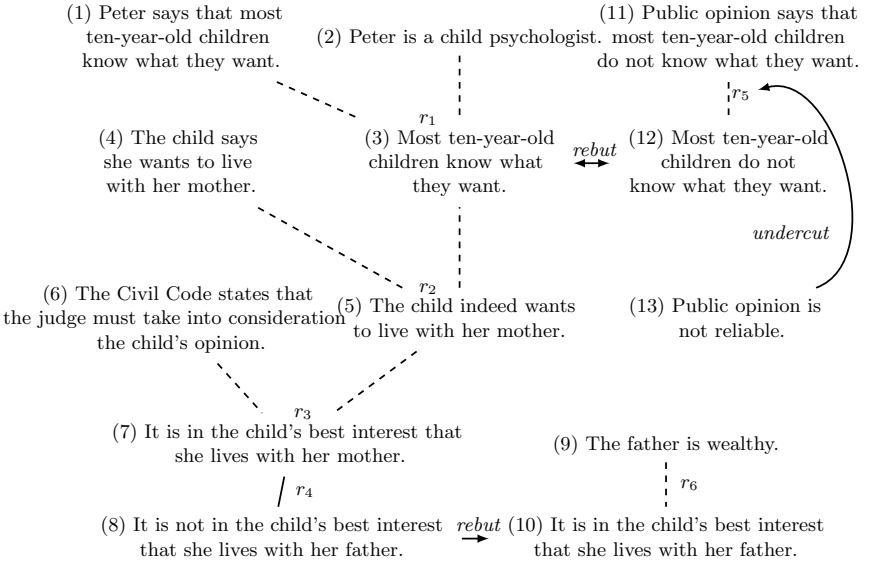


Figure 1: A dialogue about a child custody case represented as a Pollock inference graph. The solid and dotted lines (without arrowheads) are nondefeasible and defeasible inferences respectively. The arrows are defeat relations.

One model of argumentation as balancing is the Carneades Argumentation System [Gordon *et al.*, 2007]. The conception of argument graphs in Carneades is similar to Pollock's conception of an inference graph. There are nodes in the graph representing statements (propositions) and links that indicate inference relations between statements. In particular, the system distinguishes between pro and con arguments. Semantically, con arguments are instances of presumptive inference rules for negating the conclusion. If the premises of a con argument hold, this justifies rejecting the conclusion or, equivalently, accepting its logical complement. With pro and con arguments, some statements need to be ordered or otherwise aggregated to resolve the conflict. Then there are several proof standards used to balance the pros and cons. Here are three examples:

SE (Scintilla of Evidence): A statement meets this standard iff it is supported by at least one defensible pro argument.

BA (Best Argument): A statement meets this standard iff it is supported by some defensible pro argument with priority over all defensible con arguments.

DV (Dialectical Validity): A statement meets this standard iff it is supported by at least one defensible pro argument and none of its con arguments are defensible.

Example 2.8 (Child custody in Carneades). We represent part of the child custody example in Carnedes, as visualized in Figure 2. Statements are depicted as boxes and arguments as circles. For the purpose of this discussion, we assume that all the premises are ordinary without distinguishing between different types of premises. Premises are shown as edges without arrowheads. Pro arguments are indicated by circle arrowheads while con arguments are shown with standard arrowheads. Argument a_1 asserts that the child knows what she wants and she wants to live with her mother, making it a pro argument for the statement "It is in the child's best interest that she lives with her mother". In contrast, argument a_2 argues that the mother is less wealthy than the father, serving as a con argument against that statement. In this scenario, a_1 is given priority over a_2. Consequently, according to the BA proof standard, the statement "It is in the child's best interest that she lives with her mother" is accepted.

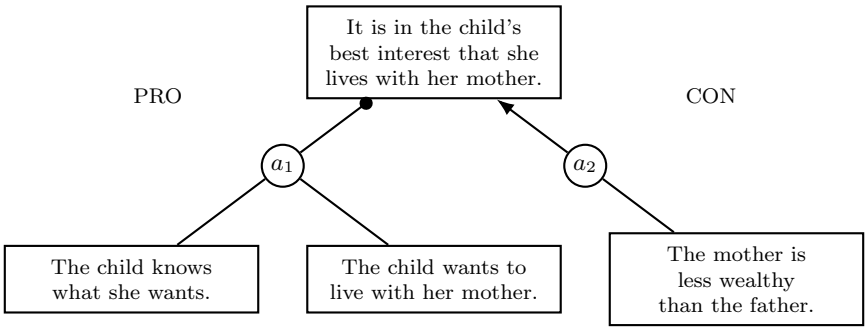


Figure 2: Child custody case represented in Carneades argument graphs

Argumentation as inference, argumentation as dialogue, and argumentation as balancing are distinct conceptualizations but they are not incompatible. A crucial question is how to move between individual reasoning (argumentation as inference and argumentation as balancing) and collective reasoning (argumentation as dialogue). Argumentation as inference can occur within an individual's mind, drawing upon a single knowledge base. However, in multiagent dialogues, each agent operates with a distinct and dynamic knowledge base. Agents employ strategic moves requiring them to learn about and understand other participants and to select or generate arguments from their knowledge base to achieve the goal of the dialogue. Some works attempt to integrate these different conceptualizations. For instance, discussion games serve as a proof procedure for abstract argumentation semantics [Caminada, 2018a], and multiagent argumentation considers agents with varying attitudes and knowledge [Arisaka *et al.*, 2022]. Moreover, several new agent argumentation semantics are inspired by social choice theory [Yu *et al.*, 2021; Baumeister *et al.*, 2021], and Carneades models legal dialogue by using critical questions from argumentation schemes while incorporating balancing to model

judgments [Gordon *et al.*, 2007].

In this section, we discussed argument models. We end again with some open questions:

1. Are there any other conceptualizations of argumentation that should be considered in the argumentation literature?
2. How should it be decided which conceptualization of argumentation to use for an application? In particular, when should argumentation as inference, dialogue, or balancing be used?
3. How are these conceptualizations related and should they be combined? For example, how should agent interaction and dialogue [Arisaka *et al.*, 2022] be introduced into Pollock's theory?
4. How should a general framework for argumentation as dialogue be designed? The formal study of argumentation-based dialogue is less substantial than the formal study of argumentation-based inference. It largely consists of a variety of different approaches and individual systems, with few unifying accounts or general frameworks [Prakken, 2018].
5. How should argumentation as balancing be represented formally? Compared with argumentation as inference and dialogue, there is little formal work on argumentation as balancing. Some examples are bipolar argumentation [Cayrol and Lagasquie-Schiex, 2013; Cayrol *et al.*, 2021; Yu *et al.*, 2020; Yu *et al.*, 2023], and an investigation into balancing operations [Knoks and van der Torre, 2023; Knocks *et al.*, 2024].

2.4 Formalizing argumentation

In this section, we discuss the large variety of formal methods in formal argumentation and their combination in applications and case studies. We show the use of nonmonotonic logic for commonsense reasoning, the integration of axiomatic and operational semantics in the Fatio dialogue system, and a combination of different reasoning methods with an example of a mother reasoning about her daughter.

Each conceptualization of argumentation comes with its own set of formal methods, as shown in Table 4 in Section 2.3. Basically, argumentation as inference uses most of the methods from graph theory (e.g., abstract argumentation [Dung, 1995]) or methods from nonmonotonic and computational logic (e.g., [Pollock, 1987; Nute, 1994; Reiter, 1980]) or causal reasoning [Giunchiglia *et al.*, 2004; Turner, 2004]. Argumentation as dialogue involves speech act theory (e.g., [FIPA, 2002]), axiomatic semantics, operational semantics, and denotational semantics, as well as game theory methods (e.g., [Moore, 1993]). In argumentation as balancing, there are methods from multi-criteria decision theory (e.g., [Amgoud, 2005; Amgoud and Vesic, 2012; Knoks and van der Torre, 2023; Knocks *et al.*, 2024]), ethical theory, legal theory, and case-based reasoning (e.g., [Yu and Gabbay, 2022]). In this section, we discuss that challenge.

Challenge 4. Formalizing argumentation.

Nonmonotonic logic was motivated by the fact that commonsense reasoning often involves incomplete or inconsistent information, in which case logical deduction is not a particularly useful reasoning model [Toulmin, 1958]. Classical logic is characterized by its monotonic nature. It asserts that if a set of statements S entails a proposition ϕ (denoted $S \vdash \phi$), any superset S' of S also entails ϕ. This principle underpins traditional logical proofs where lemmas remain valid while new premises are added. However, commonsense reasoning often allows conclusions to be retracted in the light of new information. For instance, the inference "Tweety, a bird, flies" (symbolized as $|\sim$) may be valid until additional contextual information like "Tweety is a penguin" necessitates the retraction of the initial conclusion. This approach mirrors our everyday reasoning processes, which frequently involve default rules with exceptions (e.g., $a \rightarrow x$ for strict rules and $a \Rightarrow x$ for default rules). Default rules apply unless there is evidence to the contrary requiring us to revoke our conclusions upon encountering such exceptions.

Here are some examples of the sources of nonmonotonicity [Prakken, 2010b]:

Empirical generalizations: e.g., adults are usually employed, birds can typically fly, etc.

Exceptions to legal rules: e.g., when a father dies, the child inherits, *except* when the child killed the father.

Exceptions to moral principles: e.g., one should not lie, *except* when a lie can save lives.

Conflicting information sources: experts who disagree, witnesses who contradict each other, conflicting sensory input, etc.

Alternative explanations: e.g., the grass is wet so it must have rained, but the sprinkler was on.

Conflicting reasons for actions: if we have a reason to do something, we should do it, unless we have good reasons for not doing it.

Prioritized default logic (PDL) [Brewka, 1994] is one formalization of nonmonotonic reasoning. A knowledge base in PDL contains prioritized defaults $a \overset{n}{\Rightarrow} b$ and facts, including tautologies. The notation $a \overset{n}{\Rightarrow} b$ means "if a, then normally b", with n indicating the priority level; a higher n implies a higher priority for the default rule $a \Rightarrow b$. PDL operates by selecting sets of defaults and bringing their conclusions into extensions. At each step, the default rule with the highest priority among the unapplied default rules is applied, while consistency is maintained.

Example 2.9 (Fitness lover Scot). Assume we have the following defaults and facts:

$$\left\{\begin{array}{c} \text{Defaults} \\ \hline BornInScotland \stackrel{1}{\Rightarrow} Scottish \\ Scottish \stackrel{3}{\Rightarrow} LikesWhisky \\ FitnessLover \stackrel{2}{\Rightarrow} \neg LikesWhisky \end{array}\right\} \quad \left\{\begin{array}{c} \text{Facts} \\ \hline \text{BornInScotland, FitnessLover} \end{array}\right\}$$

We can obtain the extension iteratively.

$$\begin{aligned} E_1 &= \{\text{BornInScotland}, \text{FitnessLover}\} \\ E_2 &= \{\text{BornInScotland}, \text{FitnessLover}, \neg\text{LikesWhisky}\} \\ E_3 &= \{\text{BornInScotland}, \text{FitnessLover}, \neg\text{LikesWhisky}, \text{Scottish}\} \\ E_4 &= E_3 \end{aligned}$$

We introduced the speech acts of Fatio [McBurney and Parsons, 2002] in Section 2.2, and we showed how reasons are used in a dialogue by 'question' and 'justify' moves in Section 2.3. We now reference Fatio to show how a dialogue system can make use of various formal methods: in this case, *axiomatic semantics* and *operational semantics*.

An axiomatic semantics for a programming language defines a set of axioms that the language obeys such as the pre-conditions and post-conditions for each command [Tennent, 1991]. It defines pre-conditions and post-conditions for the locutions. In Fatio, the axiomatic semantics concerns the beliefs and desires of the participating agents, which are written as $B_i\varphi$: "Agent i believes that φ is true", and $D_i\varphi$: "Agent i desires that φ be true". Central to the axiomatic semantics is a publicly viewable store to record the dialectical obligations of the participants, which is called a dialectical obligations store (DOS). The triple $(P_i, \varphi, +) \in \text{DOS}(P_i)$ denotes that participant P_i has a dialectical obligation to provide a justification or an argument in support of proposition φ, while the triple $(P_i, \varphi, -) \in \text{DOS}(P_i)$ denotes that participant P_i has a dialectical obligation to provide a justification or an argument against proposition φ.

For illustration, we list the pre- and post-conditions for *assert* and *question*, and we refer the rest of the axiomatic semantics to the original paper [McBurney and Parsons, 2004].

***assert*(P_i, φ)** *Pre-condition*: Speaker P_i wants each participant $P_j (j \neq i)$ to believe that P_i believes the proposition $\varphi \in C$.

$$((P_i, \varphi, +) \notin \text{DOS}(P_i)) \wedge (\forall j \neq i)(D_i B_j B_i \varphi).$$

Post-condition: Each participant $P_k (k \neq i)$ believes that participant P_i wants each participant $P_j (j \neq i)$ to believe that P_i believes φ.

$$((P_i, \varphi, +) \in \text{DOS}(P_i)) \wedge (\forall k \neq i)(\forall j \neq i)(B_k D_i B_j B_i \varphi).$$

Dialectical obligation: $(P_i, \varphi, +)$ is added to $\text{DOS}(P_i)$, the dialectical obligations store of speaker P_i.

question(P_j, P_i, φ) *Pre-condition*: One participant $P_i(i \neq j)$ has a dialectical obligation to support φ and participant P_j wants every other participant $P_k(k \neq j)$ to believe that P_j wants P_i to utters a *justify*$(P_i, \varphi, .)$ locution.

Post-condition: Participant P_i must utter a *justify* locution.

Dialectical obligation: No effect.

Operational semantics in Fatio is defined from a traditional computer science perspective. That means that the state of the system changes as a result of executions of commands in a programming language. To ensure automated generation of agent dialogues, participants need mechanisms to invoke specific utterances at appropriate points in the dialogue, and these mechanisms are called *agent decision mechanisms*. In this case, the commands in question are the locutions in an argumentation dialogue conducted according to the rules of the protocol. In Fatio, for example, an agent can decide whether to **Claim or Not**, whether to **React or Not**, whether to **Fold or Not**, whether to **Defend or Not** and, as a meta-level decision mechanism, whether to **Listen or Do**. There are transition rules defined for Fatio's operational semantics, and they assume that agents are equipped with decision mechanisms to initiate and respond to utterances. This enables the system to initiate utterances and respond to utterances in the dialogue, and so the states we will take to be the inputs and outputs of these decision mechanisms reflect that process.

One possible extension to the Fatio protocol is an additional semantics called *denotational semantics*, described but not explicitly defined by McBurney and Parsons [2004]. *It would link the utterances made under the protocol to the nodes and edges of a graph representing the arguments created by the participants in the course of a dialogue. This kind of graph would be similar to the argumentation graph constructed in Thomas Gordon's [1993] Pleadings Game, which is a formal structure capturing the flow and relationships between different arguments in a dialogue or argumentation context.* It would thus provide a mathematical structure to the dialogues, mapping the linguistic constructs (utterances) to a formal representation (graph) that captures the logical relationships and the dynamics of the argumentation process. This could be a way to visualize and analyze the structure of the arguments and the interplay between different participants' statements and responses in a dialogue.

Lastly, we give another example where various formal methods are combined. Neurosymbolic AI [Garcez *et al.*, 2008] combines neural and symbolic AI architectures to address the weaknesses of each, providing a robust AI system capable of reasoning, learning, and cognitive modeling. This diversity of formal methods brings many challenges to the area of formal argumentation. Consider Example 2.10 from Gabbay and Rivlin [2017].

Example 2.10 (Untidy room [Gabbay and Rivlin, 2017]). A mother goes into her teenage daughter's bedroom. Her instant impression is that it is a big mess. There is stuff scattered everywhere. The mother's feeling is that it is not like her daughter to be like this. What happened?

Conjecture: The girl may be experiencing boyfriend issues.

Further Analysis: The mother notices a collapsed shelf and realizes that the disarray is due to the shelf collapsing under excessive weight which, upon reflection, follows a logical (gravitational) pattern.

Several types of reasoning are illustrated through this scenario:

Neural network reasoning: The mess is perceived instantly, similarly to facial recognition by neural networks.

Nonmonotonic deduction: The mother deduces from the context and her knowledge that her daughter does not typically live in disarray. Thus, something extraordinary must have happened.

Abductive reasoning: She hypothesizes a plausible explanation that her daughter has social-emotional issues, which is common among teenagers.

Database AI deduction: A reevaluation leads to the understanding that the mess is due to gravitational effects rather than disorganization on the part of her daughter.

Pattern recognition: Someone accustomed to similar patterns may identify the cause as easily as they might recognize a face.

In practical reasoning, it is crucial to combine various formal methods. To deal with scenarios similar to Example 2.10, D'Avila Garcez *et al.* [2005] proposed a hybrid model of computation that allows for deduction and learning with argumentative reasoning. The model manages to combine value-based argumentation frameworks and neural-symbolic learning systems by providing a translation from argumentation networks to neural networks. Another example is the general argumentation framework presented by Williamson and Gabbay [2005]. The framework incorporates the idea of recursive causality and extends the Bayesian network formalism to cope with recursive causality. The authors discussed how support relations behave analogously to causal relations and how arguments are recursive structures; these two observations motivate the use of recursive Bayesian networks for modeling arguments.

In this section, we discussed the formalization of argumentation. As usual, we end with some open questions.

1. We have observed that every conceptualization comes with their own formal methods. What else do they depend on? For example, which formal method(s) should be chosen for a case study or application?

2. Example 2.10 also illustrated that we often need to combine reasoning and formal methods. How can various formal methods be combined in a case study? For example, how can symbolic logic be combined with network (neural and argumentation) reasoning?

3. Formal argumentation is often presented as a general way to deal with nonmonotonicity. But how should arguments be conducted when there are various sources of nonmonotonicity? For example, how should we argue in legal or ethical contexts?

4. Concepts relevant for argumentation currently include, among others: time, action, knowledge, belief, revision, deduction, learning, context, neural networks, probabilistic networks, argumentation networks, consistency, etc. How can these concepts be incorporated into existing formal models of argumentation?

3 The paradigm shift: the attack-defense perspective

In this section, we discuss and critically reflect upon the attack-defense paradigm shift. In section 3.1, we discuss the universality of attack and defense. In section 3.2, we consider the variety of nonmonotonic logics and game theoretic solution concepts. In section 3.3, we discuss reasoning with inconsistent knowledge bases. In section 3.4, we consider argumentation as dialogue that is based on other concepts besides attack, like agents, strategies, and games. And in section 3.5, we discuss Dung's attack-defense paradigm shift for balancing that is based on both attack and support. We discuss the challenges of: creating argumentation frameworks and semantics, representing nonmonotonic logics and game theoretic concepts, defining rationality postulates for new logics, generalizing Dung's attack-defense paradigm shift for dialogue, and generalizing Dung's attack-defense paradigm shift for balancing.

3.1 Universality of attack

In this section, we introduce the attack-defense paradigm shift initiated by Dung's [1995] paper, we discuss the requirement that every utterance can be attacked including claims, arguments, and attacks, and we describe the flattening of diverse extended argumentation frameworks into basic ones.

The attack-defense paradigm shift was a turning point in modern formal argumentation, marked by Dung's theory of abstract argumentation [Dung, 1995]. In this theory, the acceptability of arguments depends on the attack relations between them, not their internal structure. An argument is accepted if it is not attacked or is successfully defended — meaning all its attackers are attacked. Pre-existing ideas and methods, such as Pollock's defeasible reasoning, dialogue theories, and balancing techniques, continue to persist and influence contemporary research. Rather than being rendered obsolete, these traditional theories are reinterpreted within the context of this new paradigm.

While the central notion of Dung's theory is the acceptability or non-acceptability of arguments based on attack and defense, Dung shows that nonmonotonic logic is a special form of argumentation (more details in Section 3.2). It can be visualized as the commutative diagram in Figure 3. There are two approaches to deriving conclusions from a knowledge base. The first is a direct approach where a given logic selects a set of rules with conclusions.

The other is an indirect approach through argumentation, as shown in Figure 3 (2—4). Structured argumentation studies the process that adds the structure that turns collections of rules into arguments and assigns attack relations (2) among arguments. This gives us abstract argumentation frameworks — directed graphs where nodes represent arguments, and arrows represent attack relations. Then argumentation semantics (3) determine the acceptance status of arguments and their conclusions. To represent a given logic by structured argumentation, eventually the conclusions from both approaches must be the same.

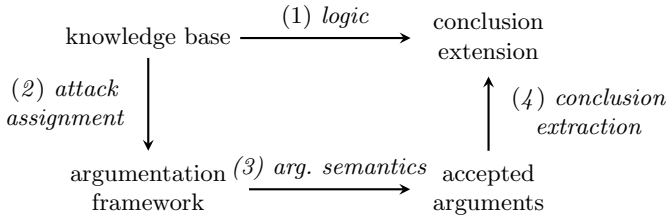


Figure 3: Commutative diagram: two approaches to nonmonotonic inference: (1) logic systems; (2)–(4) argumentation systems. With appropriate choices on elements (2) and (3), one can obtain exactly the same conclusions as for the given logic (1).

We now present informally the construction of arguments and attack relations from a knowledge base in the ASPIC+ structured argumentation frameworks [Prakken, 2010a]. A knowledge base typically consists of a set of strict rules (with a simple arrow $\rightarrow$) and a finite set of defeasible rules (using a double-lined arrow $\Rightarrow$), where each defeasible rule is assigned a priority number, denoted as $a \overset{n}{\Rightarrow} b$. The knowledge base also includes a base of evidence (BE). An argument can be constructed as follows:

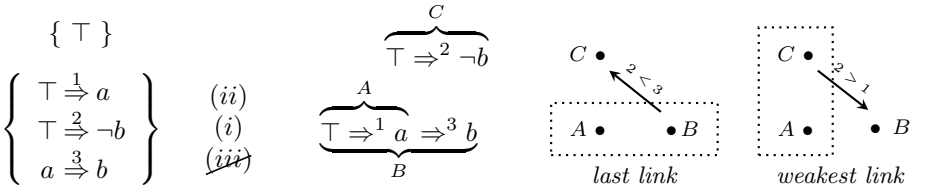
1. For each element $\alpha \in BE$, the expression $[\alpha]$ constitutes an argument having the conclusion α.
2. Let r be a rule of the form $\alpha_1, \ldots, \alpha_n \rightarrow / \Rightarrow \alpha$, where $A_1, \ldots, A_n$ are arguments with conclusions α_i (for $1 \leq i \leq n$). In this case, the expression $[A_1, \ldots, A_n \rightarrow / \Rightarrow \alpha]$ is regarded as an argument with conclusion α.

Each argument is derived by applying the steps above (1 and 2) finitely many times to ensure a structured process for argumentation within the framework.

We now use Example 3.1 to illustrate the commutative diagram, and we explain the technical details later in section 3.2.

Example 3.1 (Two approaches to nonmonotonic reasoning). Consider a knowledge base containing three defeasible rules $\overset{n}{\Rightarrow}$ as well as facts ($\{\top\}$), as in Figure 4(a). Logical approaches to defeasible reasoning select a subset of

rules whose conclusions are maximally consistent. For example, PDL [Brewka and Eiter, 1999], discussed in Section 2.4, selects the strongest applicable rules, i.e., the order $(i) \rightarrow (ii)$ in Figure 4(a), with output $\{a, \neg b\}$. While (iii) is now made applicable by a, its consequent b conflicts with $\neg b$ and cannot be selected. Argumentation approaches, in turn, build explicit *arguments* (Figure 4(b)) and represent these conflicts $(b, \neg b)$ as attacks between arguments (B and C). Observe how the arguments in Figure 4(b) activate the rules in Figure 4(a). To specifically capture PDL, one needs a selection of *attacks* (discussed in Section 3.2), such as the attack induced by the *weakest link* in Figure 4(c), which defines that the strength of an argument reflects its weakest rules. Intuitively, the jointly acceptable arguments here are $\{A, C\}$, which corresponds to the PDL extension $\{a, \neg b\}$. But in Figure 4(c), the *last link*, which defines that the strength of an argument is that of its last rule, selects the arguments $\{A, B\}$ with output $\{a, b\}$.



(a) knowledge base (b) arguments (c) attack assignments; extensions

Figure 4: The PDL approach to nonmonotonic reasoning. (a) PDL iteratively selects the strongest active rule up to the point of inconsistency. (b) Arguments build upon facts (A, C) or other arguments (B). (c) Argumentation semantics abstract from a logical structure. Attacks depend on argument strength.

While nonmonotonic logic provides a structured framework for managing conflicts and reasoning through premises, inferences, and conclusions, it becomes difficult to capture complex conflicts among a large number of arguments. Here, abstract argumentation frameworks can provide landscapes of how these arguments relate to one another via attack relations. One of the main goals of abstract argumentation theory is to answer the question: which sets of arguments can be reasonably accepted in a discussion based on a given argumentation framework? In the simple argumentation frameworks of Figure 4(c), sets of arguments could be selected by intuition. But a formal method is needed for a more complex graph, for instance where attacks among arguments can form even or odd cycles, which may be part of more complex structures like strongly connected components (SCCs). In abstract argumentation, argumentation semantics provides a way to deal with these complications.

We use graph labeling based on so-called *gunfight rules* [Caminada, 2006; Caminada and Gabbay, 2009] to determine which arguments can survive in

conflicts. The concept is straightforward: in a gunfight, one stays alive iff all attackers are dead, and one dies iff at least one attacker is still alive. Understanding this analogy essentially captures the core idea of abstract argumentation:

1. An argument is labeled *in* iff all its attackers are labeled *out*;
2. An argument is labeled *out* iff it has at least one attacker that is labeled *in*;

Example 3.2 (Argumentation framework with two cycles). Consider the argumentation framework in Figure 5, which has a set of arguments: $\{a, b, c, d, e\}$. We follow the direction of the graph. On the left, we have an even cycle: a and b attack each other. On the right, we have an odd cycle: c attacks d, d attacks e, and e attacks c. The two cycles are connected by the attack from b to c, thus the status of the arguments in the even cycle will influence the status of the arguments in the odd cycle. In the left cycle, there are two possibilities. In the first case, a is labeled *in*, then b is labeled *out*. However, there is no way to label the arguments in the odd cycle on the right. Thus, we need a third label called *undec* (undecided), indicating that one abstains from an explicit judgment whether the argument is in or out. It means that not all the attackers are labeled *out* and no attackers are labeled *in*. Therefore, c, d, e will be labeled *undec*. In the second case, b is *in*, a is *out*, and c is *out*. Then d is *reinstated* as an *in* argument because its only attacker c is *out*; we can also say b defends d. It follows that e is *out* because d is now *in*. We can label a and b with the third label *undec*, and all the arguments in the odd cycle are also labeled *undec*.

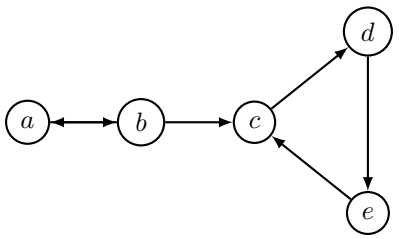
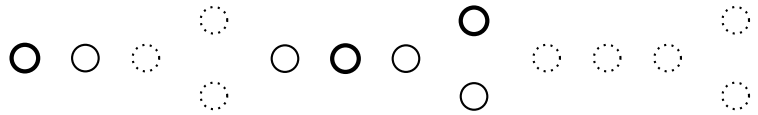


Figure 5: Argumentation framework with two cycles

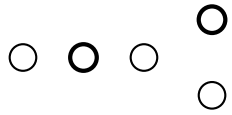
Now the question is: which labelings in Dung's theory are called extensions? We illustrate the extensions to the framework in Example 3.2 below. We use the thick nodes for *in*, normal nodes for *out*, and dotted nodes for *undec*, to obtain a visualization similar to a colored graph. We say that if a labeling is three-valued, then it is a *complete* extension, as in the first item below. A complete extension generalizes a stable extension and there is no argument labeled *undec*, i.e., there is no dotted node, as in the second item below. The unique grounded

extension is the most skeptical complete extension; only arguments that cannot avoid being accepted are labeled *in*, as in the third item. For some frameworks, there are no stable extensions. Then we can use preferred extensions, which are the maximal complete extensions, as in the fourth item.

- Three complete extensions (3-valued)

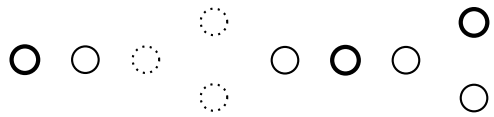


- One stable extension (2-valued)



- One grounded extension (minimal complete)

- Two preferred extensions (maximal complete)



While there have been many transformations of nonmonotonic reasoning formalisms into Dung's theory, direct usage is limited when modeling the argumentation of some realistic examples [Boella *et al.*, 2009] such as multiagent argumentation and dialogues [Yu *et al.*, 2021; Arisaka *et al.*, 2022], decision-making [Kakas and Moraitis, 2003], coalition formation [Amgoud, 2005], combining micro arguments [Toulmin, 1958], normative reasoning [Atkinson and Bench-Capon, 2005], or meta-argumentation [Boella *et al.*, 2009]. That leads to the following challenge.

Challenge 5. Creating argumentation frameworks and semantics.

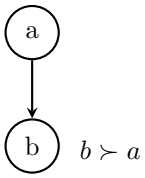
Several extensions to abstract argumentation frameworks are discussed in the second volume of the Handbook of Formal Argumentation. Figure 6 visualizes six examples of extended argumentation frameworks. Preference-based argumentation [Kaci *et al.*, 2021] introduces a preference relation between the arguments, as shown in Figure 6(a), where *a defeats b*, and *b* is preferred over *a*. Bipolar argumentation [Cayrol and Lagasquie-Schiex, 2005] defines support and attack independently. There are arguments in favor of other arguments, i.e., with a support relation, and also arguments against other arguments, i.e., with an attack relation, as shown in Figure 6(b). Weighted argumentation [Bistarelli

et al., 2021] specifies a numeric value that indicates the relative strength of an attack, as shown in Figure 6(c). Abstract agent argumentation [Yu *et al.*, 2021] extends Dung's framework with a set of agents and a relation associating arguments with agents, as shown in Figure 6(d). Value-based argumentation [Atkinson and Bench-Capon, 2021], as shown in Figure 6(e), defines values that are associated with an argument. The preference ordering of the values may depend on a specific audience. To model defeat for a specific audience: an argument A attacks an argument B for audience a if A attacks B and the value associated with B is not preferred to the value associated with A for audience a. Higher (second)-order argumentation [Cayrol *et al.*, 2021] introduces a new kind of attack which is a binary relation from arguments to attack relations, as shown in Figure 6(f).

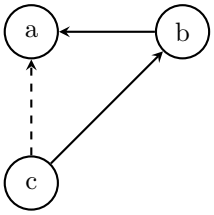
One technique that has already proven to be useful in the past for studying such extensions is a meta-argumentation methodology involving the notion of flattening [Boella *et al.*, 2009]. Flattening is a function that maps some extended argumentation frameworks into Dung frameworks. There are two main flattening techniques. One is that we keep the arguments the same while removing attacks or introducing auxiliary attacks (this is also called reductions sometimes). This technique is used in preference-based argumentation, abstract agent argumentation, bipolar argumentation, etc. The other technique is to use not only auxiliary attacks but also auxiliary arguments in higher-order argumentation.

Example 3.3 (Four reductions of preference-based argumentation). Figure 7 illustrates the differences between the four reductions from a preference-based argumentation framework to abstract argumentation frameworks [Kaci *et al.*, 2021]. The basic idea of Reduction 1 is that an attack succeeds only when the attacked argument is not preferred to the attacker. Reduction 2 enforces that one argument defeats another when the former is preferred but attacked by the latter. The idea of Reduction 3 is that if an argument is attacked by a less preferred argument, then the former should defend itself against its attacker. Reduction 4 mixes the second and the third reductions.

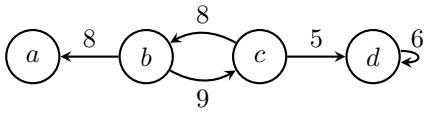
Flattening by adding auxiliary arguments is a way of implementing the methodology of meta-argumentation [Boella *et al.*, 2009]. Meta-argumentation generally involves taking into account the arguments of, e.g., lawyers, commentators, citizens, teachers, or parents (in accordance with the level of their expertise) but it can also go beyond this — the arguers and the meta-arguers can be represented by the *same* reasoners. For example, a lawyer may debate whether a suspect's argument attacks another argument, and she may also argue in a similar way about her own arguments. To give another example, people may be in the middle of an argument, but then start questioning the rules of the dialogue game, and argue about that. A further example is that of a child arguing that the argument *I was ill* attacks the argument *I have to do my homework* but then finds that the argument *I have a nice tan* attacks the



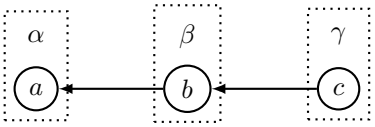
(a) **Preference-based argumentation** $\langle Ar, \rightarrow, \succeq \rangle$, where Ar is a set of arguments, $\rightarrow$ is an attack over arguments, and $\succeq$ is a preference relation over Ar.



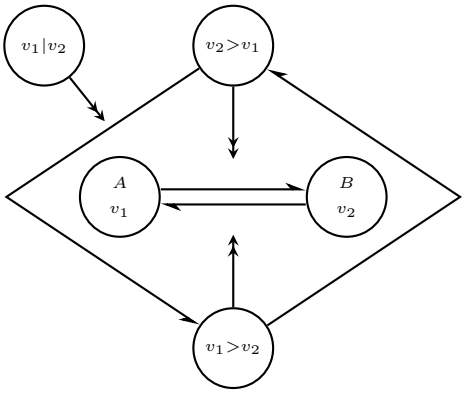
(b) **Bipolar argumentation** $< Ar, \rightarrow, \dashrightarrow >$, where Ar is a set of arguments, $\rightarrow$ is an attack over arguments, and $\dashrightarrow$ is support for arguments.



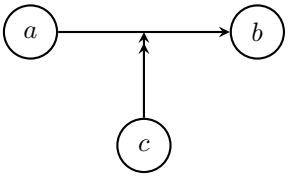
(c) **Weighted argumentation** $\langle Ar, \rightarrow, w \rangle$, where Ar is a set of arguments, $\rightarrow$ is an attack over arguments, and w is a function from Ar to $[0, 1]$.



(d) **Abstract agent argumentation** $\langle Ar, \rightarrow, S, \sqsubset \rangle$, where Ar is a set of arguments, $\rightarrow$ is an attack over arguments, S is a set of agents, and $\sqsubset$ is a binary relation associating arguments with agents.



(e) **Value-based extended argumentation** $\langle Ar, \rightarrow, V, val, P \rangle$, where Ar is a set of arguments, $\rightarrow$ is an attack over arguments, V is a set of values, val is a function mapping arguments to values, and P is a set of audiences.



(f) **Higher-order argumentation** $\langle Ar, \rightarrow, \twoheadrightarrow \rangle$, where Ar is a set of arguments, $\rightarrow$ is an attack over arguments, and $\twoheadrightarrow$ is an attack relation over arguments that can be attacked.

Figure 6: Six extensions to the argumentation framework

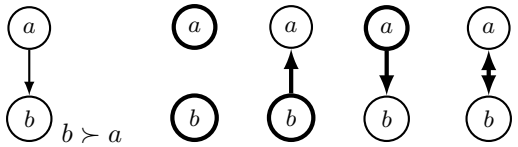


Figure 7: From left to right: the original argumentation framework, and the results after applying the four reductions respectively. The defeat relation is visualized with thick lines, and arguments that are accepted in grounded semantics also have thick lines.

argument *I was ill*.

When we flatten the extended framework, if an argument a of the extended argumentation framework also occurs in the flattened abstract argumentation framework, then we no longer refer to it as argument a but as the meta-argument "argument a is accepted", denoted as $accept(a)$. We use Example 3.4 to illustrate instantiating Dung's abstract argumentation framework by introducing meta-arguments that use flattening.

Example 3.4 (Flattening with auxiliary arguments [Boella *et al.*, 2009]). Given the higher-order argumentation framework in Figure 6(f), the flattened framework is as illustrated in Figure 8. We introduce the meta-arguments $Y_{a,b}$, which means that a is capable of attacking b, and $X_{a,b}$. which means that a does not have the capability of attacking b. We use the meta-arguments in the following way. Each $a \rightarrow b$ is replaced by $accept(a) \longrightarrow X_{a,b} \longrightarrow Y_{a,b} \longrightarrow accept(b)$. The accepted arguments are $\{accept(a), accept(c), Y_{c,Y_{a,b}}, accept(b)\}$.

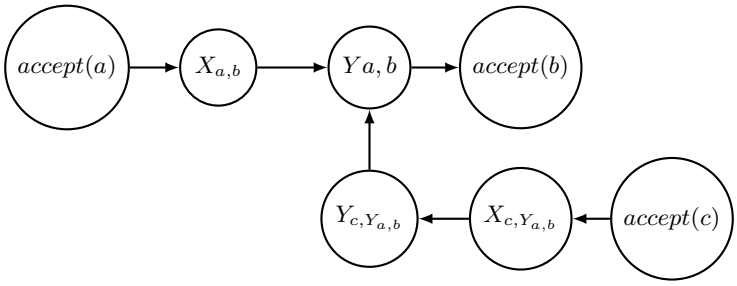


Figure 8: Flattened argumentation framework for Figure 6(f)

These examples illustrate how diverse extended argumentation frameworks build upon Dung's foundational concepts by introducing additional elements. However, the key reasoning ontology fundamentally relies on just two elements: arguments and attacks. Every utterance, be it a claim, an argument, or even an

attack, can be modeled as an argument within these frameworks. Consequently, many of these extended frameworks can be *flattened* to basic ones, reinforcing the idea that attack graphs serve as a universal model of reasoning — much like how Turing machines serve as a universal model of computation.

In this section, we discussed the attack-defense paradigm shift introduced by Dung, emphasizing that every type of utterance (claim, argument, or attack) can be attacked. We also presented how diverse extended argumentation frameworks can be flattened into basic ones, demonstrating the universality of attack. We end this section with the following questions:

1. What is an argument? What is an attack? What is the interplay between an argument and an attack?
2. Should attack always be the first-class citizen in formal argumentation? For example, a novel notion of attack-defense is adopted as a first-class citizen by Liao and van der Torre [2024]. It can represent some knowledge that cannot be represented in Dung-style argumentation, e.g., some context-sensitive knowledge in a dialogue.
3. How should the new diversity created by the attack-defense perspective be handled?
4. What does the attack-defense paradigm shift mean for argumentation as dialogue? What does the attack-defense paradigm shift mean for argumentation as balancing?
5. If we flatten an extended argumentation framework we introduce auxiliary arguments. How can we then recognize these auxiliary arguments in the instantiated Dung abstract argumentation framework? How can we deal with the original arguments and the arguments introduced later?

3.2 Representing nonmonotonic logics

In this section, we discuss: structured argumentation as a bridge from classical to nonmonotonic logic, the variety of nonmonotonic logics available, and the represention of nonmonotonic logic. We again refer to the commutative diagram in Figure 3 that we used in section 3.1 to illustrate how the same conclusions can be reached with two different approaches: the direct approach using logic and the indirect approach through the construction of argumentation frameworks, and semantics. Here, we illustrate this using the weakest versus last link principle and we continue with PDL and the weakest link.

Despite the uniqueness of classical logic, a wide variety of nonmonotonic logics are employed in different contexts. However, engaging in nonmonotonic logics means the aim is to extend classical logic rather than replace it *tout court* [Makinson, 2005]. Structured argumentation is used to classify existing nonmonotonic logics as a way to define a new nonmonotonic logic and create a bridge from classical to nonmonotonic logic. Dung [1995] provides semantics of attacks for structured argumentation. This has been used in the

ASPIC+ system by Modgil and Prakken [2013; 2014], and it has also been used to reconstruct and compare a variety of nonmonotonic logics, namely default logic [Reiter, 1980], Pollock's [1987] argumentation system, and several logic-programming semantics. More representations have been developed — for details, please refer to the work of Heyninck [2019]. However, as discussed in Section 2, there is also a diversity of natural argumentation, conceptualizations, and formal methods. Notwithstanding the initial appeal of Dung's abstract argumentation theory, there are many different kinds of argumentation frameworks and semantics. That leads to the following challenge:

Challenge 6. Representing nonmonotonic logics and solution concepts.

Before we get into approximating PDL with argumentation, let us first talk about methodologies employed to compare different nonmonotonic logics and, in particular, their use of examples. There are different approaches to the use of examples in different disciplines. In law, ethics, and linguistics, examples are central to the development and validation of theories because they help ground abstract concepts in real-world scenarios, which helps to align logical frameworks with intuitive understanding. In contrast, knowledge representation (KR) and other areas of computer science often use examples as a practical tool to test, demonstrate, and communicate the effectiveness of a formal theory rather than using them as foundational elements in theory construction.

NLP task: translating natural language into formal language. Consider the aforementioned example of the fitness-lover Scot and an additional example about a snoring professor:

The fitness-lover Scot: It is commonly assumed that if a man was born in Scotland, then he is Scottish. And if he is Scottish, we can normally deduce that he likes whiskey. However, fitness lovers normally avoid alcohol for health reasons. Stewart was born in Scotland, and he is also a fitness lover. Does he like whiskey?

The snoring professor A library has a general rule that misbehavior, such as snoring loudly, leads to denial of access. However, there is another rule that professors are normally allowed access. Bob is a professor and he is snoring loudly in the library. Should he be allowed access to the library?

NLP could be used to identify three rules for each example, and then further abstract them into these formal (default) rules with priorities: $\{\top \overset{1}{\Rightarrow} a, a \overset{3}{\Rightarrow} b, \top \overset{2}{\Rightarrow} \neg b\}$. We then have:

Fitness-lover Scot

$$\left\{\begin{array}{c} \mathit{BornInScotland} \overset{1}{\Rightarrow} \mathit{Scottish} \\ \mathit{Scottish} \overset{3}{\Rightarrow} \mathit{LikesWhisky} \\ \mathit{FitnessLover} \overset{2}{\Rightarrow} \neg \mathit{LikesWhisky} \end{array}\right\}$$

Snoring professor in the library

$$\left\{\begin{array}{c} \mathit{snores} \overset{1}{\Rightarrow} \mathit{misbehaves} \\ \mathit{misbehaves} \overset{3}{\Rightarrow} \mathit{accessDenied} \\ \mathit{professor} \overset{2}{\Rightarrow} \neg \mathit{accessDenied} \end{array}\right\}$$

KR task: after inputting some requirements, i.e., the goal of the reasoning, the system asks what you want to derive from what you have. Although the above two examples share a similar structure, there could be different reasoning requirements that lead to the selection of different rules and ultimately different conclusions. In the fitness-lover example, one might prioritize the rule $\top \overset{2}{\Rightarrow} \neg b$ and conclude that Stewart does not like whiskey. In contrast, in the snoring professor example, one might prioritize the rule $a \overset{3}{\Rightarrow} b$ and conclude that Bob should be denied access.

Logic design task: According to these requirements, the system asks what is the best logic for your application. These two examples have been used to illustrate the difference between *prescriptive* and *descriptive* reasoning in nonmonotonic reasoning and between the *weakest link* and the *last link*, which are two principles regarding how an argument draws strength from its defaults.

In Section 2.4 and Section 3.1, we briefly mentioned Brewka and Eiter's [1999] PDL. Pardo *et al.* [2024] compared structured argumentation based on the weakest link variant with that of the PDL variant. Let us start with a reminder that PDL can be understood as a greedy approach, i.e., PDL iteratively adds the strongest applicable and consistent default. Initially, people thought that using the weakest link principle to construct argumentation frameworks would capture this kind of greedy procedure. However, over time, analysis of the weakest-link-related attack assignment reveals that it is more complicated and ambiguous than it appears at first sight.

The history of the weakest link revolves around three key examples from the literature, visualized in Figure 9 and described in Examples 3.5–3.7. Note that these examples illustrate the role of formal argumentation in the context of PDL. We refer to the work of Pardo *et al.* [2024] for the formal definitions. Here, we discuss Examples 3.5–3.7 informally.

The following example illustrates the use of priorities. What does a *stronger priority* mean? Under the *prescriptive* reading, it means priority in the order of application: PDL always selects the strongest default (among those that are applicable and consistent). Under the *descriptive* reading, the priority of a default is its contribution to the overall status of any extension containing this default [Delgrande *et al.*, 2004]. The two readings clash in the most discussed example of defeasible reasoning with prioritized rules.

Example 3.5 (Weakest vs. Last link). Let $\{\top \overset{1}{\Rightarrow} a,\ a \overset{3}{\Rightarrow} b,\ \top \overset{2}{\Rightarrow} \neg b\}$ be again our defaults (Figure 9, top). The two readings of priorities give the following outputs:

(*Prescriptive*.) Based on application order, one must select $\{\top \overset{2}{\Rightarrow} \neg b, \top \overset{1}{\Rightarrow} a\}$ thereby obtaining the output $\{a, \neg b\}$, as in PDL. In fact, PDL is an implementation of the prescriptive reading. Let us call *simple weakest link* (*swl*) the strength defined by the lowest priority of an argument:

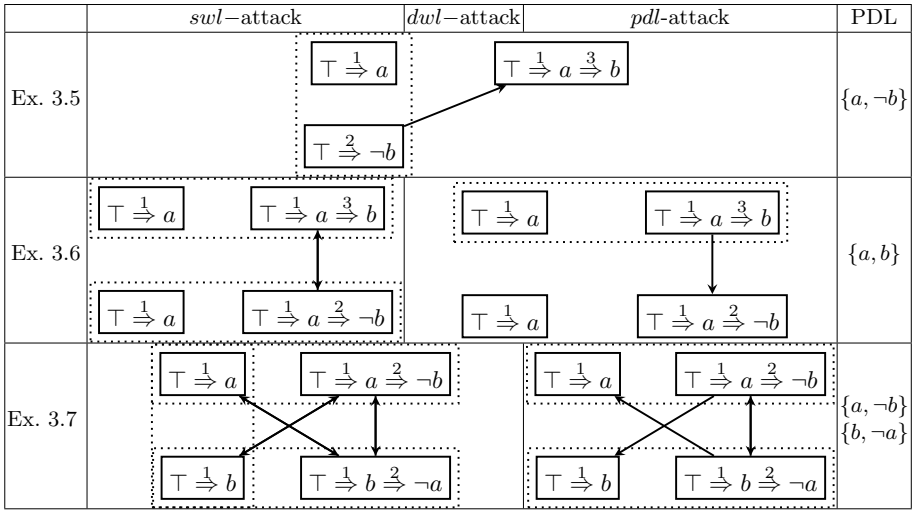


Figure 9: Approximating PDL in structured argumentation: a comparison of three attacks (columns) for three examples (rows). Columns are not marked when adjacent notions of attack agree on the induced attack relation at a given row. Dotted rectangles are argument extensions. The rightmost attacks approximate PDL better.

$$\top \overset{1}{\Rightarrow} a \overset{3}{\Rightarrow} b \quad \longmapsto \quad 1 = \min\{1,3\} \qquad \top \overset{2}{\Rightarrow} \neg b \quad \longmapsto \quad 2 = \min\{2\}$$

A comparison of the strengths in this conflict produces the attack shown in Figure 9 (top). The semantics then gives the argument selection also shown. Our three attack relations (swl, dwl, pdl) do in fact agree on the verdict for this example.[1]

(*Descriptive.*) This reading favours the set $\{\top \overset{1}{\Rightarrow} a, a \overset{3}{\Rightarrow} b\}$ as its priorities $\{1,3\}$ are more desirable than the rival ones $\{1,2\}$. *Last link* can be seen as an implementation of this reading: the contribution of a new default to a selection or argument, say $\{\top \overset{1}{\Rightarrow} a\}$, is defined by the desirability of this default (2 vs. 3 in the example). Last link thus agrees on the above preference but arrives at it through argumentative means. First, one computes argument strength:

$$\top \overset{1}{\Rightarrow} a \overset{3}{\Rightarrow} b \quad \longmapsto \quad 3 = \text{last}(1,3) \qquad \top \overset{2}{\Rightarrow} \neg b \quad \longmapsto \quad 2 = \text{last}(2)$$

Based on this, argument $\top \overset{1}{\Rightarrow} a \overset{3}{\Rightarrow} b$ attacks $\top \overset{2}{\Rightarrow} \neg b$. Using a standard argumentation semantics, one obtains the output $\{a, b\}$, not shown in Figure 9 (top).

[1]This example represents the Tweety scenario $\{penguin \to bird, \quad bird \Rightarrow flies, penguin \Rightarrow \neg flies\}$ with priorities instead of the strict rule $(\to)$. Without priorities, the solution $\{penguin, bird, \neg flies\}$ obtains from specificity (of *penguin* over *bird*): *birds fly* is overruled by the more specific *penguins do not fly*. Without specificity the solution obtains from appropriate priorities using PDL or *swl*.

The simple weakest link, though, does not always capture the prescriptive reading. In response to this, a more intuitive *disjoint* variant of the weakest link has been considered [Young *et al.*, 2016]. This variant assumes a relational measure of argument strength. It ignores all the shared defaults before searching for the weakest link between two arguments.

Example 3.6 (Simple vs. disjoint weakest link). Let $\{\top \overset{1}{\Rightarrow} a,\ a \overset{3}{\Rightarrow} b,\ a \overset{2}{\Rightarrow} \neg b\}$ define our knowledge base. Note that the two arguments $\top \overset{1}{\Rightarrow} a \overset{3}{\Rightarrow} b$ and $\top \overset{1}{\Rightarrow} a \overset{2}{\Rightarrow} \neg b$ share a default $\top \overset{1}{\Rightarrow} a$ with the lowest priority. See the middle row in Figure 9.

(Simple weakest link) Pollock's definition assigns the same strength of 1 to these two arguments. This gives the mutual *swl*-attack in Figure 9 (mid, left). Now, one argument selection $\top \overset{1}{\Rightarrow} a \overset{3}{\Rightarrow} b$ matches the PDL extension $\{a, b\}$; the other $\top \overset{1}{\Rightarrow} a \overset{2}{\Rightarrow} \neg b$, though, gives us a non-PDL extension, $\{a, \neg b\}$.

(Disjoint weakest link) The attack relation defined by *disjoint weakest link* (*dwl*) assigns strengths $3 > 2$ to the above arguments, after excluding the default they share. This generates the tie-breaking *dwl*-attack shown in Figure 9 (mid, right). This figure also shows the set of arguments selected by our semantics. The selected arguments' conclusions match the PDL output $\{a, b\}$.

Pollock's definition of weakest link *swl* [Pollock, 2001] was adopted and studied for ASPIC+ by Modgil and Prakken [2013; 2014]. Then, Young *et al.* [2016; 2017] introduced *dwl* and proved that argument extensions under the *dwl*-attack relation correspond to PDL extensions under total orders; see also the results presented by Liao et al. [2019] and Pardo and Straßer [2022]. Under total preorders, a new attack relation is needed for more intuitive outputs and a better approximation of PDL — that is, better than *dwl*.

Example 3.7 (Beyond *dwl*). Let $\{\top \overset{1}{\Rightarrow} a,\ \top \overset{1}{\Rightarrow} b,\ a \overset{2}{\Rightarrow} \neg b,\ b \overset{2}{\Rightarrow} \neg a\}$ be the defaults.

(*swl*, *dwl*) Weakest link attacks, depicted in Figure 9 (bottom, left), admit the selection of arguments $\{\top \overset{1}{\Rightarrow} a, \top \overset{1}{\Rightarrow} b\}$. This selection fits neither the prescriptive interpretation nor PDL. Selecting either default ought to be followed by the selection of a stronger default, namely $a \overset{2}{\Rightarrow} \neg b$ and $b \overset{2}{\Rightarrow} \neg a$ respectively.

(PDL) As PDL selects the strongest default one at a time, this excludes by construction the concurrent selection of $\{\top \overset{1}{\Rightarrow} a, \top \overset{1}{\Rightarrow} b\}$. The PDL-inspired attack relation in Figure 9 (bottom, right) also excludes this selection.

An important research question is then how to characterize, or at least approximate, the PDL extensions of a prioritized default theory. For total orders, an attack that characterizes PDL extensions already exists: att_{dwl} [Young *et al.*, 2016].

But for total preorders, how to characterize PDL extensions using an attack relation assignment is an open problem. Certainly, such a characterization can no longer be based on the disjoint weakest link, as shown in Example 3.8.

Example 3.8 (Disjoint weakest link vs. PDL). Example 3.7 shows a stable belief set $\{a, b\}$ under $att_{dwl}(K)$ that is not a PDL extension of K.

Example 3.9 (PDL vs. Disjoint weakest link). Let $\{\top \overset{1}{\Rightarrow} a, a \overset{2}{\Rightarrow} b, \top \overset{1}{\Rightarrow} c,\ a, c \overset{2}{\Rightarrow} \neg b\}$ define our knowledge base. Figure 10 shows that the shared rule $\top \overset{1}{\Rightarrow} a$ produces only one stable extension $\mathcal{E}$ under the disjoint weakest link, and so we have a unique stable belief set of $(AR_K, att_{dwl}(K))$:

$$\mathcal{E} = \{A, C, [A \Rightarrow b]\} \quad \longmapsto \quad S = \{a, b, c\}$$

In contrast, two PDL constructions exist for K, and so do two PDL extensions:

$$\begin{array}{lll} (\top \overset{1}{\Rightarrow} a, \quad a \overset{2}{\Rightarrow} b, \quad \top \overset{1}{\Rightarrow} c) & \longmapsto & \{a, b, c\} \\ (\top \overset{1}{\Rightarrow} c, \quad \top \overset{1}{\Rightarrow} a, \quad a, c \overset{2}{\Rightarrow} \neg b) & \longmapsto & \{a, \neg b, c\} \end{array}$$

As a consequence, disjoint weakest link cannot characterize PDL under stable semantics. Observe that att_{swl} here coincides with PDL.

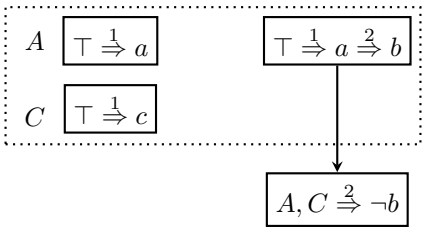


Figure 10: The stable belief set $\{a, c, b\}$ under att_{dwl} for Example 3.9. Two extensions, $\{a, c, b\}$ and $\{a, c, \neg b\}$, exist under PDL.

Attack relations have become a major subject of study in logic-based argumentation. Dung [2014; 2016] recently proposed an axiomatic method that supersedes all argumentation systems with defeasible rules. Pardo *et al.* [2024] attempted to identify an attack relation that captures PDL extensions, and they compared it with attacks based on the simple and disjoint weakest link using the eight principles advanced by Dung. They proved an impossibility theorem: representing PDL in formal argumentation should preserve a principle (attack closure), but this is incompatible with another principle (context independence).

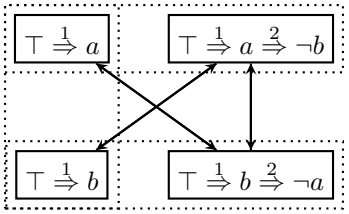


Figure 11: pPDL differs from PDL. PDL has two extensions $\{a, \neg b\}$ and $\{b, \neg a\}$. pPDL has an additional extension $\{a, b\}$. Arrows denote logical conflicts.

As seen in Examples 3.7 and 3.9, disjoint weakest link and PDL are incomparable under total preorders. As a first step towards their convergence, one can slightly modify PDL to make it closer to the disjoint weakest link. To this end, Pardo *et al.* [2024] propose *parallel* PDL (pPDL), a concurrent variant of PDL. The main novelty of pPDL is that each inductive step can concurrently select a set of defaults, rather than just one, for the technical details, we refer to the paper of Pardo *et al.* [2024].

Example 3.10 (pPDL, DWL vs. PDL). Let us use Example 3.7 to show that the default logic PDL differs from pPDL. Figure 11 illustrates the three pPDL extensions $\{a, \neg b\}$, $\{b, \neg a\}$, $\{a, b\}$, of which $\{a, b\}$ is not a PDL extension.

Although pPDL and att_{dwl} agree in this and other examples, pPDL does not always match the disjoint weakest link.

Example 3.11 (pPDL vs. DWL). Example 3.9 showed a unique stable belief set, $\{a, b, c\}$, under att_{dwl}. But there are two pPDL extensions: $\{a, b, c\}$ and $\{a, \neg b, c\}$.

To sum up, the first goal of Pardo *et al.* [2024] was to identify an attack relation that captures PDL extensions and compare it with attacks based on the simple and disjoint weakest link using the eight principles advanced by Dung [2016; 2018]. They proved which principles for attack relations are satisfied by weakest link, disjoint weakest link and PDL based attacks. Their principle-based analysis presented the difference between several kinds of attack relation assignment. They identified and explained the nature of the weakest link principle and revealed that there is still the potential to improve the weakest link attack. On this last question, they proposed pPDL (parallel PDL), a concurrent variant of PDL, and they showed by way of examples that it falls closer to the disjoint weakest link than PDL does. While the pPDL variant still does not match the disjoint weakest link, one might conjecture that some further refinement might do.

In addition to presenting the argumentation framework, Dung [1995] also investigated two examples of problems from microeconomics — cooperative game theory and matching theory. In each case, Dung showed how an appro-

priate framework can represent a given cooperative game or a given instance of the stable marriage problem, and that the sets of winning arguments in such argumentation frameworks correspond to meaningful solutions in both these domains.

Cooperative game theory studies how rational agents cooperate to form coalitions to maximize their payoffs. A coalition's payoff is measured by its value, and agents ideally cooperate to form a coalition. The Von Neumann-Morgenstern (vNM) stable set [Von Neumann and Morgenstern, 1947] is the solution concept for distributing the grand coalition's payoff and ensuring that no agent defects. Dung showed that stable extensions of an argumentation framework correspond to vNM stable sets [Dung, 1995, Theorem 37]. However, just like the stable extensions of an argumentation framework may not exist, vNM stable sets also do not always exist. Dung proposed that sets of payoff distributions that form preferred extensions could serve as an alternative solution concept because preferred extensions always exist, and therefore this is well defined for all cooperative games.

In this section, we discussed the representation of nonmonotonic logics using the attack-defense paradigm. We end with several questions concerning such representations:

1. We showed that PDL and the weakest link definitions are similar but not exactly the same. How can PDL be changed to make it fit one of the weakest link definitions? How can the weakest link be changed to fit PDL?

2. We discussed the logic of the weakest link. What is the logic that corresponds to the last link?

3. We showed various alternative formalizations of the weakest link principle. Likewise, are there variants of the last link principle?

4. PDL is only one of many logics for prioritized rules. How can all the other systems for prioritized rules be represented?

5. We discussed representation of nonmonotonic logics, but Dung also talked about logic programming and game theory. What is the relation between different solution concepts in game theory and (extended) abstract argumentation frameworks semantics?

3.3 Postulates from paraconsistent reasoning

In this section, we continue our discussion of formal argumentation as a logical framework for nonmonotonic reasoning. We consider inconsistent knowledge bases and so-called rationality postulates from paraconsistent reasoning, which is used to define new nonmonotonic logics in ASPIC+. We illustrate the postulates using the example of three persons on a two-person tandem taken from Caminada and Wu [2011].

In previous sections, we discussed universality of attack and one resultant challenge — representing existing nonmonotonic logic. In particular, we discussed representing PDL with structured argumentation, comparing attack assignments using variants of the weakest link with principles. What we showed is just the tip of the iceberg. There are numerous options based on different knowledge bases containing various types of information such as strict and/or defeasible rules. There are different methods for constructing argumentation frameworks. There are applications of distinct semantics. The combination of all these factors defines different argumentation-based logics that can be adopted or rejected, depending on their applicability in different contexts. *Rationality postulates* are a list of desiderata that structured argumentation systems should satisfy in order to be logically well-behaved [Caminada and Amgoud, 2007]. In this section, we address the following challenge:

Challenge 7. Rationality postulates for defining a new logic.

Various rationality postulates are inspired by paraconsistent reasoning. Paraconsistent logic [Da Costa, 1974; Priest, 2002] is a non-classical logical system designed to handle contradictions without leading to the collapse (or "explosion") of the entire systems (as would occur in classical logic). These logics have inspired the development of modal and nonmonotonic logics as well as various rationality postulates [Da Costa *et al.*, 2007]. Such postulates ensure that logic can handle inconsistencies without leading to the kind of trivialization where any and every conclusion becomes derivable from a set of contradictory premises. One key postulate of paraconsistent logic is noninterference, i.e., independent knowledge bases do not influence each other's outcomes. Another is avoiding contamination, i.e., the outcome of a set of formulas remains unchanged when merged with an unrelated set [Caminada *et al.*, 2012].

A side note regarding terminology: we use terms such as postulates, axioms, requirements and desiderata in a rather interchangeable manner, and they differ slightly from principles and properties. All six terms refer to the behavior of logic, the construction of an argumentation framework, and the semantics of argumentation frameworks. Abstract properties are formally specified, and in this section, postulates are treated as desiderata, akin to formal requirements in computer science. In Section 4.1, where we discuss *the principle-based methodology* in detail, postulates are regarded as more general properties, with some being desirable and others not.

There are three fundamental rationality postulates [Caminada and Amgoud, 2007]. Direct Consistency means that any extension should be consistent according to certain semantics. Indirect Consistency means that the set of the conclusions of arguments in a given extension is consistent when closed under the strict rule. Closure means that arguments with conclusions derived from arguments in an extension using strict rules should also be in the extension.

Given these postulates, the question is under what conditions do structured argumentation satisfy them. When assigning attack relations among arguments from a knowledge base, there are so-called rebuts when the conclusions

of two arguments conflict with one other. Two kinds of rebuts have been discussed in the literature: *restricted rebuts* and *unrestricted rebuts*. The intuition behind restricted rebut is: if an argument is built up with only strict rules, then the conclusion should also be strict, and the argument cannot be attacked. The intuition behind unrestricted rebut is that a conclusion is defeasible, i.e., it can be attacked iff it is built up with at least one defeasible rule. Different choices on rebuts influence how to define the argumentation formalism that derives reasonable conclusions. This exists in the ASPIC family of argumentation frameworks, including ASPIC+ [Modgil and Prakken, 2013; Modgil and Prakken, 2014], ASPIC- [Caminada *et al.*, 2014] and ASPIC-END [Dauphin and Cramer, 2018].

Example 3.12 illustrates rationality postulates, comparing unrestricted rebut and restricted rebut, and it shows the solutions to restricted rebut required to satisfy the rational postulates for this example.

Example 3.12 (Married John [Caminada and Amgoud, 2007]). Consider an argumentation system consisting of the strict rules $\{\rightarrow r, \rightarrow n, m \rightarrow hs, b \rightarrow \neg hs\}$ and the two defeasible rules $\{r \Rightarrow m, n \Rightarrow b\}$. An intuitive interpretation of this example is the following: "John wears a ring (r) on his finger. John is also a regular nightclubber (n). Someone who wears a ring on his finger is usually married (m). Someone who is a regular nightclubber is usually a bachelor (b). Someone who is married has a spouse (hs) by definition. Someone who is bachelor does not have a spouse ($\neg hs$) by definition." We can construct the following arguments:

$$\begin{array}{lll} A_1 :\rightarrow r & A_3 : A_1 \Rightarrow m & A_5 : A_3 \rightarrow hs \\ A_2 :\rightarrow n & A_4 : A2 \Rightarrow b & : A_6 : A_4 \rightarrow \neg hs \end{array}$$

If we apply unrestricted rebut, we have A_5 and A_6 attacking each other, and we obtain the grounded extension of $\{A_1, A_2, A_3, A_4\}$ with the conclusion extension $\{r, n, m, b\}$, which does not satisfy the direct consistency property. If we apply restricted rebut, the situation is even worse. Because we do not have any attack relations, we have the extension $\{A_1, A_2, A_3, A_4, A_5, A_6\}$ and the conclusion extension $\{r, n, m, b, hs, \neg hs\}$, which are not consistent.

Two solutions for argumentation systems applying restricted rebut to satisfy the rationality postulates are *closure of transposition* and *closure of contraposition*, as adopted in ASPIC+ [Modgil and Prakken, 2014].

Example 3.13 (Example 3.12 continued). Given that we have $m \rightarrow hs$ and $b \rightarrow \neg hs$ in the knowledge base, we add their "contraposed" versions: $\neg hs \rightarrow \neg m$ and $hs \rightarrow \neg b$. We can construct additional arguments: $A_7 : A_5 \rightarrow \neg b$ and $A_8 : A_6 \rightarrow \neg m$. We have that A_7 restrictively rebuts A_4, and that A_8 restrictively rebuts A_3. As a result, each set of conclusions yielded under grounded or preferred semantics satisfies the postulates of direct consistency, closure, and indirect consistency.

We use the example of three persons on a two-person tandem [Caminada and Amgoud, 2007] to take a closer look at unrestricted rebut and restricted rebut — the latter is applied where unrestricted rebut can lead to undesired behavior.

Example 3.14 (Restricted rebut vs. unrestricted rebut). Consider a knowledge base consisting of three defeasible rules, $\{\top \Rightarrow p, \top \Rightarrow q, \top \Rightarrow r\}$, and three strict rules, $\{p, q \rightarrow \neg r, p, r \rightarrow \neg q, q, r \rightarrow \neg p\}$. We can construct six arguments as shown below. If we apply unrestricted rebut, we can obtain the abstract argumentation framework on the left hand side of Figure 12. One of the complete extensions is $\{A_1, A_2, A_3\}$, yielding conclusion extension $\{p, q, r\}$. If we close this extension under strict rules, we have $\{p, q, r, \neg p, \neg q, \neg r\}$, which is not consistent. If we apply restricted rebut, we obtain the framework at the right hand side of Figure 12, where we have the complete extensions of $\{A_1, A_2, A_6\}$, $\{A_1, A_3, A_5\}$ and $\{A_2, A_3, A_4\}$. They are also consistent under the closure of strict rules.

$$A_1 : \top \Rightarrow p \qquad A_4 : q, r \rightarrow \neg p$$
$$A_2 : \top \Rightarrow q \qquad A_5 : p, r \rightarrow \neg q$$
$$A_3 : \top \Rightarrow r \qquad A_6 : p, q \rightarrow \neg r$$

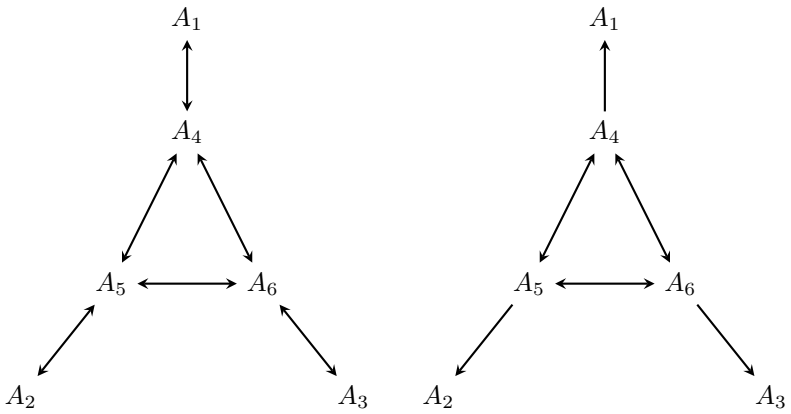


Figure 12: Resticted rebut vs. unrestricted rebut

There are more postulates. For example, *noninterference* and *crash resistance* are particularly relevant when the strict rules are derived from classical logic, and again we examine various ways of satisfying these properties. However, there have been comparatively fewer results that would establish them in systems of the ASPIC family.

Noninterference: no set of formulas can influence the entailment of an unrelated set of formulas when they are merged with a completely unrelated (syntactically disjoint) defeasible theory.

Crash resistance: no set of formulas can make an unrelated set of formulas completely irrelevant when they are merged with a completely unrelated (syntactically disjoint) defeasible theory.

A violation of non-interference means that a defeasible theory somehow influences the entailment of a completely unrelated (syntactically disjoint) defeasible theory when being merged with it. A violation of the crash resistance property is more severe, as this means that a defeasible theory influences the entailment of a completely unrelated (syntactically disjoint) defeasible theory to such an extent that the actual content of this other defeasible theory becomes irrelevant.

In this section, we discussed the use of postulates from paraconsistent reasoning in argumentation. We end with some open questions.

1. In structured argumentation, arguments can be attacked by either defeasible premises, defeasible inference rules, or the conclusion of defeasible rules. In assumption-based argumentation, there are only defeasible premises, while ASPIC+ allows all defeasibilities. How should we decide upon and clarify the various defeasibilities in structured argumentation?

2. Incorporating formal argumentation and social concepts has attracted much interest. One example is the use of an argumentative approach to normative reasoning [Dong *et al.*, 2019; Pigozzi and van der Torre, 2018; Dong *et al.*, 2021; Straßer and Arieli, 2019]. The question then is how to construct and evaluate deontic arguments.

3. Dialectical concepts like multiple agents, communication steps, or commitment stores (like those of the Fatio dialogue system) do not play a role in ASPIC+, which is more monolithic. If we want to add dialectical aspects to structured argumentation [Prakken, 2024a], how should we design an argumentation system that behaves logically?

3.4 Extensions of the attack-defense paradigm for dialogue

In this and the following sections, we discuss extensions to abstract argumentation. There are various approaches to extracting more information from frameworks, and there is a variety of qualitative and quantitative enrichments of frameworks. Semantics can be defined by reductions, selections, or adaptations of defense. In argumentation as inference, only preference is clearly linked to structured argumentation. In this section, we focus on extensions inspired by dialogue.

There are two kinds of extensions to abstract argumentation in the commutative diagram of Figure 3. The first extends the argumentation framework with

qualitative and quantitative components from the knowledge base. In section 3.1, we mentioned various examples of such extensions. The second pertains to step (3), where the argumentation semantics contains more information rather than the acceptance of arguments.

For the second type of extension, Villata *et al.* [2011], for instance, generalize the argument semantics by selecting from the graph not only a set of nodes but also a set of edges. This represents intuitively that attacks can be successful or unsuccessful. A similar kind of intuition is formalized in extended argumentation frameworks with second or higher-order attacks [Barringer *et al.*, 2005; Cayrol *et al.*, 2021]. Attacks are treated as arguments that can be attacked, and thus can be accepted and rejected too. Consider, for example, the two-three cycle framework shown in Figure 5. One possible output is a subframework where we retain only the attack from b to a in the cycle on the left, effectively reducing the complexity of the argumentation structure while maintaining specific attack relations.

Extended abstract argumentation frameworks enhance the expressive capacity of frameworks. However, it is not clear how these extensions can be constructed directly from a knowledge base while incorporating additional components such as agents, supports, numerical values or weights. One exception is preference, which is clearly linked to structured argumentation as inference. In structured argumentation, preferences play a central role in determining formal outcomes. For example, in ASPIC+ [Modgil and Prakken, 2013], the defeat relation between arguments is governed by a preference order, typically derived through mechanisms such as the weakest link or last link principles. Specifically, an attack from one argument to another only succeeds as a defeat if the attacked argument is not stronger than or strictly preferred to the attacking argument, according to the given preference relation. In abstract preference-based argumentation [Kaci *et al.*, 2021], the first reduction in Figure 7 corresponds to this type of attack assignment.

In section 2.3, we introduced argumentation as dialogue. In section 2.4, we discussed its formal methods, e.g., speech acts, game theory, axiomatic semantics, and operational semantics. Inspired by dialogue, we have the following challenge:

Challenge 8. Generalizing Dung's attack-defense paradigm for dialogue.

At the structured level, it is natural to have the role of agents. One example is Jiminy architecture [Liao *et al.*, 2023], discussed in section 2.1. Jiminy involves multiple stakeholders, each with their own knowledge base. When dilemmas and conflicts arise, the argumentation engine considers the combination of all the arguments constructed by each stakeholder. Either there is a large framework consisting of all the stakeholders' arguments and the attack relations, or all the knowledge bases are combined first, and the argumentation frameworks are constructed afterward.

At the abstract level, agent-based extensions typically introduce various aspects such as agents, coalitions, knowledge, uncertainty, support, and so on.

As a result, there are various ways to define the semantics. Below we discuss abstract agent argumentation [Yu *et al.*, 2021], which uses a minimal extension of Dung's framework as a common core. This work only introduces an abstract set of agents and arguments are associated with agents. There are four types of semantics, defined by adaptations of defense, reductions, aggregations, and selections:

Agent defense approaches adapt Dung's notion of defense to argumentation semantics.

Social approaches are based on counting the number of agents [Leite and Martins, 2011] and a reduction to preference-based argumentation [Amgoud and Cayrol, 2002].

Agent reductions take the perspective of individual agents and aggregate their individual perspectives [Giacomin, 2017].

Filtering methods are inspired by agents' knowledge or trust [Arisaka *et al.*, 2022]. They leave out some arguments or attacks because they do not belong to any agent.

Yu *et al.* [2021] have defined *individual agent defense* and *collective agent defense*. Roughly, in individual agent defense, if an agent puts forward an argument, it can only be defended by arguments from that same agent, i.e., a set of arguments E from individual agent α defends an argument c iff there exists an agent α who has argument c such that for all arguments b attacking c, there exists an argument a in E from α attacking b. Whereas with collective agent defense, a set of agents α can do that, i.e., a set of arguments E defends c collectively iff for all arguments b attacking c, there exists an agent α who has c and an argument a in E from a set of agents α such that a attacks b. Example 3.15 illustrates these two agent defenses.

Example 3.15 (Individual agent defense vs. collective agent defense). In Figure 6(d), argument c defends argument a, but it does not individually agent-defend it because c and a come from different agents. Consider another abstract agent framework visualized in Figure 13. Here, $\{c_1, c_2\}$ collectively agent-defend argument a, but they do not individually agent-defend it.

Social semantics is based on a reduction to preference-based argumentation for each argument, by counting the number of agents that have those arguments. It thus interprets agent argumentation as a kind of voting procedure. Example 3.16 illustrates social reduction.

Example 3.16 (Social reduction). Consider the agent argumentation framework visualized in Figure 14. Arguments a and b both belong to agent α, b also belongs to agent β, and a attacks b. In that situation, argument b is preferred to argument a because it is held by more agents. We can then apply the four reductions from preference-based argumentation framework to abstract argumentation framework, followed by application of Dung's semantics.

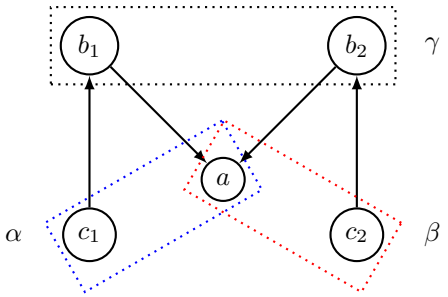


Figure 13: $\{c_1, c_2\}$ collectively agent-defending a

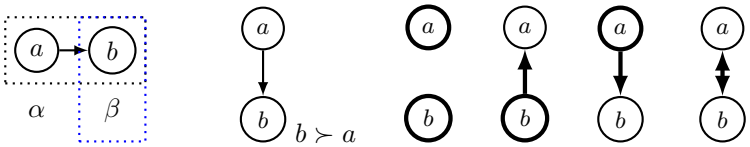


Figure 14: Social reduction: the left hand is an abstract agent argumentation framework, the middle is the corresponding preference-based framework, and the right hand are four corresponding abstract argumentation frameworks (as discussed in Example 3.3)

Agent reductions take the perspective of each agent and obtain the semantics accordingly. Intuitively, agents prefer their own arguments over the arguments of others. Thus, for each agent, there is a preference-based argumentation framework. In social reduction semantics, there is a unique abstract argumentation framework after each reduction. However, in agent reduction semantics, we obtain a set of abstract argumentation frameworks — one for each agent. The final step is to take the union of all these frameworks to form a combined abstract argumentation framework. Example 3.17 illustrates agent reductions.

Example 3.17 (Agent reduction). Reconsider the abstract argumentation framework in Figure 14. Agent β prefers argument b over argument a. Thus, we get the same preference-based framework as depicted in Figure 14, but for a very different reason to that of social reduction. For agent α, argument a and b are equivalent. To compute the agent extensions, we take the union of the reductions for each agent.

Agent filtering semantics remove arguments that do not belong to an agent, or they remove attacks that do not belong to an agent. An attack belongs to an agent if both the attacking and attacked arguments belong to that agent. Example 3.18 illustrates agent filtering semantics.

Example 3.18 (Agent filtering). Consider the two abstract agent frameworks visualized in Figure 15. For the framework on the left, we might say that argument a is not known because it doesn't belong to any agent. And for

the framework on the right, we might say that the attack is unknown because no agent holds both arguments a and b. The filtering methods remove such unknown arguments and unknown attacks. This is followed by the application of Dung's semantics.

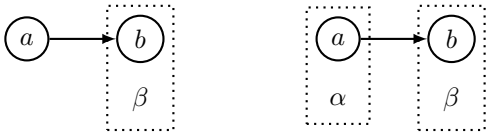


Figure 15: Unknown argument and unknown attack

There are several aspects of dialogue beyond associating arguments with agents that can be represented in abstract argumentation. One significant aspect is to make *time* explicit: unlike inference, dialogue inherently unfolds over time, with the dynamic argumentation framework evolving as the dialogue progresses. Dialogue can also be *strategic* — sometimes it is advantageous for an agent to not reveal certain arguments (this is discussed further in Section 4.3). A prototypical example of this is the content of the Miranda warning: "Anything you say can and will be used against you in a court of law". Example 3.19 illustrates how a suspect's argument can be strategically turned against him/her in a dialogue.

Example 3.19 (Dialogue between accuser and suspect [Okuno and Takahashi, 2009]). Let Pr and Op be the players involved in the following argumentation dialogue (Pr and Op denote, respectively, a proponent and an opponent):

$\mathbf{Pr_0}$: "You killed the victim."

$\mathbf{Op_1}$: "I did not commit murder! There is no evidence!"

$\mathbf{Pr_1}$: "There is evidence. We found your ID card near the scene."

$\mathbf{Op_2}$: "That is not evidence! I had my ID card stolen!"

$\mathbf{Pr_2}$: "It is you who killed the victim. Only you were near the scene at the time of the murder."

$\mathbf{Op_3}$: "I did not go there. I was at facility A at that time."

$\mathbf{Pr_3}$: "At facility A? Then, it is impossible that you had your ID card stolen because facility A does not allow any person to enter without an ID card."

In this example, the opponent tries to defend himself with the claim "I had my ID card stolen!" (Op_2). However, the proponent strategically uses this very claim against the opponent (Pr_3), arguing that if the opponent was at facility

A, it would have been impossible for his ID card to have been stolen because the facility does not permit entry without an ID card. This demonstrates how an argument can backfire in a strategic dialogue.

In this section, we discussed extending the attack-defense paradigm for dialogue, particularly with agents. We end the section with the following questions:

1. In this section, we discussed abstract agent argumentation, and we provided various semantics. How should a theory of structured agent argumentation be designed?

2. Strategic dialogue goes beyond argumentation as inference by incorporating agency. How should dialogue strategies be designed? And should they be evaluated?

3. What is the next step required to bridge the gap between 1) Dung's attack-defense paradigm and 2) strategic argumentation and dialogue?

4. There are many kinds of dialogues. What are the main components of a dialogue? For example, what are the components of persuasion dialogue systems like Fatio?

5. For all these kinds of dialogue, what makes a good dialogue? For example, what is a successful Fatio dialogue? Does a successful dialogue happen when someone is convinced of an argument they did not hold previously or does it happen when the parties agree about where they disagree?

3.5 Extensions of the attack-defense paradigm for balancing

In this section, we continue our discussion of extensions to abstract argumentation, focusing on extensions inspired by argumentation as balancing.

Argumentation as balancing brings to mind a double pan scale. The pros go on one pan and the cons go on the other. The pros and cons may have relative weights, and one needs to balance them from a utilitarian lens to determine the status of the issues, e.g., what action to take. Balancing finds applications in ethics and the law. In the legal domain, balancing is a metaphoric term that is generally used to describe an important conceptual operation [Aleinikoff, 1986]. In many conflicts, there is something to be said in favor of two or more outcomes. Whatever result is chosen, someone will be advantaged and someone will be disadvantaged; some policies will be promoted at the expense of some others. Hence it is often said that a "balancing operation" must be undertaken, with the "correct" decision seen as the one yielding the greatest net benefit. In medical ethics, for example, there are models of clinical ethics case consultation that often refer to 'balancing' or 'weighing' moral considerations [McDougall *et al.*, 2020].

Challenge 9. Generalizing Dung's attack-defense paradigm for balancing.

At an abstract level, it seems that pro and con arguments and the relations between them can be represented intuitively in bipolar argumentation frameworks discussed by Cayrol and Lagasquie-Schiex [2005; 2009; 2010; 2013], extending abstract argumentation framework with support relation that is independent of attack. Figure 16 illustrates three bipolar argumentation frameworks, where attack relations are depicted by solid arrows, and support relations are depicted by dashed arrows. Similarly to abstract agent argumentation semantics, there are also three types of bipolar argumentation semantics defined by Yu *et al.* [2023].

The defense-based approach defines new notions of defense using both support and attack.

The selection-based approach utilizes support only for selecting some of the extensions provided in Dung's semantics.

The reduction-based approach introduces indirect attacks based on interpretations of support.

There are three new defense based on both attack and support relations, called defense_1, defense_2, and defense_3, all of which have additional requirements for Dung's defense. Defended_1 requires that the argument defends (in Dung's theory) another argument also supports it. Defended_2 requires that a defender is supported. Moreover, defended_3 requires not only that the attackers are attacked, but also that all supporters of the attackers are attacked as well. We illustrate the three defense in Figure 16.

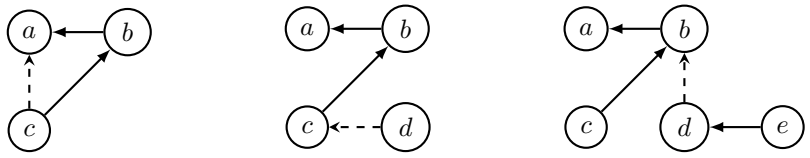


Figure 16: Three bipolar argumentation frameworks (BAFs) illustrating the three defense notions. In the left hand framework, $\{c\}$ defends_1 a. In the middle framework, $\{c, d\}$ defends_2 a. In the right hand framework, $\{c, e\}$ defends_3 a.

The selection-based approach uses support during the post-processing step for Dung's theory of abstract argumentation [Gargouri *et al.*, 2020], i.e., first Dung's semantics are obtained, then support can be used to select extensions from Dung's semantics. One way selects the extensions that have the largest number of internal supports. This reflects the idea that for a coalition, the more internal supports they have, the more cohesive they are. The other way is to select the extensions that receive the most support from outside. This reflecting the idea that the more support a coalition receives, the stronger it is. It thus interprets support as a kind of voting procedure. We say that argument b in E

is internally supported if b receives support from arguments in E. Argument b in E is externally supported if b receives support from arguments that are outside E.

Example 3.20 (Selection-based approach to bipolar argumentation). Consider the bipolar argumentation framework on the right hand side of Figure 17. There are four extensions to Dung's stable semantics: $\{\{a, d\}, \{a, c\}, \{b, d\}, \{b, c\}\}$. By following the selection based on internal supports, $\{\{a, c\}\}$ is the stable semantics, while by following selection based on external supports, $\{\{a, d\}\}$ is the stable semantics.

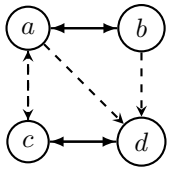


Figure 17: A bipolar argumentation framework

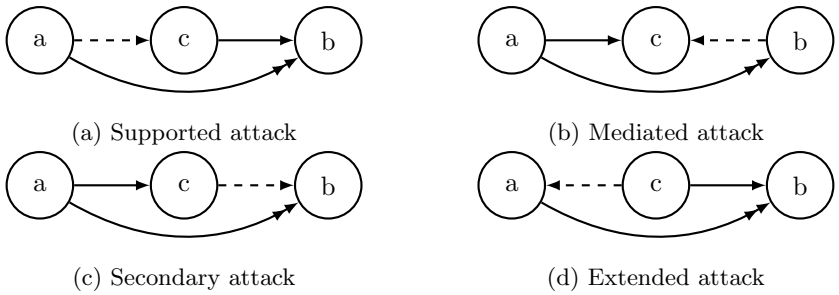
The reduction-based approach has been studied extensively by Cayrol and Lagasquie-Schiex [2005; 2009; 2013], and support is used as pre-processing for Dung semantics. The corresponding abstract argumentation frameworks are reduced by adding indirect attacks from the interaction between attacks and supports with different interpretations, i.e., deductive support and necessary support. Based on these two interpretations, four reductions have been discussed introducing additional attacks.

Deductive support [Boella *et al.*, 2010] captures the intuition that if a supports b, then the acceptance of a implies the acceptance of b. Based on deductive interpretation, there are two kinds of additional attacks:

Supported attack and mediated attack. For example, in Figure 18(a), a supports c, and c attacks b. Acceptance of a implies acceptance of c, and acceptance of c implies non-acceptance of b. So, acceptance of a implies non-acceptance of b. Thus, the supported attack from a to b is introduced, depicted as a double-headed arrow. Similarly, the mediated attack is visualized in Figure 18(b).

Necessary support [Nouioua and Risch, 2010] captures the intuition that if a supports b, then the acceptance of a is necessary to obtain the acceptance of b, or equivalently, the acceptance of b implies the acceptance of a.

Secondary attack and extended attack For example, in Figure 18.(c), a attacks c, c supports b. The acceptance of a implies the non-acceptance of c and the non-acceptance of c implies the non-acceptance of b; so, the acceptance of a implies the non-acceptance of b. Thus, the secondary attack from a to b is introduced, depicted as a double-headed arrow. Similarly, the extended attack is visualized in Figure 18.(d).



(a) Supported attack (b) Mediated attack

(c) Secondary attack (d) Extended attack

Figure 18: Four kinds of reductions of bipolar argumentation frameworks

We now use the child custody example to illustrate reduction-based semantics.

Example 3.21 (Child custody in bipolar argumentation). Consider the bipolar framework visualized below. The figure should be read as follows. Normal arrows are attack relations, dashed arrows are support relations, a double box represents a prima facie argument which is self-supporting, and a single box represents a standard argument that does not support itself. Our focus is on how to interpret the support from (OP): "Child wants to live with her mother" to (M): "Child's best interest is that she lives with her mother." For a comprehensive analysis, we refer to the work of Yu *et al.* [2020]. The supporting argument (OP) might have a special status because of the rules of the Civil Code: *the judge has to take the child's opinion into consideration when deciding about custody.* Analysis of this rule shows how various interpretations of the support interpretations relate to legal interpretations. We assume that the child wants to live with her mother (OP). What does this mean? One can say that the obligation to take argument (OP) into consideration means that (OP) is a prima facie argument and thus has to be accepted. If it is a prima facie argument, (M) receives the evidentiary support it needs. But this in itself doesn't decide how argument (OP) affects the extension. The extension depends on how we interpret the support relation between (OP) and (M): deductive or necessary. It seems very intuitive to interpret the support relation as deductive: the obligation to take the child's opinion into consideration is apparently very much in line with what deductive support means: if we accept the child's opinion (which is prima facie) then we have to accept (M) too. We assume the support from W: "Father is wealthy because he inherited" to F: "Child's best interest is that she lives with her father", is not deductive.

Such an analysis contributes to the discussion on the formalization of legal interpretation in the following way. The role of interpretation is crucial in the law, but it is also a source of criticism of the use of logic-based methods for modeling legal reasoning. For example, Leith warns that the knowledge engineer's interpretation when formalizing norms is necessarily premature because the authority for interpreting the law has been assigned to the judiciary [Prakken,

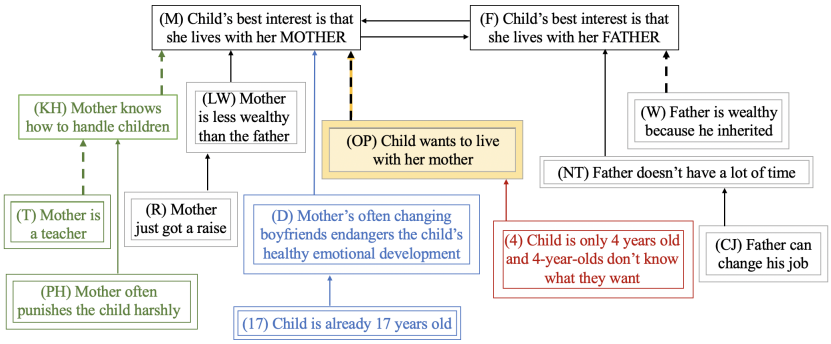


Figure 19: A child custody deliberation with possible arguments and their relations in a divorce case [Yu *et al.*, 2020].

2013]. Addressing this criticism, the literature on legal interpretation has discussed the possibility that legal knowledge-based systems contain alternative syntactic formalizations. It has been observed that while, on the syntactic level, formalization commits us to a given interpretation, on the conceptual level, classification of factual situations as legal concepts is not an issue of logical form [Prakken, 2013]. Alternatively, we can restrict the investigation by saying that "the only aspects of legal reasoning that can be formalized are those aspects that concern the following problem: *given* a particular interpretation of a body of legal knowledge, and *given* a particular description of some legal problem, what are the general rational patterns of reasoning with which a solution to the problem can be obtained?" [Prakken, 2013]. If a formal framework offers the different interpretations itself, though, then using it might be directly exploitable to the comparison of the different possibilities and routes of reasoning given each interpretation.

It has been argued that for the validation of a bipolar argumentation theory, so-called theory-based validation is preferable to empirical validation [Prakken, 2020; Polberg and Hunter, 2018], which itself is preferable to intuition-based validation. Nevertheless, in this context, the principle-based analysis discussed in Section 4.1 complements these validation methods. The theory of formal argumentation needs to be complemented with examples and case studies concerning the use of the theory.

In this section, we discussed extending the attack-defense paradigm for balancing. We end this section with the following questions:

1. The attack-defense paradigm introduced the distinction between structured and abstract argumentation. For every extension, we need to decide whether to represent it at the structured or abstract level. For example, should justification (as in Fatio) be expressed as in structured argumentation (within the argument) or as a support (among arguments)?

2. When we introduce a new concept like support at the abstract level, it can be interpreted in various ways at the structured level. For example, what does support mean other than inferential relation (e.g., in ASPIC+)?

3. How can we better represent argumentation as balancing (e.g., in law and ethics)? For example, how should the pros and cons be aggregated? What is the role of weights?

4. How should other aspects of balancing be incorporated? For example, how should argument strength be represented and evaluated? It is important to distinguish between different kinds of argument strength, in particular logical, dialectical and rhetorical argument strength [Prakken, 2024a].

5. In the previous and present sections, we discussed extended abstract argumentation inspired by dialogue and balancing. Which other inspirations can be utilized to design extensions of abstract argumentation? For example, how can natural argumentation inspire extensions of abstract argumentation?

4 After the paradigm shift: the computational turn

In this section, we transition from the paradigm shift in formal argumentation to computational argumentation. Section 4.1 introduces principle-based analysis as a methodology for handling the diversity of argumentation models at a higher level of abstraction, providing a systematic approach to designing and choosing methods for different computational contexts. In Section 4.2, we focus on compositionality principles such as locality, modularity, and transparency, which play a central role from the attack-defense perspective and are exploited in algorithmic strategies like divide and conquer. Section 4.3 discusses the relationship between explanation and argumentation, highlighting, for example, strategic argumentation, discussion games and reason-based models for understanding argumentation as dialogue, inference, and balancing, respectively. Finally, Section 4.4 addresses integrating argumentation techniques with existing and emerging technologies, showcasing the potential of distributed argumentation systems and their applications in diverse technological contexts.

4.1 A principle-based analysis

In this section, we turn towards computational argumentation by discussing methodology. Principle-based analysis is a methodology for managing the diversity of argumentation, such as when selecting among existing methods or designing new ones. Principles describe formal argumentation at a higher level of abstraction, and a wide range of principles exists across all models of argumentation.

The principle-based approach is also called the axiomatic approach and the postulate-based approach. Principles are properties, while postulates are normally desirable properties or requirements. This approach is particularly useful

when there is a diversity of alternative methods. It has been successfully applied in various areas. For example, Alchourrón-Gärdenfors-Makinson (AGM) postulates [Alchourrón *et al.*, 1985] have been applied in belief revision operations to ensure rationality, and axiomatic principles are applied when searching for and choosing suitable voting rules for various contexts.

The principle-based approach has also been used to describe formal argumentation at a higher level of abstraction. Abstraction in mathematics is the process of extracting the underlying structures, patterns or properties of a mathematical concept. In software engineering and computer science, abstraction is the process of generalizing concrete details. In formal argumentation, one form of abstraction is to focus on the attack and defense relations between arguments rather than their internal structures. The attack-defense relation is used to define the functions of semantics. Principles can be defined as sets of such functions and are represented as constraints on such functions. Principles can be used to compare or define new functions.

The challenge addressed in this section is:

Challenge 10. Conducting a principle-based analysis of argumentation.

There is a diversity of principles and postulates in all models of argumentation, particularly in the context of argumentation as inference, less so in argumentation as dialogue and balancing. To illustrate, let us reuse the commutative diagram in Figure 20 featuring examples of principles that are used for different purposes. For step (1), there are Kraus-Lehmann-Magidor (KLM) principles [Kraus *et al.*, 1990] that a logical inference relation ought to satisfy. For structured argumentation in steps (2-4), there are axiomatic analyses of various attack relation assignments among arguments [Dung, 2016; Dung and Thang, 2018; Pardo *et al.*, 2024], as discussed in Section 3.2. For the whole commutative diagram, rationality postulates are used to ensure that the conclusions drawn at the end of the overall process have desirable properties [Caminada and Amgoud, 2005; Caminada and Ben-Naim, 2007; Caminada, 2018b], as discussed in Section 3.3. For step (3), there is a diversity of semantics available for the abstract argumentation framework. Baroni and Giacomin [2007] classified argumentation semantics based on a set of principles, which was extended by van der Torre and Vesic [2018]. For diverse extended argumentation frameworks, with even more semantics. There are the principles-based analysis of ranking-based semantics, multiagent argumentation [Yu *et al.*, 2021], and bipolar argumentation [Yu *et al.*, 2023].

There are three steps in a principle-based approach [van der Torre and Vesic, 2018].

Define the function that will be the object of the study. For instance, abstract argumentation semantics are functions that map graphs to sets of sets of graph nodes.

Define the principles — examine the relations between functions and principles to see if the semantics satisfy the principles.

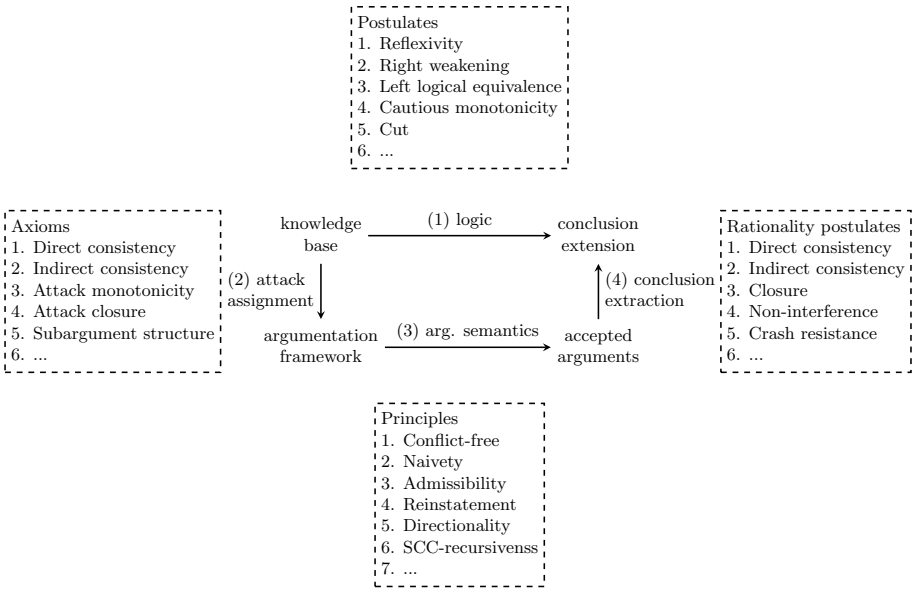


Figure 20: Principle-based analysis in commutative diagram

Classify and study the sets of principles — study the relations between principles. For example, a set of principles may imply another principle. Or there may be incompatibilities among principles. Or there may be a set of principles that characterizes a semantics.

There are three main branches of abstract argumentation semantics.

Admissibility-based semantics (AB) uses gunfight rules requiring that an extension E defends itself against all attackers [Dung, 1995], i.e., whenever each argument attacking E from the outside is itself attacked by some element of E.

Weak admissibility-based semantics (WA) only requires that an extension E defends itself against reasonable arguments [Baumann *et al.*, 2020].

Non-admissibility based semantics (NA) requires an extension E to be a maximal conflict-free set of arguments. The most prominent example of non-admissibility based semantics is CF2 semantics [Baroni *et al.*, 2005].

We illustrate the above three branches of semantics with Example 4.1 below.

Example 4.1 (Three branches of abstract argumentation semantics). Consider the three frameworks in Figure 21. For framework (a), the only extension in AB semantics is the empty set whereas under the CF2 semantics,

b is accepted. To get the desirable properties of directionality and strongly-connected-component (SCC) recursiveness (discussed further below), CF2 is defined in terms of a local function that computes the maximal conflict-free subsets for each strongly connected component of a framework. Under WA semantics, the set of weakly preferred extensions of framework (a) is $\{\{b\}\}$ because its only attacker a is self-attacking. It is like a "zombie", it is there but it can do no harm [Baumann *et al.*, 2020]. For framework (b), the set $\{d\}$ is not admissible because it does not defend itself from b. Nevertheless, under WA semantics, d is acceptable because b is part of an odd-length cycle of arguments that are never accepted, and so it does not pose an actual threat. These extensions are listed in Table 6.

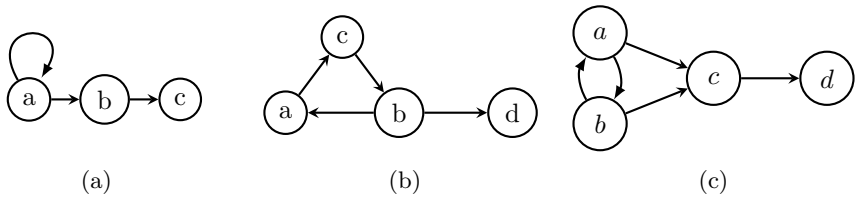


Figure 21: Three argumentation frameworks

Semantics	**(a)**	**(b)**	**(c)**
AB	$\{\emptyset\}$	$\{\emptyset\}$	$\{\{a,d\},\{b,d\}\}$
NA(CF2)	$\{\{b\}\}$	$\{\{a,d\},\{c,d\},\{b\}\}$	$\{\{a,d\},\{b,d\}\}$
WA	$\{\{b\}\}$	$\{\{d\}\}$	$\{\{a,d\},\{b,d\}\}$

Table 6: Three semantics applied to the frameworks in Figure 21. AB = admissibility-based; NA = non-admissibility based; WA = weak-admissibility based.

To compare the diverse semantics, we can categorize formal argumentation principles into three types: traditional principles (the most discussed), variants of traditional principles, and new principles specifically designed for emerging semantics. Below, we provide examples of these principles to illustrate how they are used. This will enable us to compare the different branches of argumentation semantics as well as the different agent argumentation semantics described in Section 3.4.

We list some of the traditional principles that have been used to compare these semantics.

Conflict-freeness states that every extension under semantics is a conflict-free set, i.e., there is no attack relation among the arguments in the extension.

Admissibility is satisfied by a semantics if and only if every extension under that semantics is admissible.

Naivety states that every extension under the semantics is a maximal conflict-free set.

Directionality states that an argument a should be affected only by a's attacker. The arguments that only receive an attack from a should not have any effect on the state of a.

SCC (strongly-connected-component) recursion states that extension construction carried out in an initial SCC do not depend on those concerning the other ones, while they directly affect the choices about the subsequent SCCs and so on.

Modularity states that the semantics of a framework can be obtained by the semantics of the smaller parts of that framework.

Example 4.2. Given the framework shown in Figure 22, the complete and weak complete semantics are the same: $\{\emptyset\}$, while $\emptyset$ is not a maximal conflict-free set in this framework, e.g., $\{a\}$ is also conflict-free.

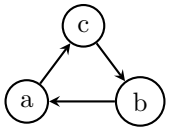


Figure 22: Complete and weak complete semantics do not satisfy naivety

In Section 3.4, we discussed abstract agent argumentation [Yu *et al.*, 2021]. The analysis focused on the four traditional Dung semantics (complete, preferred, stable, and grounded) denoted as TR. Two new concepts, individual defense and collective defense, have been introduced. By applying these concepts to each of the four traditional semantics, there are two new variants for each: one based on individual defense and the other on collective defense. This results in a total of eight distinct defense-based semantics, denoted as Sem_1 and Sem_2. Additionally, social agent semantics, which prioritizes arguments supported by more agents, produce sixteen semantics through four reductions (denoted as $SR_1 - SR_4$) from preference-based argumentation frameworks to abstract argumentation frameworks. Agent reduction semantics, which considers the perspective of individual agents, also yields sixteen semantics through the four reductions, denoted as $AR_1 - AR_4$. Lastly, agent filtering semantics, inspired by limited knowledge, introduces eight additional semantics, denoted as OR and NBR. Altogether, this results in a total of fifty-two semantics.

The paper provides a full analysis of fifty-two agent semantics, including Dung's traditional semantics, with seventeen principles. The results of

principle-based analysis are typically summarized in tables, as seen in Table 7, 8, 9. These principles are categorized into three groups: five traditional principles (Table 7), four variations of these traditional principles (Table 8), and eight new principles specifically designed for agent-based argumentation (Table 9).

Semantics	P1	P2	P3	P4	P5
TR	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGP}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$
Sem_1	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGP}$	$\mathbb{S}$	$\mathbb{S}$
Sem_2	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGP}$	$\mathbb{S}$	$\mathbb{S}$
SR_1	$\times$	$\times$	$\mathbb{CGP}$	$\times$	$\times$
SR_2	$\mathbb{CGPS}$	$\times$	$\times$	$\times$	$\times$
SR_3	$\times$	$\mathbb{CGPS}$	$\mathbb{CGP}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$
SR_4	$\mathbb{CGPS}$	$\mathbb{G}$	$\times$	$\times$	$\mathbb{G}$
AR_1	$\times$	$\times$	$\mathbb{CGP}$	$\times$	$\mathbb{S}$
AR_2	$\mathbb{CGPS}$	$\times$	$\times$	$\times$	$\times$
AR_3	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGP}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$
AR_4	$\mathbb{CGPS}$	$\mathbb{G}$	$\times$	$\times$	$\mathbb{G}$
OR	$\mathbb{CGPS}$	$\times$	$\mathbb{CGP}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$
NBR	$\times$	$\times$	$\mathbb{CGP}$	$\times$	$\mathbb{S}$

Table 7: Comparison of abstract agent argumentation semantics and traditional principles. When a principle is never satisfied by a certain reduction for all semantics, we use the $\times$ symbol, and we use a question mark to represent an open problem. P1 refers to Principle 1, and the same convention holds for all the others. P1 = conflict-free, P2 = admissibility, P3 = directionality, P4 = SCC-recursiveness, P5 = modularity.

For example, the agent admissibility principle is a variation of the traditional admissibility principle, with agent defense replacing the standard notion of defense. Since admissibility can be applied to either individual defense or collective defense, this gives rise to two distinct agent admissibility principles. Similarly, the agent SCC-recursiveness principles are adapted to reflect the concepts of individual and collective defense, resulting in two corresponding principles. Additionally, eight entirely new agent principles have been introduced to address the unique aspects of agent-based argumentation, and these are shown in Table 9.

Below we list the eight new principles. Some of them are expected to be satisfied by all of the semantics. Some can be used to distinguish between different semantics, since only certain semantics satisfy or do not satisfy certain principles.

Principle 10: AgentAdditionPersistence states that if more agents adopt an argument that is already accepted, this does not affect the extension.

Semantics.	P6	P7	P8	P9
TR	×	×	×	×
Sem_1	$\mathbb{CGP}$	$\mathbb{CGP}$	$\mathbb{S}$	$\mathbb{S}$
Sem_2	×	$\mathbb{CGP}$	$\mathbb{S}$	$\mathbb{S}$
SR_1	×	×	×	×
SR_2	×	×	×	×
SR_3	×	×	×	×
SR_4	×	×	×	×
AR_1	×	×	×	×
AR_2	×	×	×	×
AR_3	×	×	×	×
AR_4	×	×	×	×
OR	×	×	×	×
NBR	×	×	×	×

Table 8: Comparison of the reductions and agent admissibility principles, and agent SCC-recursion. P6 = agent admissibility$_1$, P7 = agent admissibility$_2$, P8 = agent SCC-recursiveness$_1$, P9 = agent SCC-recursiveness$_2$.

Principle 11: AgentAdditionUniversalPersistence reflects the same idea as principle 10 but is based on the assumption that a is accepted in all extensions.

Principle 12: PermutationPersistence reflects a principle we expect to hold for all agent semantics — anonymity. If we permute the agents, it does not affect the extensions. This principle is analogous to language independence for arguments, as defined by Baroni and Giacomin [2007].

Principle 13: MergeAgent states that if the arguments of two agents do not attack each other, we can merge these agents into one single agent. The principle does not hold for agent defense semantics because new agent defenses may be created.

Principle 14: RemovalAgentPersistence states that if two agents have the same arguments, we can remove one of these agents without changing the extensions. This represents the opposite of social semantics, where the number of agents does make a difference.

Principle 15: AgentNumberEquivalence is inspired by social agent semantics. It states that where there are two argumentation frameworks with the same arguments and attacks, if for every argument the number of agents holding that argument is the same, then the extensions are the same.

Principle 16: ConflictfreeInvolvement is inspired by agent reduction semantics. It states that if the set of an agent's arguments is conflict-free, then there is an extension containing those arguments.

Principle 17: RemovalArgumentPersistence is inspired by OrphanReduction semantics. It states that if we have arguments that do not belong to any agents, then they can be removed from the framework without affecting the extensions.

In the resulting Table 9, all the agent semantics satisfy P12. Perhaps surprisingly, neither social agent semantics nor agent reduction semantics satisfy P10 while trivial reduction semantics, social agent semantics, and agent filtering semantics satisfy P13. Moreover, all agent semantics except social agent semantics satisfy P14. No semantics satisfy P16. As expected, only OrphanRemoval satisfies P17. The only semantics that are not distinguished yet concern the use of different preference reductions, or different Dung semantics. To distinguish between these, the principles proposed in preference-based argumentation [Kaci *et al.*, 2021] and in Dung's semantics can be used [Baroni and Giacomin, 2007; van der Torre and Vesic, 2018]. In that sense, the principle-based analyses can complement one other.

Sem.	P10	P11	P12	P13	P14	P15	P16	P17
TR	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	×	×
Sem_1	$\mathbb{S}$	$\mathbb{S}$	$\mathbb{CGPS}$	×	$\mathbb{CGPS}$	×	×	×
Sem_2	$\mathbb{S}$	$\mathbb{S}$	$\mathbb{CGPS}$	×	$\mathbb{CGPS}$	×	×	×
SR_1	×	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	×	$\mathbb{CGPS}$	×	×
SR_2	×	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	×	$\mathbb{CGPS}$	×	×
SR_3	×	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	×	$\mathbb{CGPS}$	×	×
SR_4	×	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	×	$\mathbb{CGPS}$	×	×
AR_1	×	$\mathbb{CGPS}$	$\mathbb{CGPS}$	×	$\mathbb{CGPS}$	×	×	×
AR_2	×	$\mathbb{CGPS}$	$\mathbb{CGPS}$	×	$\mathbb{CGPS}$	×	×	×
AR_3	×	$\mathbb{CGPS}$	$\mathbb{CGPS}$	×	$\mathbb{CGPS}$	×	×	×
AR_4	×	$\mathbb{CGPS}$	$\mathbb{CGPS}$	×	$\mathbb{CGPS}$	×	×	×
OR	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	×	$\mathbb{CGPS}$
NBR	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	×	×	×

Table 9: Comparison between the reductions and new agent principles. P10 = AgentAdditionPersistence, P11 = AgentAdditionUniversalPersistence, P12 = PermutationPersistence, P13 = PermutationPersistence, P14 = RemovalAgentPersistence, P15 = AgentNumberEquivalence, P16 = ConflictfreeInvolvement, Principle 17 = RemovalArgumentPersistence.

Finally, the principle-based approach to formal argumentation may lead to the study of impossibility and possibility results. For instance, Arrow's impossibility theorem in voting and social choice theory demonstrates that no voting

system can simultaneously satisfy the whole set of seemingly reasonable criteria — non-dictatorship, unrestricted domain, Pareto efficiency, and independence of irrelevant alternatives — when there are three or more options available. This kind of result highlights the inherent trade-offs involved in designing systems that attempt to balance competing principles. Similarly, as discussed in Section 3.2, any attempt to realize PDL in ASPIC+ should preserve the definitional principle of attack closure. The impossibility theorem explains how this is incompatible with context independence [Pardo *et al.*, 2024]. These impossibility results are crucial because they reveal the boundaries of what can be achieved within a given formal system. Additionally, they also guide researchers to either accept certain trade-offs or seek alternative approaches that might circumvent these limitations.

In this section, we discussed principle-based analysis and, as usual, we list several research questions about that topic.

1. How can we provide guidance to users who are not experts in formal or computational argumentation on how to use the theory of argumentation for their needs? Can we develop a user guide for the theory of argumentation?

2. How do we decide which conceptualization of formal argumentation to use for an application (argumentation as inference, dialogue or balancing), and how do we connect or combine these conceptualizations?

3. What needs to be changed to move from constructing comparison tables (as shown in Tables 7-9) to characterization, or proving possibility and impossibility results? For example, how to characterize last vs. weakest link in structured argumentation, or characterize various kinds of abstract argumentation semantics?

4. Which methodology can be developed for formal and computational argumentation to guide the search for and design of principles? For example, the principle of resolution was defined by Baroni and Giacomin [2007], then the resolution-based semantics were defined and studied by Baroni *et al.* [2011].

5. How can we combine principles from various extended argumentation frameworks? For example, reductions in preference-based argumentation often remove attacks whereas reductions in bipolar argumentation often add attacks.

4.2 Algorithmic argumentation

In this section, we consider the role of principles in algorithmic argumentation. Algorithmic argumentation, as illustrated in Table 1, refers to a step-by-step procedure or set of rules designed to perform a specific task or solve a particular argumentation problem. We focus mainly on the calculation of argumentation

semantics. On the one hand, compositionality principles play a central role in the attack-defense perspective. On the other hand, algorithms and other computational approaches exploit these principles. We illustrate this by discussing locality, modularity, and transparency principles on one side, and "divide and conquer" and robustness principles on the other.

Traditionally, Dung's abstract argumentation frameworks are viewed as monolithic entities where various semantics are applied globally to determine which arguments are acceptable. While this unified approach preserves generality, it has been shown to be computationally intractable. That complexity presents the following challenge:

Challenge 11. Designing efficient algorithms for argumentation semantics.

The idea of compositionality is that an abstract argumentation framework is broken down into interacting smaller subframeworks. This motivates a *local focus* and further investigation into locality and modularity principles in abstract argumentation. Related principles are, for example, directionality, SCC-recursiveness, and decomposability.

The directionality property corresponds to the idea that the attack relation encodes a form of dependency and that arguments can affect one other only by following the direction of the attacks. This consideration can be extended from individual arguments to sets of arguments. If a set of arguments is unattacked (i.e., it does not receive attacks from arguments outside the set) it should not be affected by the rest of the argumentation framework. In more formal terms, given an argumentation semantics, projecting the semantics of the global framework onto an unattacked set should result in the semantics producing an evaluation of an argumentation framework consisting of only that unattacked set.

Example 4.3 illustrates directionality and why stable semantics does not satisfy this property.

Example 4.3 (Directionality [Baroni and Giacomin, 2007]). Consider the stable semantics of the argumentation framework in Figure 5. For the subframework consisting of $\{a, b\}$, the stable semantics is $\{\{a\}, \{b\}\}$. The stable semantics of the whole framework is $\{\{b, d\}\}$ whose projection — the unattacked set $\{a, b\}$ — is $\{b\}$. However, $\{\{b\}\}$ does not coincide with the stable semantics of the unattacked framework. This is a counterexample proving that stable semantics does not satisfy the directionality.

The property of SCC-recursiveness [Baroni *et al.*, 2005] is based on decomposition along the SCCs of the argumentation framework. Different from directionality, SCC-recursiveness has a direct constructive interpretation. The structure of the argumentation frameworks drives the incremental definition of extensions step by step. Without complex technicalities, we illustrate SCC-recursiveness with Example 4.4.

Example 4.4 (SCC-recursiveness [Baroni *et al.*, 2005]). Consider again the framework in Figure 5.

Step 1 Partition the argumentation framework into SCCs. There are two SCCs in the framework: $S_1 = \{a, b\}$ and $S_2 = \{c, d, e\}$. Here, S_1 is identified as the initial SCC as it does not depend on S_2.

Step 2 Construct extensions incrementally using a base function. Determine the possible extensions within each initial SCC using a semantic-specific base function. This function returns the extensions for argumentation frameworks consisting of a single SCC. For the SCC of $S_1 = \{a, b\}$, the base function provides two possible partial candidate extensions: $\{a\}$ and $\{b\}$.

Step 3 For each possible extension determined in Step 2, apply the reinstatement principle. This involves suppressing the nodes directly attacked within subsequent SCCs and considering the distinction between defended and undefended nodes. Let's take candidate extension $\{b\}$. Here, argument c in S_2 cannot be included in the extension because it is attacked by argument b. Therefore, only $\{d, e\}$ can be taken into consideration.

Step 4 Recursively apply the steps on restricted frameworks. Consider the subframework $(\{d, e\}, \{(d, e)\})$. This subframework is again partitioned into SCCs, resulting in $S_1' = \{d\}$ and $S_2' = \{e\}$. Considering $S1' = \{d\}$, since e is attacked by d, e is excluded. Hence, the only extension left is $\emptyset$. Thus, the final extension is $\{b\} \cup \{d\} \cup \emptyset = \{b, d\}$.

The idea of decomposability is to break down an abstract argumentation framework arbitrarily into interacting smaller subframeworks called modules. Input/Output frameworks have been defined on this basis [Baroni *et al.*, 2014]. Each module can be described as a black box whose Input/Output behavior — specifically referring to the labeling — fully determines its role in the system's global behavior. That makes it possible to describe and analyze the framework's global behavior in terms of the combination of the local behaviors of its constituent modules. Each local behavior can be characterized individually. This characterization is independent of the internal details of other modules. Instead, it focuses only on the connections and mutual interactions between the module and the other modules. Additionally, if two modules have the same input and output behavior, they are interchangeable in a way that does not influence the global behavior.

Example 4.5 illustrates the interfaces of subframeworks.

Example 4.5 (Interface [Baroni *et al.*, 2014]). Given the abstract argumentation framework (AF) visualized in Figure 23 and the subframeworks induced by the sets $\{a, b, c\}$ and $\{d\}$, these subframeworks are denoted as $\mathrm{AF}_{\downarrow\{a,b,c\}}$ and $\mathrm{AF}_{\downarrow\{d\}}$. The subframeworks interact with one other through the attacks $a \to d$ and $d \to a$ respectively. For $\mathrm{AF}_{\downarrow\{a,b,c\}}$, the interface, or input argument, is argument d, while for $\mathrm{AF}_{\downarrow\{d\}}$, the input argument, or interface, is argument a.

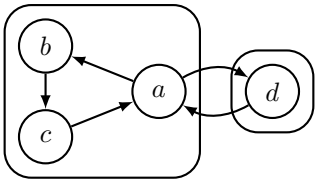


Figure 23: A partition of an abstract argumentation framework

A local function determines the labeling of a subframework based on the labeling of external input arguments, ensuring that the internal labeling of a subframework is consistent with its external influences. Example 4.6 illustrates how external input arguments enforce the internal labelings.

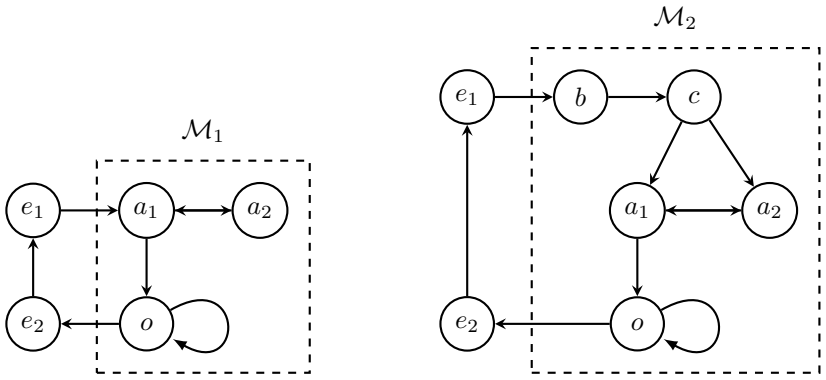
Example 4.6 (External input arguments enforcement [Baroni *et al.*, 2014]). Consider the argumentation framework in Example 4.5. If we apply the local function to subframework $\mathrm{AF}_{\downarrow\{a,b,c\}}$ with the external argument d, if d is labeled as *in*, the resulting labeling of the subframework is {(a, out), (b, in), (c, out)}. If d is labeled as *out*, the resulting labeling of the subframework is $\{\{(a, \text{undec}), (b, \text{undec}), (c, \text{undec})\}\}$. If d is labeled as *undec*, the resulting labeling of the subframework is $\{\{(a, \text{undec}), (b, \text{undec}), (c, \text{undec})\}\}$. We can apply the same analysis for the subframework $\mathrm{AF}_{\downarrow\{a,b,c\}}$ with external argument a.

The property of *decomposability* states that given an arbitrary partition of an argumentation framework into a set of subframeworks, the outcomes produced by a given semantics can be obtained as a combination of the outcomes produced by a local counterpart applied separately on each subframework and vice versa.

Example 4.7 (Decomposability [Baroni *et al.*, 2014]). Considering again the argumentation framework in Example 4.5 and the partition $\{\{a, b, c\}, \{d\}\}$, the decomposability of the complete semantics requires a local function such that the labelings of AF are exactly those obtained by the union of the compatible labelings of $\mathrm{AF}_{\downarrow\{a,b,c\}}$ and $\mathrm{AF}_{\downarrow\{d\}}$ given by the local function itself. The labeling $\{(a, \text{out}), (b, \text{in}), (c, \text{out})\}$ is compatible with $\{(d, \text{in})\}$. The first is obtained with d labeled *in*, and the latter is obtained with a labeled *out*. On the other hand, the labeling $\{(a, \text{out}), (b, \text{in}), (c, \text{out})\}$ is not compatible with, e.g., $\{(d, \text{out})\}$. Overall, exactly two global labelings arise from combining the compatible outcomes — $\{(a, \text{undec}), (b, \text{undec}), (c, \text{undec}), (d, \text{undec})\}$ and $\{(a, \text{out}), (b, \text{in}), (c, \text{out}), (d, \text{in})\}$, which corresponds to the complete labelings of the whole AF.

The property of *transaparency* states that if two modules have the same Input/Output behavior, then we can replace one with the other without influencing the framework's global behavior. This ensures that the invariant part of the framework is unaffected. Example 4.8 illustrates the transparency property.

Example 4.8 (Transparency [Baroni *et al.*, 2014]). Consider the argumentation frameworks AF_1 and AF_2 shown in Figure 24. $\mathcal{M}_1$ and $\mathcal{M}_2$ have the same Input/Output behavior, i.e., they are equivalent under preferred semantics. The invariant set of the replacement is the set $\{e_1, e_2\}$. However, after the replacement of $\mathcal{M}_1$ by $\mathcal{M}_2$ in AF_1, the preferred extension changes. In fact, the preferred labelings of AF_1 are $\{(a_1, in), (a_2, out), (o, out), (e_2, in), (e_1, out)\}$ and $\{(a_1, out), (a_2, in), (o, undec), (e_2, undec), (e_1, undec)\}$, while $\{(b, in), (c, out), (a_1, in), (a_2, out), (o, out), (e_2, in), (e_1, out)\}$ is the only preferred labeling of AF_2.



(a) AF_1 (b) AF_2

Figure 24: Preferred semantics does not satisfy transparency

Now, we illustrate how to exploit modularity, which leads us to the concept of summarization [Baroni *et al.*, 2014]. Summarization allows a complex part of an argumentation process, such as the analysis and discussion of factual evidence in a legal case, to be replaced by a more synthetic representation. This process focuses on the facts that have an actual impact on the decision, leaving out unnecessary details. The main concern with summarization is ensuring that, as the argumentation framework is simplified, the overall outcome remains consistent and preserved.

We use Example 4.9 to illustrate how summarization works.

Example 4.9 (Summarization [Baroni *et al.*, 2014]). Consider the argumentation frameworks AF_1 and AF_2 shown in Figure 25. AF_2 can be obtained from AF_1 by "summarizing" the component $\mathcal{M}_1$, including the arguments a_1, a_2, a_3 and a_4, with the component $\mathcal{M}_2$, including the arguments a_1 and a_2. Then, e_1 and e_2 are the same in the two frameworks, i.e. e_1 is labeled *in* and e_2 is labeled *out*. More generally, consider a finite sequence of n arguments $a_1, \ldots, a_n$ such that each argument attacks the subsequent one, i.e. a_i attacks a_{i+1} with $1 \leq i < n$, and suppose that only a_1 can receive further attacks from other arguments and that only a_n can attack other arguments. Then, it

is intuitive to see that the "black-box behavior" of a sequence of arguments of this kind, whose external "terminals" are a_1 and a_n, only depends on whether n is even or odd. In fact, the behavior of any even-length sequence is the same as where $n = 2$ (if a_1 is *in* then a_n is *out*, if a_1 is *out* then a_n is *in*, if a_1 is *undec* then a_n is *undec*), while for any odd-length sequence, the behavior is the same as for a_1 alone (if n is an odd number, a_n necessarily gets the same label as a_1).

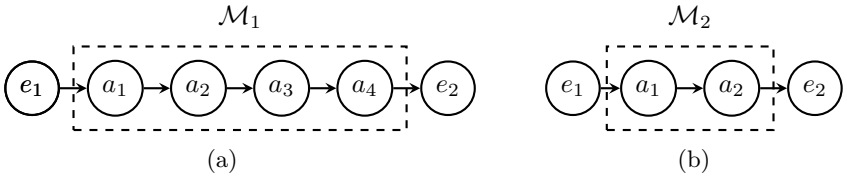


Figure 25: Summarizing a chain of arguments

The first area where locality and modularity principles find application is in the development of algorithms for efficiently computing argumentation semantics, particularly through divide-and-conquer strategies. By leveraging locality, one can focus on specific parts of an argumentation framework and thus reduce the computational burden. There are three locality- and modularity-based approaches that demonstrate how these principles can enhance the efficiency of computing semantics in dynamic, static, and partial argumentation frameworks. For instance, when only partial semantics are required — such as in scenarios where the status of certain arguments needs to be queried — algorithms can be designed to focus solely on the relevant arguments, disregarding those that do not impact the outcome. Similarly, in dialogues where new arguments are introduced, the computation can be streamlined by ignoring the effects of irrelevant arguments. For a comprehensive overview of these approaches, we refer to the work of Baroni *et al.* [2018].

Other principles used in the design of algorithms include robustness principles [Rienstra *et al.*, 2020], which deal with the behavior of a semantics when the argumentation framework changes due to the addition or removal of an attack between two arguments. Robustness principles have been classified into two kinds: persistence principles and monotonicity principles. The former deal with the question of whether a labeling in an argumentation framework under a given semantics persists after a change. The latter deal with the question of whether new labelings are created after a change. They are listed as follows:

XY addition persistence: a labeling of an argumentation framework in which x is labeled X and y is labeled Y is still a labeling of F after adding an attack from x to y.

XY removal persistence: a labeling of an argumentation framework in

which x is labeled X and y is labeled Y is still a labeling of F if removing the attack from x to y.

XY addition monotonicity: if in all labelings of an argumentation framework, x is labeled X and y is labeled Y, then adding an attack from x to y does not lead to new labelings.

XY removal monotonicity: if in all labelings of an argumentation framework, x is labeled X and y is labeled Y, then removing an attack from x to y does not lead to new labelings.

Persistence and monotonicity principles are also useful for addressing enforcement problems [Baumann, 2012] in abstract argumentation. This is about the problem of determining minimal sets of changes to an argumentation framework in order to enforce some result, such as the acceptance of a given set of arguments. Because persistence and monotonicity principles can be used to determine which changes to the attack relation of an argumentation framework do or do not change its evaluation, these principles can be used to guide the search for sets of changes that address the enforcement problem. This idea has already been used for extension enforcement under grounded semantics [Niskanen *et al.*, 2018].

In this section, we have discussed compositionality principles, algorithms and other computational approaches that exploit principles. We end this section with the following questions:

1. Most algorithms are developed for abstract argumentation and for argumentation as inference. What are the computational tasks for structured argumentation, for argumentation as dialogue, and for argumentation as balancing?

2. Apart from algorithms, what other tools does computer science have to offer, e.g., analysis of computational complexity, efficient implementation of algorithms?

3. As discussed in this section, principle-based analysis is a bridge between formal and algorithmic argumentation. Which principles are particularly useful? And how can we use these principles to speed up computation?

4. What else can we learn from artificial intelligence? The rise of machine learning and foundation models is changing the landscape of argumentation. How could these new approaches speed up computation?

5. Which topics need to be addressed first in computational argumentation? Should we address algorithms for formal argumentation concepts or focus our attention on the challenges of natural argumentation?

4.3 Explanation and argumentation

In this section, we discuss some relations between computational argumentation and explanation. Strategic argumentation explains argumentation as dialogue, discussion games explain argumentation as inference, and reason-based explanation can be used for argumentation as balancing. We illustrate the explanation for argumentation through the example of discussion games for grounded semantics.

In recent years, the field of explainable artificial intelligence (XAI) [Longo *et al.*, 2024] has gained significant attention due to the increasing complexity and opacity of AI systems, particularly when it comes to systems being potentially deployed in critical decision-making areas such as healthcare, finance, and the law. The main focus is usually on making the reasoning behind the decisions or the predictions made by the AI system [Phillips *et al.*, 2021] more understandable and transparent.

The relationship between explanation and argumentation can be seen from different perspectives. On the one hand, an explanation for argumentation mostly concerns the question of whether a certain argument or claim can be accepted (or not) and why. This has been studied not only at the abstract level [Ulbricht and Wallner, 2021] but also at the structured level [Borg and Bex, 2024]. On the other hand, explanation through argumentation is often intuitively seen as reasonable [Sklar and Azhar, 2018]. For example, it can clarify the decision-making process of an AI system through argumentation procedures. This can be done in a static manner by illustrating the argument inference process or showing the relations between arguments, or it can be done through an interactive dialogue that explains the reasoning [Castagna *et al.*, 2024b]. In this section, we talk about the following challenge:

Challenge 12. Explaining argumentation.

Take the example of argumentation as inference, which is about how reasonable decisions or conclusions can be reached by constructing for and against arguments and then evaluating those arguments. It makes it possible to understand decisions by tracing exactly why a particular conclusion was reached. It also makes it possible to see how certain decisions it relates to other potential conclusions. Explanations are often found to be argumentative. Mercier and Sperber [2017] highlighted that the effectiveness of interactive argumentation in changing people's minds, at least for simple arguments, stems from the chance to address counterarguments during discussions. Participants can raise and rebut counterarguments, which makes the exchange more dynamic. Contrast this with one-sided messaging campaigns, where counterarguments are generated but remain unaddressed [Altay *et al.*, 2022]. Interactive argumentation can involve a form of dialogue where users interact with an AI system, asking for clarifications or further information, and the system responds with explanations. In this sense, explanation is intertwined with dialogue — a conversation where arguments are presented, challenged, and defended, as in the Fatio design [McBurney and Parsons, 2004].

Strategic argumentation (see Governatori *et al.*'s [2021] overview) can be used to explain argumentation as dialogue. By analyzing the strategies employed by an agent, it is possible to understand how and why that agent chooses to disclose certain arguments or information during a debate in order to achieve a specific objective and prevent the opposing party from gaining an advantage.

To give an example, Arisaka *et al.* [2022] propose an abstract agent argumentation model that distinguishes the global argumentation of judges from the local argumentation of accused persons, prosecutors, defense lawyers, witnesses and experts. All the "local" agents have partial knowledge of the arguments and attacks of the other "local" agents, on which basis they decide autonomously whether to accept or reject their own arguments and whether to bring their own arguments forward in court. The arguments accepted by the judge are based on a game-theoretic equilibrium among the argumentation of the other agents. The theory can be used to distinguish between the various direct and indirect ways in which an agent's arguments can be used against his/her other arguments. The global abstract agent argumentation framework is viewed differently by the different agents.

Example 4.10 (Murder at Facility C). There was a murder at Facility C. Acc has been accused of the crime. There is a witness Wit and a prosecutor Prc. Acc has two arguments in mind:

a_1 He was at Facility A on the day of the murder [this is a fact Acc knows].

a_2 He is innocent [this is Acc's claim].

Prc entertains the following arguments:

a_6 Only Acc could have killed the victim [this is Prc's claim].

Meanwhile, Wit has certain beliefs on the basis of which he has three arguments:

a_3 Acc stayed home on the day of the murder, having previously lost his ID card [this Wit originally believes to be a fact].

a_4 Acc could enter any facility provided he had his ID card on him [this is a fact known to Wit].

a_5 Acc could not have been at Facility C at the time of the murder [this is Wit's claim].

Further, the relationship between the three arguments is such that a_3 attacks a_4, which attacks a_5. Altogether, these arguments by the three agents form the argumentation framework in Figure 26(a). Acc, Prc and Wit reveal their own internal argumentation, partially or elaborately, for the judge to evaluate. But since agents may come to learn the arguments of other agents if, say, they are expressed before they present their own arguments, it is possible that they take the additional information into account when deciding which arguments

to present. In this example, both Prc and Acc have the characteristic of being unaware agents. Prc has no reason to drop argument a_6, and neither does Acc, as he sees no benefit in admitting that a_6. However, how Wit responds to the fact known to Acc (a_1) could prove crucial to whether he is found innocent or guilty.

Case A. Suppose Wit is unaware and open-minded. Wit presents what he believes, i.e., his local argumentation framework (see Figure 26(a).) She locally accepts a_3 and a_5. The judge evaluates all the arguments, concluding that a_2 is not acceptable, i.e., Acc is guilty. The judge starts his inference with Acc's acceptable a_1 and proceeds to reject a_2. The two arguments a_3 and a_5 accepted by Wit are not accepted by the judge. This illustrates indirect use of an argument against Acc.

Case B. Suppose Wit is unaware and closed-minded. Instead of presenting all the reasoning he had developed in his local argumentation, Wit states the following key points concisely: that Acc stayed home on the day of the murder, and that Acc could not have been at Facility C (see $\mathcal{F}_i$). Omission of a fact known to Wit (a_4), which Wit perhaps considers irrelevant to the criminal case, changes the judge's decision completely. In $\mathcal{F}_i$, a_5 is seen an argument that is globally acceptable. That argument rejects a_6 in favor of a_2.

Case C. Suppose Wit learns a_1 beforehand. Wit realizes that a_3, which she thought was a fact, is not actually true. She no longer claims a_5 in her local argumentation, but she nevertheless discloses her entire original argumentation (see Figure 26(a)). Her conclusion that a_4 is acceptable concurs with the judge's view on the matter, and the judge ultimately concludes that a_2 shall be rejected.

Case D. Suppose again that Wit learns a_1 beforehand but that she mentions the key arguments concisely. She states that entry into any facility requires an ID card (see $\mathcal{F}_j$). Here again, the judge has no objection to the evidence that might have been provided by Proc. As such, a_6 is accepted, which proves that Acc is guilty. This illustrates direct use of an argument by Wit against Acc.

We now move on to discussion games designed to explain argumentation as inference. As discussed in Section 4.2, calculating semantics, or determining whether a specific argument is present in some or all labelings, can be computationally expensive. Discussion games provide an alternative to that. Discussion games [Caminada, 2017] take place between two parties, typically called the "proponent" and the "opponent", who argue about whether a particular argument within a formal argumentation framework should be accepted. Discussion games can be seen as proof procedures for the argumentation semantics they are associated with, e.g., grounded semantics, preferred semantics, or stable

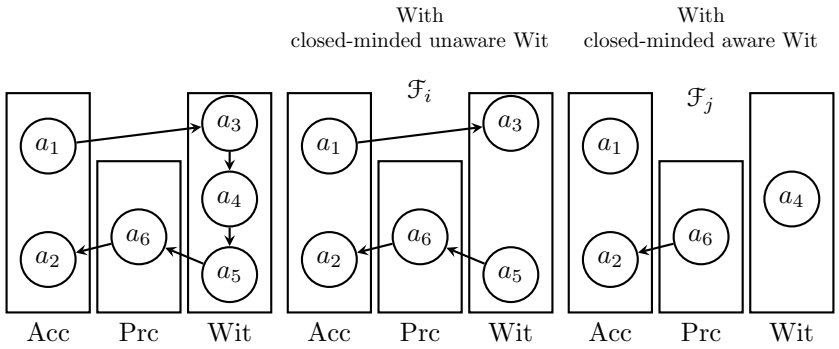


Figure 26: Left: Argumentation by an accused (Acc), by a witness (Wit), and by a prosecutor (Prc). Middle: multiagent semantics with open-minded unaware Wit ($\mathcal{F}_i$). Right: multiagent semantics with closed-minded aware Wit ($\mathcal{F}_j$).

semantics. These games provide a local explanation, focussing on the admissibility of arguments.

A discussion game for grounded semantics is won by one agent iff a particular argument is labeled *in*. There are two players: Proponent (P) and Opponent (O). There are four moves with respect to arguments A and B:

P: HTB(A): The labeling of A is *in*.

O: CB(A): Maybe the labeling of A is *out* in every complete labeling.

O: CONCEDE(A): Agree that the labeling of A is *in* in every complete labeling.

O: RETRACT(A) The labeling of A is *out* in every complete labeling.

The following are the discussion rules on grounded semantics.
P: **HTB(A):** Either this is the first move, or:

- the previous move was **CB(B)**, where A attacked B, and:
- no **CONCEDE** or **RETRACT** move is applicable.

O: **CB(A):** A is an attacker of the last **HTB(B)** statement, which has not yet been conceded, and:

- the directly preceding move was not a **CB** statement,
- argument A has not yet been **RETRACT**ed, and
- no **CONCEDE** or **RETRACT** move is applicable.

O: **CONCEDE(A):** There has been a **HTB(A)** statement in the past, and

- every argument attacking **HTB(A)** has been **RETRACT**ed, and
- **CONCEDE(A)** has not yet been moved.

O: **RETRACT(A):** There has been a **CB(A)** statement in the past, and:

- there exists an argument attacking **CB(A)** that has been **CONCEDE**ed, and
- **RETRACT(A)** has not yet been moved.

General rule: No "**HTB or CB repeats**" are allowed. **HTB(A)** is only allowed once, **CB(A)** is only allowed once. For any A, not both of**HTB(A)** and **CB(A)** are allowed.

We use Example 4.11 to illustrate how grounded discussion games work.

Example 4.11 (Grounded discussion game). Given the abstract argumentation framework visualized in Figure 27, the grounded discussion game for argument C proceeds as follows:

(1) $P : HTB(c)$	(3) $P : HTB(a)$	(5) $O : RETRACT(b)$
(2) $O : CB(b)$	(4) $O : CONCEDE(a)$	(6) $O : CONCEDE(c)$

P wins the grounded discussion game for argument c, and c is labeled *in* in the grounded labeling.

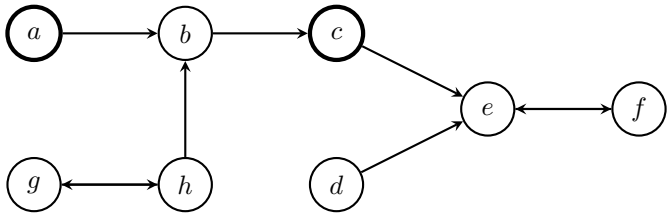


Figure 27: An abstract argumentation framework where c is being discussed

Yet, despite its potential advantages, the use of argumentation as balancing for explanatory purposes has not been adequately explored. This method employs reason-based [Knoks and van der Torre, 2023] and scale-based balancing as a decision model. Thus, the explanation could provide an overview of the pros and cons concerning a decision, like judges do when explaining their rulings.

Explanations are available to not just the experts who designed the system (which is then called a white box [Vilone and Longo, 2021]), but also to lay people — non-experts who may not understand all the intricacies of certain models. This issue is about how to personalize explanations, which is particularly relevant given the diversity of backgrounds, contexts, mental states, emotions, and

abilities of subjects receiving explanations generated by AI systems (humans such as patients and healthcare professionals, or virtual intelligent autonomous agents). To this end, new forms of knowledge representation should be envisioned and synergistically integrated to enable argumentation reasoning over that.

In this section, we discussed argumentation and explanation. We end this section with the following open questions:

1. Argumentation and explanation can be related to each other in various ways. What is the role of argumentation in explanation, and what is the role of explanation in argumentation? For example, how can the Fatio dialogue protocol be extended with explanation dialogues?

2. Dialogue is often cited as a distinctive feature of argumentation that can be used for interactive explanation in human-computer interactions, but other features may also be relevant. How about, for example, using balancing or inference in explanation?

3. On the topic of dialogue, it is often observed that argumentation is persuasion and that there are many other kinds of dialogues types. How can different dialogue types be integrated into dialogue systems? For example, how can information seeking be included in argumentation?

4. Since explanations are often personal, expressed in natural language, and use common sense reasoning, foundation models and chatbots have been promoted as part of the explanation toolbox. How can we use LLMs in argumentation to incorporate context, mental states and emotions?

5. Explanation techniques can be evaluated in terms of the degree to which they improve a system's goals, i.e., how does the combination of argumentation and explanation techniques improve system behavior?

4.4 Argumentation technology

We conclude our discussion of computational argumentation by discussing the integration of argumentation techniques with existing and emerging technologies in computer science like NLP, LLMs, distributed argumentation technology, and dialogue technology. We illustrate the integration of these technologies by using as an example the integration of argumentation with blockchain technology into the architecture of the IHiBO.

Recent years have seen remarkable advancements in deep learning, particularly with the development and deployment of LLMs. This presents a significant opportunity to integrate argumentation. Argumentation is inherently suitable for enhancing the reasoning and conversational capabilities of LLMs [Bezou-Vrakatseli, 2023; Castagna *et al.*, 2024c]. Argumentation provides a formal mechanism for capturing interactions between agents, and it manages the information conflicts that arise during these interactions. This makes it an potent-

cial tool for improving the logical consistency and depth of responses generated by LLMs.

Additionally, LLMs prompts to reevaluate the relationship between abstract and structured argumentation. Traditionally, structured arguments were necessary because the attack relations were defined based on the internal structure of the arguments, as discussed in Section 3.2. However, with LLM capabilities, it becomes possible to retain arguments in their natural language form and use an LLM to extract the underlying argumentation framework. This approach allows argumentation to be integrated more naturally with conversational models because LLMs provide the contextual understanding needed to facilitate these processes.

Furthermore, continuous improvements in computational power, together with the capabilities of foundational models like LLMs, have opened up new avenues for integrating argumentation into more complex systems. In such systems, argumentation can synergize with other technologies, enhancing the overall functionality and enabling more sophisticated applications. Such integration will not only advance the field of computational argumentation but will also push the boundaries of what can be achieved in AI-driven reasoning and decision-making systems. In this section, we discuss the following challenge:

Challenge 13. Integrating argumentation with technologies.

Distributed argumentation technology [Yu, 2023] is a computational approach that incorporates argumentation reasoning mechanisms within multiagent systems. An instantiation of distributed argumentation technology is *Intelligent Human-Input-Based Blockchain Oracle (IHiBO)* [Yu *et al.*, 2022]. The motivation for IHiBO comes from fund management for the securities market. Figure 28 shows a toy fund management procedure. Investors first pool their money together and then fund managers conduct investment research and prepare the specific plan for the investment portfolio. Fund managers invest securities on behalf of their clients (investors) according to their research and the final decision in the investment plan. The investment generates returns, and the returns are passed down to investors. Fund managers play an important role in the investment and financial world as they give investors peace of mind that their money is in the hands of experts. However, reality is not always as hoped for and investors are supposed to know (but do not actually know) where their money is going, why, and how much is the true profit.

A significant aspect of IHiBO is its human-input-based oracle, which bridges the gap between a blockchain and real-world data, allowing human experts to input information into the blockchain. IHiBO was envisioned for use in fund management, where managers provide their arguments in terms of the investment plan for stocks. Specifically, IHiBO utilizes multiagent abstract argumentation frameworks to model decision-making processes, which are then implemented by smart contracts and stored on a blockchain. The integration of argumentation reasoning and blockchain makes the decision-making process more transparent and traceable.

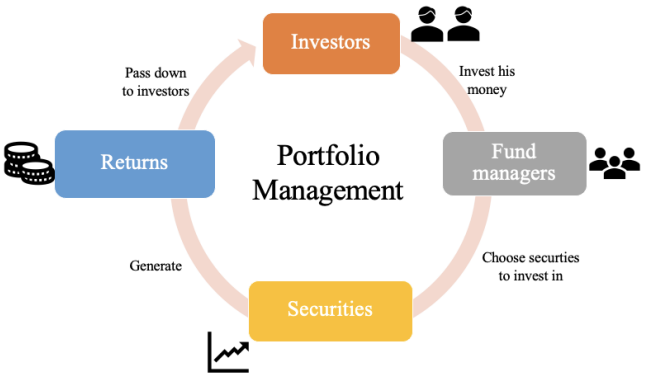


Figure 28: A fund investment process

IHiBO leverages a two-layer distributed ledger technology (DLT), shown in Figure 29, to ensure security and immutability of data while maintaining efficiency and scalability. The secondary layer, a private permissioned blockchain, accessible only to authorized users, facilitates the decision-making smart contract. This layer maintains privacy and reduces transaction costs, providing a balanced approach to data security and operational efficiency. The main layer is a public permissionless blockchain such as Ethereum, where the smart contract for executing stock transactions is invoked by the output of the decision-making process. This two-layered design is particularly important in fund management, where decisions may involve sensitive information. IHiBO's architecture supports not only multiagent abstract argumentation but also other kinds of reasoning that can be encoded in smart contracts.

In this section, we have discussed he integration of argumentation techniques with existing and emerging technologies, such as IHiBO. We end this section with the following questions.

1. What can technologies do for argumentation, and what can argumentation do for technologies? For example, what can foundational models do for (natural, formal, and computation) argumentation? How can LLMs be used to develop technologies like IHiBO 2.0? Another discussed example outlines the potential roles of computational models of legal argumentation [Prakken, 2024b]: as tools for guiding prompt engineering, as benchmarks for evaluating the outputs of legal generative AI, and as symbolic alternatives to legal generative AI, that could be integrated as conversational interfaces.

2. As we emphasized in this chapter, conceptualizations of argumentation can take the form of inference, dialogue or balancing, and these models have their own formal methods. Do they also have their own technologies?

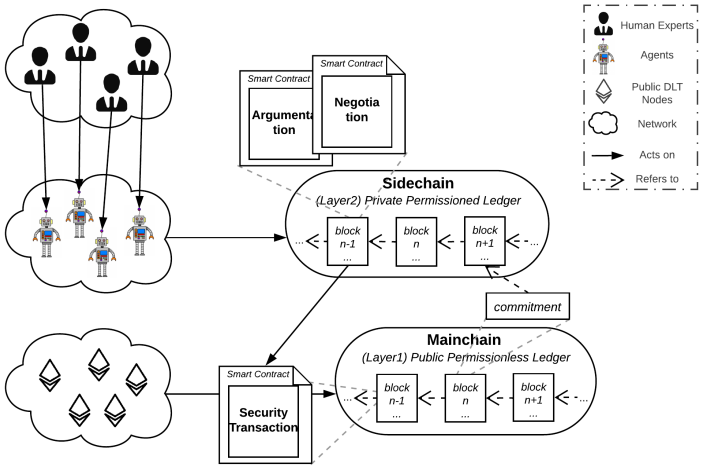


Figure 29: The architecture of the IHiBO framework. DLT = distributed ledger technology.

3. In the attack-defense paradigm shift, commutative diagrams play a central role. How can these technologies be integrated to make structured argumentation diagrams commute? How should various technologies be integrated with structured and abstract argumentation?

4. Another recurring discussion in this chapter is methodology, e.g., how can we develop a user guide about the new formal methods? Likewise, we may ask: how can we use new argumentation technologies like IHiBO?

5. This is just the beginning of the use of the attack-defense paradigm shift in argumentation technology. For a start, how can we use algorithmic argumentation methods in argumentation technologies?

5 Summary

This chapter has discussed the evolving landscape of argumentation, exploring its natural forms, the paradigm shift initiated by Dung, and subsequent advancements in computational approaches.

Natural argumentation is rooted in both theoretical and practical reasoning, with formal argumentation grounded in philosophical and mathematical foundations. This foundational approach is essential for representing, managing, and resolving conflicts in various disciplines. For instance, the Jiminy ethical governor, which operates across six layers of conflict, exemplifies the complexity and depth of formal reasoning in ethical decision-making. However, natural argumentation is inherently diverse, reflecting the complexities of human thought and communication. This diversity is evident in psychological evaluations of arguments, where understanding and generating arguments involves intricate

cognitive processes. Foundation models are increasingly employed to construct and decode arguments, highlighting the importance of questions, particularly "why" questions, in uncovering weaknesses and justifying decisions. These questions play a crucial role in frameworks like the Fatio dialogue protocol, emphasizing the centrality of inquiry in argumentation. Argumentation can be conceptualized in various ways such as inference, dialogue, or balancing, each offering distinct perspectives and applications. For example, a divorce court case can be modeled differently depending on the chosen conceptual framework, which demonstrates the flexibility of argumentation theories. Additionally, formalizing argumentation through a variety of methods allows for a combination of different reasoning techniques, for instance, in the representation of practical scenarios like a mother reasoning with her daughter with mixed formal methods. This highlights the practical utility of formal argumentation for navigating complex real-world situations.

The paradigm shift in argumentation was significantly influenced by the attack-defense framework introduced by Dung [1995]. This framework marked a turning point in argumentation theory by emphasizing that every utterance, whether a claim, argument, or attack, can be contested. This led to a more comprehensive and universal approach to analyzing arguments. Structured argumentation has served as a bridge between classical and nonmonotonic logic, representing various logic and game theory concepts. This is particularly evident in the design of nonmonotonic logic, where rationality postulates from paraconsistent reasoning are crucial. Examples like the weakest versus last link principles and PDL illustrate the depth and versatility of formal argumentation in designing and representing these logics. Furthermore, extensions to abstract argumentation frameworks have been developed so that we can extract more information and incorporate qualitative and quantitative elements such as bipolarity, preferences, and so on. These extensions, inspired by dialogue and balancing, have enriched Dung's abstract argumentation frameworks and allow for a more nuanced understanding and modeling of complex argumentative situations.

Computational argumentation has advanced significantly with the development of the principle-based approach, which handles the diversity of argumentation models by providing a higher level of abstraction. This approach is essential for selecting appropriate methods and designing new ones. It ensures that the diverse landscape of argumentation models can be navigated and applied effectively. Compositionality principles such as locality, modularity and transparency are central to the attack-defense perspective in computational argumentation. These principles are exploited by algorithms and computational techniques to enhance their efficiency, robustness, and scalability, as seen in the divide and conquer approach which breaks down complex frameworks into manageable components. The relationship between explanation and argumentation was also discussed. Strategic argumentation explains dialogues, discussion games clarify inference, and underdeveloped reason-based explana-

tions are used for balancing. These connections underscore the importance of argumentation for making AI systems more transparent and understandable. Additionally, the integration of argumentation techniques with emerging technologies such as distributed argumentation technology has expanded the potential applications of argumentation in areas like blockchain and AI. For instance, the IHiBO architecture integrates argumentation with blockchain technology to enhance transparency and trust in decision-making processes.

In summary, this chapter presents an overview of argumentation: past achievements, the current state of the art, and future directions. We discussed the three pillars in the context of natural argumentation before discussing the attack-defense paradigm shift initiated by Dung and advancements in computational argumentation that are shaping the future of the field.

6 Acknowledgments

All authors acknowledge financial support from the Luxembourg National Research Fund (FNR) — L. van der Torre through the project The Epistemology of AI Systems (EAI) (C22/SC/17111440), L. van der Torre and R. Markovich through the projects Logical Methods for Deontic Explanations (LODEX) (INTER/DFG/23/17415164/LoDEx) and Study of the Limits, Problems and Risks Associated with Autonomous Technologies (INTEGRAUTO) (INTER/AUDACE/21/16695098) and all authors through the project Deontic Logic for Epistemic Rights (DELIGHT) (O20/14776480). R. Markovich and L.Yu are also supported by the University of Luxembourg's Marie Speyer Excellence Grant of 2024 Formal Analysis of Discretionary Reasoning (MSE-DISCREASON).

BIBLIOGRAPHY

[Alchourrón *et al.*, 1985] Carlos E Alchourrón, Peter Gärdenfors, and David Makinson. On the logic of theory change: partial meet contraction and revision functions. *The journal of symbolic logic*, 50(2):510–530, 1985.

[Aleinikoff, 1986] T Alexander Aleinikoff. Constitutional law in the age of balancing. *Yale lj*, 96:943, 1986.

[Alkaissi and McFarlane, 2023] Hussam Alkaissi and Samy I McFarlane. Artificial hallucinations in chatgpt: implications in scientific writing. *Cureus*, 15(2), 2023.

[Altay *et al.*, 2022] Sacha Altay, Marlène Schwartz, Anne-Sophie Hacquin, Aurélien Allard, Stefaan Blancke, and Hugo Mercier. Scaling up interactive argumentation by providing counterarguments with a chatbot. *Nature Human Behaviour*, 6(4):579–592, 2022.

[Amgoud and Cayrol, 2002] Leila Amgoud and Claudette Cayrol. Inferring from inconsistency in preference-based argumentation frameworks. *International Journal of Approximate Reasoning*, 29(2):125–169, 2002.

[Amgoud and Vesic, 2012] Leila Amgoud and Srdjan Vesic. On the use of argumentation for multiple criteria decision making. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 480–489. Springer, 2012.

[Amgoud, 2005] Leila Amgoud. An argumentation-based model for reasoning about coalition structures. In *International Workshop on Argumentation in Multi-Agent Systems*, pages 217–228. Springer, 2005.

[Amgoud, 2009] Leila Amgoud. Argumentation for decision making. *Argumentation in artificial intelligence*, pages 301–320, 2009.

[Arisaka *et al.*, 2022] Ryuta Arisaka, Jérémie Dauphin, Ken Satoh, and Leendert van der Torre. Multi-agent argumentation and dialogue. *IfCoLog Journal of Logics and Their Applications*, 9(4):921–954, 2022.

[Atkinson and Bench-Capon, 2005] Katie Atkinson and Trevor Bench-Capon. Legal case-based reasoning as practical reasoning. *Artificial Intelligence and Law*, 13:93–131, 2005.

[Atkinson and Bench-Capon, 2021] Katie Atkinson and Trevor Bench-Capon. Value-based argumentation. *Handbook of Formal Argumentation*, 2:397–441, 2021.

[Aumann, 2016] Robert J Aumann. *Agreeing to disagree*. Springer, 2016.

[Austin, 1975] John Langshaw Austin. *How to do things with words*. Harvard University Press, 1975.

[Awad *et al.*, 2018] Edmond Awad, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon, and Iyad Rahwan. The moral machine experiment. *Nature*, 563(7729):59–64, 2018.

[Baroni and Giacomin, 2007] Pietro Baroni and Massimiliano Giacomin. On principle-based evaluation of extension-based argumentation semantics. *Artificial Intelligence*, 171(10-15):675–700, 2007.

[Baroni *et al.*, 2005] Pietro Baroni, Massimiliano Giacomin, and Giovanni Guida. Scc-recursiveness: a general schema for argumentation semantics. *Artificial Intelligence*, 168(1-2):162–210, 2005.

[Baroni *et al.*, 2011] Pietro Baroni, Paul E Dunne, and Massimiliano Giacomin. On the resolution-based family of abstract argumentation semantics and its grounded instance. *Artificial Intelligence*, 175(3-4):791–813, 2011.

[Baroni *et al.*, 2014] Pietro Baroni, Guido Boella, Federico Cerutti, Massimiliano Giacomin, Leendert van der Torre, and Serena Villata. On the input/output behavior of argumentation frameworks. *Artificial Intelligence*, 217:144–197, 2014.

[Baroni *et al.*, 2018] Pietro Baroni, Massimiliano Giacomin, and Beishui Liao. Locality and modularity in abstract argumentation. *Handbook of formal argumentation*, pages 937–979, 2018.

[Barringer *et al.*, 2005] Howard Barringer, Dov Gabbay, and John Woods. Temporal dynamics of support and attack networks: From argumentation to zoology. *Mechanizing Mathematical Reasoning: Essays in Honor of Jörg H. Siekmann on the Occasion of His 60th Birthday*, pages 59–98, 2005.

[Baumann *et al.*, 2020] Ringo Baumann, Gerhard Brewka, and Markus Ulbricht. Revisiting the foundations of abstract argumentation–semantics based on weak admissibility and weak defense. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 2742–2749, 2020.

[Baumann, 2012] Ringo Baumann. What does it take to enforce an argument? Minimal change in abstract argumentation. In *ECAI 2012*, pages 127–132. IOS Press, 2012.

[Baumeister *et al.*, 2021] Dorothea Baumeister, Daniel Neugebauer, and Jörg Rothe. Collective acceptability in abstract argumentation. *Handbook of Formal Argumentation, Volume 2*, 2021.

[Bengel *et al.*, 2024] Lars Bengel, Lydia Blümel, Tjitze Rienstra, and Matthias Thimm. Argumentation-based probabilistic causal reasoning. In *Conference on Advances in Robust Argumentation Machines*, pages 221–236. Springer, 2024.

[Bezou-Vrakatseli, 2023] Elfia Bezou-Vrakatseli. Evaluation of LLM reasoning via argument schemes. *Online Handbook of Argumentation for AI*, 4, 2023.

[Bistarelli *et al.*, 2021] Stefano Bistarelli, Francesco Santini, et al. Weighted argumentation. *Handbook of Formal Argumentation, Volume 2*, 2021.

[Black *et al.*, 2021] Elizabeth Black, Nicolas Maudet, and Simon Parsons. Argumentation-based dialogue. *Handbook of Formal Argumentation, Volume 2*, 2021.

[Boella *et al.*, 2009] Guido Boella, Dov M Gabbay, Leendert van der Torre, and Serena Villata. Meta-argumentation modelling I: Methodology and techniques. *Studia Logica*, 93:297–355, 2009.

[Boella *et al.*, 2010] Guido Boella, Dov M Gabbay, Leendert van der Torre, and Serena Villata. Support in abstract argumentation. In *Proceedings of the Third International Conference on Computational Models of Argument (COMMA'10)*, pages 40–51. Frontiers in Artificial Intelligence and Applications, IOS Press, 2010.

[Borg and Bex, 2024] AnneMarie Borg and Floris Bex. Minimality, necessity and sufficiency for argumentation and explanation. *International Journal of Approximate Reasoning*, page 109143, 2024.

[Brewka and Eiter, 1999] G. Brewka and T. Eiter. Preferred answer sets for extented logic programs. *Artif. Intell.*, 109:297–356, 1999.

[Brewka, 1994] Gerhard Brewka. Reasoning about priorities in default logic. In Barbara Hayes-Roth and Richard E. Korf, editors, *Proc. of the 12th National Conference on AI*, volume 2, pages 940–945. AAAI Press / The MIT Press, 1994.

[Budzynska *et al.*, 2018] Katarzyna Budzynska, Serena Villata, et al. Processing natural language argumentation. *Handbook of formal argumentation*, 1:577–627, 2018.

[Caminada and Amgoud, 2005] Martin Caminada and Leila Amgoud. An axiomatic account of formal argumentation. In *AAAI*, volume 6, pages 608–613, 2005.

[Caminada and Amgoud, 2007] Martin Caminada and Leila Amgoud. On the evaluation of argumentation formalisms. *Artificial Intelligence*, 171(5-6):286–310, 2007.

[Caminada and Ben-Naim, 2007] Martin Caminada and Jonathan Ben-Naim. *Postulates for paraconsistent reasoning and fault tolerant logic programming.* PhD thesis, Department of Information and Computing Sciences, Utrecht University, 2007.

[Caminada and Gabbay, 2009] Martin WA Caminada and Dov M Gabbay. A logical account of formal argumentation. *Studia Logica*, 93:109–145, 2009.

[Caminada and Pigozzi, 2011] Martin Caminada and Gabriella Pigozzi. On judgment aggregation in abstract argumentation. *Autonomous Agents and Multi-Agent Systems*, 22:64–102, 2011.

[Caminada and Wu, 2011] Martin Caminada and Yining Wu. On the limitations of abstract argumentation. In *Proceedings of the 23rd Benelux Conference on Artificial Intelligence (BNAIC 2011)*, pages 59–66, 2011.

[Caminada *et al.*, 2012] Martin WA Caminada, Walter A Carnielli, and Paul E Dunne. Semi-stable semantics. *Journal of Logic and Computation*, 22(5):1207–1254, 2012.

[Caminada *et al.*, 2014] Martinus Wigbertus Antonius Caminada, Sanjay Modgil, and Nir Oren. Preferences and unrestricted rebut. *Computational Models of Argument*, 2014.

[Caminada, 2006] Martin Caminada. Semi-stable semantics. *COMMA*, 144:121–130, 2006.

[Caminada, 2017] Martin Caminada. Argumentation semantics as formal discussion. *Journal of Applied Logics*, 4(8):2457–2492, 2017.

[Caminada, 2018a] Martin Caminada. Argumentation semantics as formal discussion. *Handbook of Formal Argumentation*, 1:487–518, 2018.

[Caminada, 2018b] Martin Caminada. Rationality postulates: Applying argumentation theory for non-monotonic reasoning. *Handbook of Formal Argumentation, Volume 1*, pages 771–796, 2018.

[Castagna *et al.*, 2024a] Federico Castagna, Nadin Kökciyan, Isabel Sassoon, Simon Parsons, and Elizabeth Sklar. Computational argumentation-based chatbots: a survey. *Journal of Artificial Intelligence Research*, 80:1271–1310, 2024.

[Castagna *et al.*, 2024b] Federico Castagna, Peter McBurney, and Simon Parsons. Explanation–question–response dialogue: An argumentative tool for explainable AI. *Argument & Computation*, (Preprint):1–23, 2024.

[Castagna *et al.*, 2024c] Federico Castagna, Isabel Sassoon, and Simon Parsons. Can formal argumentative reasoning enhance LLMs performances? *arXiv preprint arXiv:2405.13036*, 2024.

[Cayrol and Lagasquie-Schiex, 2005] 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.

[Cayrol and Lagasquie-Schiex, 2009] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Bipolar abstract argumentation systems. In *Argumentation in Artificial Intelligence*, pages 65–84. Springer, 2009.

[Cayrol and Lagasquie-Schiex, 2010] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Coalitions of arguments: A tool for handling bipolar argumentation frameworks. *International Journal of Intelligent Systems*, 25(1):83–109, 2010.

[Cayrol and Lagasquie-Schiex, 2013] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Bipolarity in argumentation graphs: Towards a better understanding. *International Journal of Approximate Reasoning*, 54(7):876–899, 2013.

[Cayrol *et al.*, 2021] Claudette Cayrol, Andrea Cohen, and Marie Christine Lagasquie Schiex. Higher-order interactions (bipolar or not) in abstract argumentation: a state of the art. 2021.

[Cerutti *et al.*, 2021] Federico Cerutti, Marcos Cramer, Mathieu Guillaume, Emmanuel Hadoux, Anthony Hunter, and Sylwia Polberg. Empirical cognitive studies about formal argumentation. *Handbook of Formal Argumentation, Volume 2*, 2021.

[Coleman, 1984] James S Coleman. Micro foundations and macrosocial behavior. *Angewandte Sozialforschung anc AIAS Informationen Wien*, 12(1-2):25–37, 1984.

[Cramer and Guillaume, 2018a] Marcos Cramer and Mathieu Guillaume. Directionality of attacks in natural language argumentation. In *CEUR Workshop Proceedings*. RWTH Aachen University, Aachen, Germany, 2018.

[Cramer and Guillaume, 2018b] Marcos Cramer and Mathieu Guillaume. Empirical cognitive study on abstract argumentation semantics. In *Computational Models of Argument*, pages 413–424. IOS Press, 2018.

[Cramer and Guillaume, 2019] Marcos Cramer and Mathieu Guillaume. Empirical study on human evaluation of complex argumentation frameworks. In *Logics in Artificial Intelligence: 16th European Conference, JELIA 2019, Rende, Italy, May 7–11, 2019, Proceedings 16*, pages 102–115. Springer, 2019.

[Da Costa *et al.*, 2007] Newton CA Da Costa, Décio Krause, and Otávio Bueno. Paraconsistent logics and paraconsistency. In *Philosophy of logic*, pages 791–911. Elsevier, 2007.

[Da Costa, 1974] Newton CA Da Costa. On the theory of inconsistent formal systems. *Notre dame journal of formal logic*, 15(4):497–510, 1974.

[Danry *et al.*, 2023] Valdemar Danry, Pat Pataranutaporn, Yaoli Mao, and Pattie Maes. Don't just tell me, ask me: AI systems that intelligently frame explanations as questions improve human logical discernment accuracy over causal AI explanations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2023.

[Dauphin and Cramer, 2018] Jérémie Dauphin and Marcos Cramer. ASPIC-END: structured argumentation with explanations and natural deduction. In *Theory and Applications of Formal Argumentation: 4th International Workshop, TAFA 2017, Melbourne, VIC, Australia, August 19-20, 2017, Revised Selected Papers 4*, pages 51–66. Springer, 2018.

[D'Avila Garcez *et al.*, 2005] Artur S D'Avila Garcez, Dov M Gabbay, and Luis C Lamb. Value-based argumentation frameworks as neural-symbolic learning systems. *Journal of Logic and Computation*, 15(6):1041–1058, 2005.

[Delgrande *et al.*, 2004] James Delgrande, Torsten Schaub, Hans Tompits, and Kewen Wang. A classification and survey of preference handling approaches in nonmonotonic reasoning. *Computational Intelligence*, 20(2):308–334, 2004.

[Dong *et al.*, 2019] Huimin Dong, Beishui Liao, Reka Markovich, and Leendert van der Torre. From classical to non-monotonic deontic logic using ASPIC+. In *Logic, Rationality, and Interaction: 7th International Workshop, LORI 2019, Chongqing, China, October 18–21, 2019, Proceedings 7*, pages 71–85. Springer, 2019.

[Dong *et al.*, 2021] Huimin Dong, Réka Markovich, and Leendert van der Torre. Towards AI logic for social reasoning. *arXiv preprint arXiv:2110.04452*, 2021.

[Dung and Thang, 2018] Phan Minh Dung and Phan Minh Thang. Fundamental properties of attack relations in structured argumentation with priorities. *Artificial Intelligence*, 255:1–42, 2018.

[Dung, 1995] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial intelligence*, 77(2):321–357, 1995.

[Dung, 2014] Phan Minh Dung. An axiomatic analysis of structured argumentation for prioritized default reasoning. volume 263 of *Frontiers in Artificial Intelligence and Applications*, pages 267–272. IOS Press, 2014.

[Dung, 2016] Phan Minh Dung. An axiomatic analysis of structured argumentation with priorities. *Artif. Intell.*, 231:107–150, 2016.

[FIPA, 2002] FIPA. Communicative act library specification. *http://www. fipa. org/specs/-fipa00037*, 2002.

[Gabbay and Rivlin, 2017] Dov M Gabbay and Lydia Rivlin. Heal2100: human effective argumentation and logic for the 21st century. The next step in the evolution of logic. *IFCoLog Journal of Logics and Their Applications*, 2017.

[Garcez *et al.*, 2008] Artur SD'Avila Garcez, Luis C Lamb, and Dov M Gabbay. *Neural-symbolic cognitive reasoning*. Springer Science & Business Media, 2008.

[Gargouri *et al.*, 2020] Anis Gargouri, Sébastien Konieczny, Pierre Marquis, and Srdjan Vesic. On a notion of monotonic support for bipolar argumentation frameworks. In *20th International Conference on Autonomous Agents and MultiAgent Systems*, 2020.

[Georgeff *et al.*, 1999] Michael Georgeff, Barney Pell, Martha Pollack, Milind Tambe, and Michael Wooldridge. The belief-desire-intention model of agency. In *Intelligent Agents V: Agents Theories, Architectures, and Languages: 5th International Workshop, ATAL'98 Paris, France, July 4–7, 1998 Proceedings 5*, pages 1–10. Springer, 1999.

[Giacomin, 2017] Massimiliano Giacomin. Handling heterogeneous disagreements through abstract argumentation (extended abstract). In *PRIMA 2017: Principles and Practice of Multi-Agent Systems*, pages 3–11, 2017.

[Gillioz and Zufferey, 2020] Christelle Gillioz and Sandrine Zufferey. *Introduction to experimental linguistics*. John Wiley & Sons, 2020.

[Giunchiglia *et al.*, 2004] Enrico Giunchiglia, Joohyung Lee, Vladimir Lifschitz, Norman McCain, and Hudson Turner. Nonmonotonic causal theories. *Artificial Intelligence*, 153(1-2):49–104, 2004.

[Gordon *et al.*, 2007] Thomas F Gordon, Henry Prakken, and Douglas Walton. The Carneades model of argument and burden of proof. *Artificial Intelligence*, 171(10-15):875–896, 2007.

[Gordon, 1993] Thomas F Gordon. The Pleadings Game: An exercise in computational dialectics. *Artificial Intelligence and Law*, 2:239–292, 1993.

[Gordon, 2018] Thomas F Gordon. Towards requirements analysis for formal argumentation. In Pietro Baroni, Dov Gabbay, Massimiliano Giacomin, and Leendert van der Torre, editors, *Handbook of formal argumentation, Volume 1*, pages 145–156. College Publictions, 2018.

[Governatori *et al.*, 2021] Guido Governatori, Michael J Maher, and Francesco Olivieri. Strategic argumentation. *Handbook of Formal Argumentation, Volume 2*, 2021.

[Hakim *et al.*, 2019] Fauzia Zahira Munirul Hakim, Lia Maulia Indrayani, and Rosaria Mita Amalia. A dialogic analysis of compliment strategies employed by Replika chatbot. In *Third International conference of arts, language and culture (ICALC 2018)*, pages 266–271. Atlantis Press, 2019.

[Hamblin, 1970] C. L. Hamblin. Fallacies. *Tijdschrift Voor Filosofie*, 33(1):183–188, 1970.

[Heyninck, 2019] Jesse Heyninck. *Investigations into the logical foundations of defeasible reasoning: an argumentative perspective*. PhD thesis, Ruhr University Bochum, Germany, 2019.

[Hinton and Wagemans, 2023] Martin Hinton and Jean HM Wagemans. How persuasive is AI-generated argumentation? An analysis of the quality of an argumentative text produced by the GPT-3 AI text generator. *Argument & Computation*, 14(1):59–74, 2023.

[Kaci *et al.*, 2021] Souhila Kaci, Leendert van Der Torre, Srdjan Vesic, and Serena Villata. Preference in abstract argumentation. *Handbook of Formal Argumentation, Volume 2*, 2021.

[Kakas and Moraitis, 2003] Antonis Kakas and Pavlos Moraitis. Argumentation based decision making for autonomous agents. In *Proceedings of the second international joint conference on Autonomous agents and multiagent systems*, pages 883–890, 2003.

[Kant, 1993] Immanuel Kant. *Groundwork of the Metaphysics of Morals*. Hackett Publishing Company, Indianapolis, 3rd edition, 1993. [1785].

[Kissine, 2013] Michail Kissine. Speech act classifications. *Pragmatics of speech actions*, 173:202, 2013.

[Knocks *et al.*, 2024] Aleks Knocks, Muyun Shao, Leendert ver der Torre, Vincent De Wit, and Liuwen Yu. A principle-based analysis for numerical balancing. In *Logics for New-Generation Artificial Intelligence (LNGAI2024)*. College Publications, United Kingdom, 2024.

[Knoks and van der Torre, 2023] Aleks Knoks and Leendert van der Torre. Reason-based detachment. In *Logics for New-Generation Artificial Intelligence (LNGAI2023)*. College Publications, London, United Kingdom, 2023.

[Kraus *et al.*, 1990] Sarit Kraus, Daniel Lehmann, and Menachem Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial intelligence*, 44(1-2):167–207, 1990.

[Leite and Martins, 2011] João Leite and João G. Martins. Social abstract argumentation. In Toby Walsh, editor, *IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16-22, 2011*, pages 2287–2292. IJCAI/AAAI, 2011.

[Liao and van der Torre, 2024] Beishui Liao and Leendert van der Torre. Attack-defense semantics of argumentation. In *Computational Models of Argument*, pages 133–144. IOS Press, 2024.

[Liao *et al.*, 2019] Beishui Liao, Nir Oren, Leendert van der Torre, and Serena Villata. Prioritized norms in formal argumentation. *Journal of Logic and Computation*, 29(2):215–240, 2019.

[Liao *et al.*, 2023] Beishui Liao, Pere Pardo, Marija Slavkovik, and Leendert van der Torre. The Jiminy advisor: Moral agreements among stakeholders based on norms and argumentation. *Journal of Artificial Intelligence Research*, 77:737–792, 2023.

[Longo *et al.*, 2024] Luca Longo, Mario Brcic, Federico Cabitza, Jaesik Choi, Roberto Confalonieri, Javier Del Ser, Riccardo Guidotti, Yoichi Hayashi, Francisco Herrera, Andreas Holzinger, et al. Explainable artificial intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions. *Information Fusion*, 106:102301, 2024.

[Makinson, 2005] David Makinson. *Bridges from classical to nonmonotonic logic*. King's College, 2005.

[Markovich, 2019] Réka Markovich. On the formal structure of rules in conflict of laws. In *JURIX*, pages 199–204, 2019.

[McBurney and Parsons, 2002] Peter McBurney and Simon Parsons. Games that agents play: A formal framework for dialogues between autonomous agents. *Journal of Logic, Language and Information*, 11(3):315–334, 2002.

[McBurney and Parsons, 2004] Peter McBurney and Simon Parsons. Locutions for argumentation in agent interaction protocols. In *International Workshop on Agent Communication*, pages 209–225. Springer, 2004.

[McBurney *et al.*, 2021] Peter McBurney, Simon Parsons, et al. Argument schemes and dialogue protocols: Doug walton's legacy in artificial intelligence. *FLAP*, 8(1):263–290, 2021.

[McDougall *et al.*, 2020] Rosalind McDougall, Cade Shadbolt, and Lynn Gillam. The practice of balancing in clinical ethics case consultation. *Clinical Ethics*, 15(1):49–55, 2020.

[Mercier and Sperber, 2017] Hugo Mercier and Dan Sperber. *The enigma of reason*. Harvard University Press, 2017.

[Modgil and Prakken, 2013] Sanjay Modgil and Henry Prakken. A general account of argumentation with preferences. *Artif. Intell.*, 195:361–397, 2013.

[Modgil and Prakken, 2014] Sanjay Modgil and Henry Prakken. The $ASPIC^+$ framework for structured argumentation: a tutorial. *Argument Comput.*, 5(1):31–62, 2014.

[Moore, 1993] David John Moore. *Dialogue game theory for intelligent tutoring systems*. PhD thesis, Leeds Metropolitan University, 1993.

[Musi and Palmieri, 2022] Elena Musi and Rudi Palmieri. The fallacy of explainable generative AI: evidence from argumentative prompting in two domains. 2022.

[Nash Jr, 1950] John F Nash Jr. Equilibrium points in n-person games. *Proceedings of the national academy of sciences*, 36(1):48–49, 1950.

[Niskanen *et al.*, 2018] Andreas Niskanen, Johannes P Wallner, and Matti Järvisalo. Extension enforcement under grounded semantics in abstract argumentation. In *Sixteenth International Conference on Principles of Knowledge Representation and Reasoning*, 2018.

[Nouioua and Risch, 2010] Farid Nouioua and Vincent Risch. Bipolar argumentation frameworks with specialized supports. In *2010 22nd IEEE International Conference on Tools with Artificial Intelligence*, volume 1, pages 215–218. IEEE, 2010.

[Nute, 1994] Donald Nute. Defeasible logic. *Handbook of logic in artificial intelligence and logic programming*, 3:353–395, 1994.

[Okuno and Takahashi, 2009] Kenichi Okuno and Kazuko Takahashi. Argumentation system with changes of an agent's knowledge base. In *Twenty-First International Joint Conference on Artificial Intelligence*. Citeseer, 2009.

[Pardo and Straßer, 2022] Pere Pardo and Christian Straßer. Modular orders on defaults in formal argumentation. *Journal of Logic and Computation*, 2022.

[Pardo *et al.*, 2024] Pere Pardo, Liuwen Yu, Chen Chen, and Leendert van der Torre. Weakest link, prioritised default logic and principles in argumentation. *Journal of Logic and Computation*, 2024. Forthcoming.

[Phillips *et al.*, 2021] P Jonathon Phillips, P Jonathon Phillips, Carina A Hahn, Peter C Fontana, Amy N Yates, Kristen Greene, David A Broniatowski, and Mark A Przybocki. Four principles of explainable artificial intelligence. 2021.

[Pigozzi and van der Torre, 2018] Gabriella Pigozzi and Leendert van der Torre. Arguing about constitutive and regulative norms. *Journal of Applied Non-Classical Logics*, 28(2-3):189–217, 2018.

[Polberg and Hunter, 2018] Sylwia Polberg and Anthony Hunter. Empirical evaluation of abstract argumentation: Supporting the need for bipolar and probabilistic approaches. *Int. J. Approx. Reason.*, 93:487–543, 2018.

[Pollock, 1987] John L Pollock. Defeasible reasoning. *Cognitive science*, 11(4):481–518, 1987.

[Pollock, 1992] John L Pollock. How to reason defeasibly. *Artificial Intelligence*, 57(1):1–42, 1992.

[Pollock, 1994] John L Pollock. Justification and defeat. *Artificial Intelligence*, 67(2):377–407, 1994.

[Pollock, 1995] John L Pollock. *Cognitive carpentry: A blueprint for how to build a person*. Mit Press, 1995.

[Pollock, 2001] John L Pollock. Defeasible reasoning with variable degrees of justification. *Artificial intelligence*, 133(1-2):233–282, 2001.

[Pollock, 2009] John L Pollock. A recursive semantics for defeasible reasoning. In *Argumentation in artificial intelligence*, pages 173–197. Springer, 2009.

[Pollock, 2010] John L Pollock. Defeasible reasoning and degrees of justification. *Argument and Computation*, 1(1):7–22, 2010.

[Prakken and Sartor, 2015] Henry Prakken and Giovanni Sartor. Law and logic: A review from an argumentation perspective. *Artificial intelligence*, 227:214–245, 2015.

[Prakken, 2010a] Henry Prakken. An abstract framework for argumentation with structured arguments. *Argument & Computation*, 1(2):93–124, 2010.

[Prakken, 2010b] Henry Prakken. Slides on nonmonotonic logic for commonsense reasoning, 2010.

[Prakken, 2013] Henry Prakken. *Logical tools for modelling legal argument: a study of defeasible reasoning in law*, volume 32. Springer Science & Business Media, 2013.

[Prakken, 2018] Henry Prakken. Historical overview of formal argumentation. In *Handbook of formal argumentation*, pages 73–141. College Publications, 2018.

[Prakken, 2020] Henry Prakken. On validating theories of abstract argumentation frameworks: the case of bipolar argumentation frameworks. In *Proceedings of the 8th Workshop on Computational Models of Natural Argument (CMNA 2020), Perugia, Italy (and online)*. CEUR-WS.org, 2020.

[Prakken, 2024a] Henry Prakken. An abstract and structured account of dialectical argument strength. *Artificial Intelligence*, 335:104193, 2024.

[Prakken, 2024b] Henry Prakken. On evaluating legal-reasoning capabilities of generative ai. In *Proceedings of the 24th Workshop on Computational Models of Natural Argument*, pages 100–112, 2024.

[Priest, 2002] Graham Priest. Paraconsistent logic. In *Handbook of philosophical logic*, pages 287–393. Springer, 2002.

[Rahwan *et al.*, 2003] Iyad Rahwan, Sarvapali D Ramchurn, Nicholas R Jennings, Peter McBurney, Simon Parsons, and Liz Sonenberg. Argumentation-based negotiation. *The Knowledge Engineering Review*, 18(4):343–375, 2003.

[Rawls, 2001] John Rawls. *Justice as fairness: A restatement*. Harvard University Press, 2001.

[Reiter, 1980] Raymond Reiter. A logic for default reasoning. *Artificial intelligence*, 13(1-2):81–132, 1980.

[Rienstra *et al.*, 2020] Tjitze Rienstra, Chiaki Sakama, Leendert van der Torre, and Beishui Liao. A principle-based robustness analysis of admissibility-based argumentation semantics. *Argument & Computation*, 11(3):305–339, 2020.

[Ross *et al.*, 2023] Steven I Ross, Fernando Martinez, Stephanie Houde, Michael Muller, and Justin D Weisz. The programmer's assistant: Conversational interaction with a large language model for software development. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*, pages 491–514, 2023.

[Roth and Verheij, 2004] Bram Roth and Bart Verheij. Cases and dialectical arguments–an approach to case-based reasoning. In *On the Move to Meaningful Internet Systems 2004*, pages 634–651. Springer, 2004.

[Russell and Norvig, 2010] Stuart J Russell and Peter Norvig. *Artificial intelligence: A modern approach*. London, 2010.

[Savage, 1972] Leonard J Savage. *The foundations of statistics*. Courier Corporation, 1972.

[Schindler *et al.*, 2020] Samuel Schindler, Anna Drozdzowicz, and Karen Brøcker. *Linguistic intuitions: Evidence and method*. Oxford University Press, 2020.

[Searle, 1979] John R Searle. *Expression and meaning: Studies in the theory of speech acts*. Cambridge University Press, 1979.

[Sklar and Azhar, 2018] Elizabeth I Sklar and Mohammad Q Azhar. Explanation through argumentation. In *Proceedings of the 6th International Conference on Human-Agent Interaction*, pages 277–285, 2018.

[Straßer and Arieli, 2019] Christian Straßer and Ofer Arieli. Normative reasoning by sequent-based argumentation. *Journal of Logic and Computation*, 29(3):387–415, 2019.

[Sycara, 1990] Katia P Sycara. Persuasive argumentation in negotiation. *Theory and decision*, 28:203–242, 1990.

[Tennent, 1991] Robert D Tennent. Semantics of programming languages. *(No Title)*, 1991.

[Toulmin, 1958] Stephen E Toulmin. *The uses of argument*. Cambridge university press, 1958.

[Traum, 1999] David R Traum. Speech acts for dialogue agents. In *Foundations of rational agency*, pages 169–201. Springer, 1999.

[Turner, 2004] Hudson Turner. Strong equivalence for causal theories. In *Logic Programming and Nonmonotonic Reasoning: 7th International Conference, LPNMR 2004 Fort Lauderdale, FL, USA, January 6-8, 2004 Proceedings 7*, pages 289–301. Springer, 2004.

[Ulbricht and Wallner, 2021] Markus Ulbricht and Johannes P Wallner. Strong explanations in abstract argumentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 6496–6504, 2021.

[van der Lee *et al.*, 2021] Chris van der Lee, Albert Gatt, Emiel van Miltenburg, and Emiel Krahmer. Human evaluation of automatically generated text: Current trends and best practice guidelines. *Computer Speech & Language*, 67:101151, 2021.

[van der Torre and Vesic, 2018] Leendert van der Torre and Srdjan Vesic. The principle-based approach to abstract argumentation semantics. In Pietro Baroni, Dov Gabbay, Massimiliano Giacomin, and Leendert van der Torre, editors, *Handbook of formal argumentation, Volume 1*, pages 797–838. College Publictions, 2018.

[Van Laar and Krabbe, 2018] Jan Albert Van Laar and Erik CW Krabbe. The role of argument in negotiation. *Argumentation*, 32(4):549–567, 2018.

[Villata *et al.*, 2011] Serena Villata, Guido Boella, and Leendert van der Torre. Attack semantics for abstract argumentation. In *IJCAI*. IJCAI/AAAI, 2011.

[Vilone and Longo, 2021] Giulia Vilone and Luca Longo. Classification of explainable artificial intelligence methods through their output formats. *Machine Learning and Knowledge Extraction*, 3(3):615–661, 2021.

[Visser *et al.*, 2020] Jacky Visser, John Lawrence, and Chris Reed. Reason-checking fake news. *Communications of the ACM*, 63(11):38–40, 2020.

[Von Neumann and Morgenstern, 1947] John Von Neumann and Oskar Morgenstern. Theory of games and economic behavior, 2nd rev. 1947.

[Walton and Krabbe, 1995] Douglas Walton and Erik CW Krabbe. *Commitment in dialogue: Basic concepts of interpersonal reasoning*. State University of New York Press, 1995.

[Walton *et al.*, 2008] Douglas Walton, Christopher Reed, and Fabrizio Macagno. *Argumentation schemes*. Cambridge University Press, 2008.

[Walton, 1990] Douglas N Walton. What is reasoning? What is an argument? *The Journal of Philosophy*, 87(8):399–419, 1990.

[Walton, 2013] Douglas Walton. *Argumentation schemes for presumptive reasoning*. Routledge, 2013.

[Weinstock, 2006] Michael P Weinstock. Psychological research and the epistemological approach to argumentation. *Informal Logic*, 26(1):103–120, 2006.

[Williamson and Gabbay, 2005] Jon Williamson and Dov Gabbay. Recursive causality in Bayesian networks and self-fibring networks. *Laws and models in the sciences*, pages 173–221, 2005.

[Young *et al.*, 2016] Anthony P. Young, Sanjay Modgil, and Odinaldo Rodrigues. Prioritised default logic as rational argumentation. In Catholijn M. Jonker, Stacy Marsella, John Thangarajah, and Karl Tuyls, editors, *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, pages 626–634. ACM, 2016.

[Young *et al.*, 2017] Anthony P. Young, Sanjay Modgil, and Odinaldo Rodrigues. On the interaction between logic and preference in structured argumentation. In Elizabeth Black, Sanjay Modgil, and Nir Oren, editors, *Theory and Applications of Formal Argumentation - 4th International Workshop*, volume 10757, pages 35–50. Springer, 2017.

[Yu and Gabbay, 2022] Liuwen Yu and Dov Gabbay. Case-based reasoning via comparing the strength order of features. In *International Workshop on Explainable, Transparent Autonomous Agents and Multi-Agent Systems*, pages 143–151. Springer, 2022.

[Yu *et al.*, 2020] Liuwen Yu, Réka Markovich, and Leendert Van Der Torre. Interpretations of support among arguments. In *Legal Knowledge and Information Systems*, pages 194–203. IOS Press, 2020.

[Yu *et al.*, 2021] Liuwen Yu, Dongheng Chen, Lisha Qiao, Yiqi Shen, and Leendert van der Torre. A Principle-based Analysis of Abstract Agent Argumentation Semantics. In *Proceedings of the 18th International Conference on Principles of Knowledge Representation and Reasoning*, pages 629–639, 11 2021.

[Yu *et al.*, 2022] Liuwen Yu, Mirko Zichichi, Markovich Réka, Najjar Amro, et al. Intelligent human-input-based blockchain oracle (IHiBO). In *Proceedings of the 14th International Conference on Agents and Artificial Intelligence (ICAART 2022)*, pages 1–12. SCITEPRESS, 2022.

[Yu *et al.*, 2023] Liuwen Yu, Caren Al Anaissy, Srdjan Vesic, Xu Li, and Leendert van der Torre. A principle-based analysis of bipolar argumentation semantics. In *European Conference on Logics in Artificial Intelligence*, pages 209–224. Springer, 2023.

[Yu, 2023] Liuwen Yu. *Distributed Argumentation Technology*. PhD thesis, 2023.