Tesis doctoral por compendio de publicaciones

Arguments to believe and beliefs to argue

Epistemic logics for argumentation and its dynamics

Antonio Yuste-Ginel

Director: Alfredo Burrieza Muñiz

Programa de Doctorado Estudios Avanzados en Humanidades

UNIVERSIDAD DE MÁLAGA

Departamento de Filosofía

Facultad de Filosofía y Letras

Febrero de 2022

UNIVERSIDAD DE MÁLAGA

AUTOR: Antonio Yuste Ginel

https://orcid.org/0000-0002-4380-3095

EDITA: Publicaciones y Divulgación Científica. Universidad de Málaga

DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD DE LA TESIS PRESENTADA PARA OBTENER EL TÍTULO DE DOCTOR

D./Dña ANTONIO YUSTE GINEL
Estudiante del programa de doctorado Estudios Avanzados en Humanidades de la Universidad de Málaga, autor/a de la tesis, presentada para la obtención del título de doctor por la Universidad de Málaga, titulada: ARGUMENTS TO BELIEVE AND BELIEFS TO ARGUE. EPISTEMIC LOGICS FOR ARGUMENTATION AND ITS DYNAMICS

Realizada bajo la tutorización de ALFREDO BURRIEZA MUÑIZ y dirección de ALFREDO BURRIEZA MUÑIZ (si tuviera varios directores deberá hacer constar el nombre de todos)

DECLARO QUE:

La tesis presentada es una obra original que no infringe los derechos de propiedad intelectual ni los derechos de propiedad industrial u otros, conforme al ordenamiento jurídico vigente (Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia), modificado por la Ley 2/2019, de 1 de marzo.

Igualmente asumo, ante a la Universidad de Málaga y ante cualquier otra instancia, la responsabilidad que pudiera derivarse en caso de plagio de contenidos en la tesis presentada, conforme al ordenamiento jurídico vigente.

En Málaga, a 8 de FEBRERO de 2022

YUSTE GINEL ANTONIO - Fdo.: Doctorando/a	Firmado por BURRIEZA MUÑIZ ALFREDO -el día 10/02/2022 con un certificado emitido por AC FNMT Usuarios Fdo.: Tutor/a
Firmado por BURRIEZA MUÑIZ ALFREDO -el día 10/02/2022 con un certificado emitido por AC FNMT Usuarios Fdo.: Director/es de tesis	

Edificio Pabellón de Gobierno. Campus El Ejido. 29071
Tel.: 952 13 10 28 / 952 13 14 61 / 952 13 71 10
E-mail: doctorado@uma.es

UNIVERSIDAD DE MÁLAGA

Departamento de Filosofía

Autorización de depósito e informe de director de tesis doctoral

Por la presente, y como director de tesis de Antonio Yuste-Ginel, le concedo a este último mi visto bueno para que deposite su tesis *Arguments to believe and beliefs to argue. Epistemic logics for argumentation and its dynamics*. La tesis, elaborada en la modalidad compendio de publicaciones, cumple todos los requisitos para la obtención del título de doctor. En particular, el manuscrito introduce, motiva, presenta y discute una serie de resultados originales en la intersección entre la lógica epistémica dinámica y la argumentación formal.

Firmado por BURRIEZA MUÑIZ ALFREDO - el día 10/02/2022 con un certificado emitido por AC FNMT Usuarios

Fdo.: Alfredo Burrieza Muñiz.

Contents

Abbreviations of contributions' titles

Paper I	*Epistemic attitudes and persuasive argumentation* (joint work with Carlo Proietti)
Paper II	*Dynamic epistemic logics for abstract argumentation* (joint work with Carlo Proietti)
Paper III	*On the epistemic logic of incomplete argumentation frameworks* (joint work with Andreas Herzig)
Paper IV	*Multi-agent abstract argumentation frameworks with incomplete knowledge of attacks* (joint work with Andreas Herzig)
Paper V	*Basic beliefs and argument-based beliefs in awareness epistemic logic with structured arguments* (joint work with Alfredo Burrieza)
Paper VI	*An awareness epistemic framework for belief, argumentation and their dynamics* (joint work with Alfredo Burrieza)

a mis padres, Concha y Antonio, con mucho amor

Agradecimientos

La ciencia es una empresa colectiva. La parte minúscula, casi invisible, de tan gigantesca empresa que representa esta tesis no se libra de dicha caracterización. Así, las primeras líneas dedicadas a expresar mi gratitud están dirigidas a los coatures de los artículos que componen el corazón de esta tesis: Alfredo Burrieza, Andreas Herzig y Carlo Proietti. Primero, he de dar las gracias a Alfredo que, además de figurar como autor en dos de estas contribuciones, ha llevado a cabo la tediosa tarea de tutorizar y dirigir a un estudiante tan disperso como yo. Desde el principio, su apoyo y guía han sido fundamentales para no perderme en el abismo. Segundo, a Andreas, no solo por ser un excelente tutor académico y colaborador durante y después de mi estancia en Toulouse (lo que no fue nada fácil, debido a la pandemia), sino por ofrecerme un hogar. Tercero, pero solo por orden alfabético y no de importancia, he de agradecer a Carlo todo su trabajo y la inmensidad de cosas que he aprendido de él y gracias a él. Ojalá que algún día llegue a escribir con la claridad y estilo que tú lo haces.

Además de mis colaboradores directos, hay mucha gente cuyo trabajo ha hecho posible esta tesis de una u otra forma. Fernando Velázquez Quesada ocupa un lugar especial en esta categoría, ya que me recibió en mi estancia en Ámsterdam con gran hospitalidad, escuchando mis ideas ingenuas y orientándome con maestría en la bibliografía. Jose Pedro Úbeda Rives me acompañó en mis primeros pasos en la investigación, mediante la dirección de mi trabajo final de máster, y elaboró también generosos y detallados comentarios sobre algunos de los artículos que figuran en esta tesis. Debo extender mi enorme agradecimiento profesional a las/os investigadoras/es que han accedido a formar parte del tribunal de esta tesis o a actuar como evaluadores de la misma, ya sea como titulares o suplentes, aprovechando además para expresar mi admiración por su trabajo: Cristina Barés, Ringo Baumann, Claudia Fernández, Hans van Ditmarsch, Matthieu Fontaine, Davide Grossi, Jean-Guy Mailly, Manuel Ojeda-Aciego, y Chenwei Shi.

Continúo agradeciendo su amor y comprensión durante este periodo a toda mi familia, la biológica y la elegida, empezando por Giulia, mi compañera existencial, que lo ha llenado todo de luz, y me ha enseñado que la bondad desarma. A mis padres, a los que dedico este trabajo, Concha y Antonio, a mis hermanas, Candela y Marta, a mi jovial abuela Concha, a la Tits (que hay solo una), y a todos los demás. Gracias a todos mis amigos, a los que quiero con locura (afortunadamente, conforman una lista demasiado larga como para explicitarla aquí).

Termino mencionando a la gente maravillosa que he conocido estos años dentro de

o gracias a la academia, o que ya conocía pero con la que he tenido el placer de volver a compartir tiempo. A mi eterno compañero, primero de colegio, luego de carrera y finalmente de despacho: Pablo García-Barranquero. Al resto de compis de doctorado, con especial mención a mi hermana académica Claudia Fernández, Carlos Aguilera, a Andrés Ortigosa y a Antonio Rovi. A mis camaradas de *Epojé o Muerte*, por colectivizar el sufrimiento que conlleva atravesar un doctorado en filosofía. A toda la gente que me han permitido conocer mis dos estancias predoctorales: Arthur y Michelle, Laura, Kelly, Saúl, Emily, María José, Clara, ambos Victors, Dazhu, Fausto, y un largo etcétera. A todo el departamento de filosofía de la UMA, con especial mención a Antonio Diéguez, por tolerarme como compañero de oficina, a Pedro Chamizo, por permitirme impartir unas horas de filosofía del lenguaje, a Francisco Roldán, al que deseo una pronta recuperación, por su incansable trabajo, y a Jorge Costa, por pintar de colores la parte más oscura de la pandemia. Por último, le doy las gracias al personal de la cafetería de psicología, por su excelente servicio diario y por su alegría.

Acknowledgements

Science is a collective enterprise. The tiny, almost invisible part of such a huge enterprise represented by this dissertation does not fall out of this characterization. Therefore, my first words of gratitude are dedicated to the coauthors of the contributions that form the core of this work: Alfredo Burrieza, Andreas Herzig and Carlo Proietti. I should first thank Alfredo, not only for coauthoring two of the papers, but also for facing the tedious task of supervising the disperse student that I usually am. Right from the start, his support and guidance have been crucial for me not to be lost in the abyss. Second, I would like to thank Andreas, both for being an excellent supervisor and colleague during and after my visit to the IRIT (which was not that easy, due to the pandemic), as well as for being a great host. Last but no least, I must thank Carlo for all his work and for the huge amount of things that I have learnt from him. I wish to write as clearly and stylishly as you do some day in a remote future.

In addition to direct collaborators, there are many people whose work made this thesis possible in one or another way. Fernando Velázquez Quesada has been notably important within this category, as he received me during my visit to the ILLC with great hospitality, listening to my vague ideas and guiding me masterly through the vast literature on DEL and awareness logic. José Pedro Úbeda Rives greatly helped me during my first research experience, by supervising my master thesis. He also elaborated detailed comments about some of the works that constitute this thesis. I should extend my gratitude to all the researches that generously accepted the invitation to be part of the committee or to act as reviewers: Cristina Barés, Ringo Baumann, Claudia Fernández, Hans van Ditmarsch, Matthieu Fontaine, Davide Grossi, Jean-Guy Mailly, Manuel Ojeda-Aciego, and Chenwei Shi. I also take advantage to express my great admiration for their work, which provided solid foundations for the thoughts presented in this dissertation.

Let me continue these words of acknowledgement by thanking all my family, the biological and the chosen one, for their incalculable love and understanding during these years. First of all, thanks to Giulia, my existential partner, who illuminated everything in my life and taught me that kindness is always disarming. I should also thank my parents, Concha and Antonio, my dear sisters, Candela and Marta, my cheerful grandma, Concha, my very unique Tits, and everyone else. Infinitely many thanks to all my friends (luckily, this list is too long so as to make it explicit here). I deeply love you, guys.

I close this –surely insufficient– lines by mentioning all the wonderful people I met these years thanks to the studies that I am about to conclude. Thanks to my eternal mate:

first classmate at school, then at the philosophy degree and finally my officemate: Pablo García-Barranquero. To all my other colleagues in the PhD program, with special mention to my academic sister, Claudia Fernández, as well as to Carlos Aguilera, Andrés Ortigosa, and Antonio Rovi. To my comrades of *Epojé o Muerte*, for making collective the suffering that entails going through a PhD in philosophy. To all the great people I met during my two predoctoral research visits: Arthur, Michelle, Laura, Kelly, Dazhu and Fausto in Amsterdam; and Saúl, Emily, María José, Clara, and both Victors in Toulouse. Thanks to all the Department of Philosophy of UMA, with a special mention to Antonio Diéguez, for tolerating me as a rather messy officemate, to Pedro Chamizo, for letting me being his teaching assistant in the Philosophy of Language course, to Francisco Roldán, for his tireless work (wishing him an early recovery), and to Jorge Costa, for colouring the most obscure days of the pandemic. Finally, I should also thank all the staff of the cafeteria at the Faculty of Psychology, for their excellent service and their daily joy.

Chapter 1

Introduction

In a motto, this dissertation is about combining two well-known families of formalisms for knowledge representation, namely, *Epistemic Logic* (EL) and its dynamic extensions, on the one side, and *Formal Argumentation* (FA) on the other. This general goal is motivated by a strong (but perhaps rather trivial) intuition: the informal notions that EL deals with (e.g., *belief*, *knowledge* and related *epistemic attitudes*,[1] among others) and the ones that FA deals with (e.g., *argument*, *argument strength*, *conflict*, or *acceptability*) are strongly intertwined in a variety of senses. Let us illustrate this idea with a couple of simple examples.

Example 1 (The chocolate bar robbery)**.** *Anne is trying to convince Bob that she did not eat his chocolate candy. In order to do so, she thinks of two possible alibis. First, she can say that she is allergic to peanuts (as the stolen candy contained peanuts). Second, she can say that she was in her office at the time that the robbery took place. Anne believes that Bob has access to the security cameras of her office but he cannot check her medical records. Consequently, Anne decides to use her alleged allergy as an alibi.*

Example 2 (Anne and the weather)**.** *Anne is in her office at the University of Málaga. There are no windows in the room, as she is a miserable PhD student. She wonders whether it is raining outside. She first asks a colleague who answers "well, the sky looked cloudless when I came here, a couple of hours ago". After that, she opens her browser and checks the weather forecast in Málaga. The forecast says there is 80 % probability of rain. Anne comes to believe that it is raining outside.*

The crucial notions taking part in these examples, notably the ones of *arguing* and *believing*, have attracted the attention of diverse disciplines from Antiquity to nowadays. They are especially present in core topics of both contemporary philosophy and ongoing research in artificial intelligence. Formulated from a different perspective, the main objective of this dissertation is studying some of the possible relations among these notions

[1] We use *epistemic attitude* to jointly refer to the notions of belief, knowledge and related ones (subtypes of the previous ones, opinions, intuitions, etc). In general, vague terms, we can think of an epistemic attitude as a cognitive relation among an intelligent being (for instance, a human individual) and some kind of object (for instance, but not always, a proposition).

from a logical point of view, that is to say, focusing on systematic reasoning about these relations. And this is going to be done by combining EL and FA.

The choice of our methodology is arguably natural. *Epistemic logic* (Hintikka, 1962; Meyer and van der Hoek, 1995; Fagin et al., 2004), together with its dynamic extensions, known as *Dynamic Epistemic Logic* (DEL) (van Ditmarsch et al., 2007; van Benthem, 2011), provide well-known tools for qualitatively representing epistemic attitudes (mainly knowledge and belief) as well as their dynamics. *Formal argumentation* (Baroni et al., 2018b; Gabbay et al., 2021), on its side, is the broad research field (at the crossroads of linguistics, artificial intelligence, logic and formal philosophy) where formal representations of argumentative notions are investigated. As it will be argued later on, the notion of *awareness*, as treated in the EL tradition since Fagin and Halpern (1987), can be used as a theoretical bridge among both research areas, EL and FA (or, alternatively, among both sets of notions: epistemic and argumentative ones). This is doable modulo a slight but conceptually relevant change: jumping from the well-studied *awareness of sentences* to the less standard *awareness of arguments*. We will come back to methodological aspects later on, but before that, let us analyse more in detail the research problem investigated through this work.

1.1 Research problem

Believing and arguing are two central dimensions of the cognitive architecture of human beings that appear intertwined in their daily life. More in general, it seems rather easy to accept that some of the fundamental notions of argumentation theory –such as the very notion of *argument*, *conflict among arguments*, or *argument acceptability*– and some of the central notions of epistemology –such as *belief*, *knowledge* or *justification*– are also strongly connected. This connection happens in, at least, two different ways; those that are respectively captured in examples 1 and 2. Generalizing Example 1, we can assert that

C1 *the evaluation that an agent performs of her available arguments is influenced by her epistemic attitudes*.

Or, in other words, *argument evaluation is conditioned by the formation of epistemic attitudes*. This thesis is rather general and vague. Indeed, it can actually be specified in several ways. For instance, as a special, rhetorical case of it we have that:[2]

$\mathsf{C1}_{\mathsf{rhetoric}}$ *Higher-order epistemic attitudes conditions rhetoric argument evaluation.*

This is the interpretation of C1 illustrated with Example 1, according to which if an agent is looking for her best argument to persuade an opponent, she should choose (and she in fact usually does) the one that she *believes* to be more persuasive, and this in turn depends on what she believes of her interlocutor's mental states. In the example, Anne decides to

[2]We use *rhetoric* and derived terms to refer to all aspects of argumentation that take persuasion as its main focus.

use the allergy alibi due to her belief about Bob's argumentative situation (that is, about the arguments Bob is aware of).

However, $\mathsf{C1}_{\mathsf{rhetoric}}$ is not the only possible interpretation of C1. For instance, if we look at Example 2, it seems clear that Anne considers strictly stronger the argument based on the weather forecast than the one based on her colleague's opinion, but why? One possible answer, that illustrates the second reading of C1 that we will deal with, is that Anne does not trust her colleague's testimony, as she suspects that he is lying compulsively or trying to make a joke, while she does believe in the veracity of the forecast information displayed in her screen.[3] The general evaluative principle underlying this kind of explanation is:

$\mathsf{C1}_{\mathsf{epistemic}}$ *arguments with believed (respectively, known) premisses are to be preferred to arguments with premisses that are not believed (respectively, that are unknown).*

Concurrently, in Example 2, Anne comes to believe that it is raining out there because she considers the forecast argument as strictly stronger than her college testimony. The example is rather straightforward and things can get much more complicated. For instance, a third argument questioning the veracity of the information displayed in Anne's screen could come into play (imagine, for example, that she gets an email in her mobile phone warning her of the presence of fake forecast banners caused by malicious software infecting the University's computers). Evidently, this could affect her belief formation process, e.g., by forcing her to suspend her judgement, or by reinstating the authority of her colleague's testimony and thus making her believe that it is raining. In any case, the principle according to which Anne forms her belief based on her assessment of the arguments that are available at each moment can be formulated in short as *belief formation is conditioned by argument evaluation*. More in detail, and generalizing from belief to any epistemic attitude, we can assert that:

C2 *a reasonable epistemic agent should take into account her available arguments, as well as their strength, in order to form her epistemic attitudes towards a given topic*.

A deeper understanding of these principles relating argumentation and epistemic attitudes requires a more precise characterization of the involved notions (belief, argument, argument strength, etc). But, before providing such a characterization, let us make our main goal a bit more specific: it consists in looking at (possible different readings of) C1 and C2 from a logical point of view, that is, focusing on how to model them in order to perform precise reasoning, through the joint use of EL and FA as the main tool. Besides giving examples and pointing out their intuitive appeal, we will not argue in favour of C1 or C2; but rather accept them as starting points. Moreover, for most of the dissertation we focus on a particular epistemic attitude: *belief* –although sometimes narrowing our analysis to knowledge– and on a particular aspect of argumentation: argument evaluation processes –although sometimes jumping to argument generation and communication.

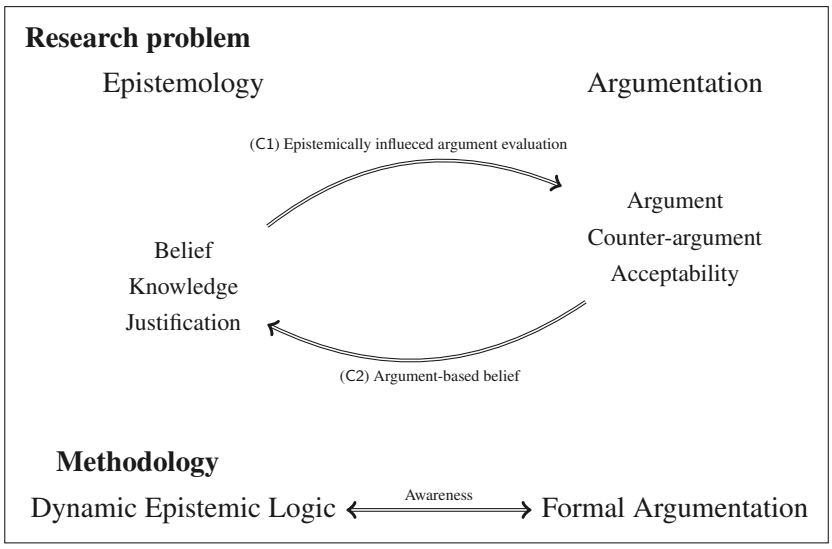


Figure 1.1: Conceptual map: panoramic view of the dissertation. The main research problem is depicted in the top part while the methodology is depicted in the bottom part.

Figure 1.1 sums up in a conceptual map what we have said so far, depicting the research problem and methodology of this dissertation.

1.2 Beliefs and arguments

Up to now, we have assumed a common, vague enough understanding of our two central notions: *belief* and *argument*. We will explain in detail how we have modelled them in this thesis, but before that, let us expose some of the intuitions and main concepts that underlie these models.

Belief is an ubiquitous notion in epistemology and cognitive sciences, since intelligent agents are supposed to hold beliefs in order to interact successfully with their environment. Among the heterogeneous options to model belief mathematically (see Genin and Huber (2021)), we have chosen standard EL (Fagin et al., 2004; Meyer and van der Hoek, 1995), sometimes enriched with tools imported from *awareness logic* (Fagin and Halpern, 1987). In the picture provided by EL, belief can be paraphrased as *true in all doxastic alternatives considered by the agent*, that is, in all situations that the agent is not able to differentiate from the current world using her available information.

Qualitative and full belief. The first feature that distinguishes EL from other modelling options is that it sees belief as a *completely qualitative notion*, meaning that it contains

[3] An alternative explanation is that Anne does believe that both pieces of information are true, but she considers more reliable the inference link of one of them.

no numeric information to measure its strength, as opposed, for instance, to the long-standing tradition working on the subjective interpretation of probabilities, in which each proposition is assigned a probability, informally representing the degree of credence, and belief is then defined as credence over some given threshold. Besides, and differently to other qualitative models, *EL beliefs are not gradual* at all, so that something is either believed or not, and there is no further distinction between stronger and weaker beliefs –as, for instance, in the closely related *plausibility models* of Baltag and Smets (2008).

The object of belief. What is this *something* that is believed? In other words, what is the object of belief? In EL, this question can receive two possible answers: propositions or sentences. A *proposition* is typically a semantic object whose primary property is the ability to bear truth values. In the picture of EL equipped with Kripke semantics, a proposition can be understood as a set of *possible worlds* or situations which is true if and only if it has the actual situation as a member (see Stalnaker (1976) for a detailed explanation of this conception of propositions). Hence, a proposition is believed if and only if it is a subset of the set of doxastic alternatives of the agent. Alternatively, we can also say that the primary objects of EL beliefs are *sentences*, these are linguistic/syntactic objects, usually represented as formulas of a given language. This answer nicely corresponds to the paraphrase of belief that we announced previously: a sentence is believed if and only if it is true in all doxastic alternatives of the agent. Note that, under the assumption that agents are perfect reasoners (something that comes with the conceptual pack of standard EL) both options are equivalent, since a perfect reasoner believes that a sentence φ is true if and only if she also believes that every sentence ψ that is logically equivalent to φ is true.

Higher-order beliefs. Perhaps one of the most appealing features of EL models is that they provide a clear, compact picture of arbitrarily *higher-order beliefs*. In other words, we can express not only what each agent believes about the external world, but also what she believes that other agents believe, and what she believes that other agents believe that other agents believe, etc. As we have mentioned, these higher-order attitudes have an important role when an agent assesses the persuasive power of a given argument.

We now move to present some of the features underlying our view of arguments.

What is an argument? This is the primary question of informal logic (Groarke, 2017), understood as a branch of argumentation theory. Of course, we do not attempt to say anything significant here for the general debate, but rather to sketch three answers that underlie the most common mathematical representations of arguments.

An argument is a pair. In informal logic, an argument is sometimes defined "as an attempt to provide evidence in favour of some point of view" (Groarke, 2017), where evidence is usually understood in a rather loose sense. According to this definition, an argument can be naturally divided in at least two well-differentiated components: the *premisses* (also called *reason(s)*) that provide (or attempt to provide) some kind of support for

the *conclusion* (or *claim*). This explains that many formal representations of arguments model them as pairs of the form (Γ, φ) where Γ and φ are respectively a set of formulas and a formula of a previously fixed formal language, standing respectively for the premisses and the conclusion of the argument. This picture makes perfect sense when the set of *inference rules* employed in the construction of the argument is homogeneous in nature and primacy. More in particular, the representation of arguments as pairs (premisses, conclusion) fits well to the assumption that the only available way of performing inference (going from one piece of information to other) is some sort of deduction (Besnard and Hunter, 2018).

An argument is a syntactic tree. Nevertheless, if the set of inference rules is stratified, e.g., according to their level of reliability, or simply divided into defeasible rules and deductive rules, then the representation of an argument should make explicit what links are being used at each step of the argument. One of the main reasons for doing so is that there might be different arguments to conclude φ departing from a set of premisses Γ, and some of them might be stronger than others due to the links they contain. As an illustration of this way of representing arguments mathematically, suppose that $\twoheadrightarrow$ stands for strict (deductive) inference while $\Rightarrow$ stands for defeasible inference, we can then form the tree-like argument

$$\langle\langle\langle \mathsf{Bird}\rangle, \langle \mathsf{Bird} \rightarrow \mathsf{Wings}\rangle \twoheadrightarrow \mathsf{Wings}\rangle \Rightarrow \mathsf{Flies}\rangle$$

whose initial premisses are the formulas Bird and Bird $\rightarrow$ Wings, and it uses a strict inference rule (*modus ponens*) for concluding Wings deductively from the initial premisses, and a defeasible rule for concluding Flies defeasibly from the intermediary conclusion Wing.[4]

An argument is a node in a directed graph. Or, said in simpler words, it really does not matter (for some purposes) to exactly account for what an argument is, but rather for how it relates (dialectically) with other arguments. To put it in Dung (1995)'s words:

"Here [in this work], an argument is an abstract entity whose role is solely determined by its relations to other arguments. No special attention is paid to the internal structure of the arguments."

The three different answers to the question "*What is an argument?*" that we have just sketched constitute fundamental design choices when mathematically modelling an argument. This thesis focuses on the second and the third one, but the first one is extensively discussed in the literature and it can be thought of as the minimal account of the structure of an argument.

How strong is an argument? The second fundamental question that articulates our intuitive guidelines to argumentation asks for the strength of arguments. Following the work of Beirlaen et al. (2018), that builds upon decades of tradition, the very notion of *argument strength* can be analysed (as far as formal approaches are concerned) as split into three different dimensions or tiers.

[4] All these notions, imported from ASPIC$^+$ arguments, will be precisely defined in Chapter 2.

The *support dimension* deals with the question of how strong is the support provided by the premisses and inference rules of an argument to accept its conclusion(s). This dimension analyses arguments individually and it can be qualified as *internal*.

The *dialectical dimension* of argument strength looks at the dialectical relations that exist between arguments. For instance, but not exclusively, relations of *attack*, *defence*, *support* or *defeat*.

The *evaluative dimension* provides general recipes for selecting sets of acceptable arguments within a set that stand on dialectical relations between them. For instance, a basic evaluative principle is that we should not accept two arguments whose conclusions are logically incompatible or conflictive (this is a form of *conflict-freeness*).

Note that while structured representations of arguments (e.g., arguments-as-pairs and arguments-as-trees) enable the study of the support dimension and of the dialectical one, abstract representations of arguments (arguments-as-points) suffice to account for the evaluative dimension.

1.3 How to read this dissertation

We close this introduction by providing an outline of the rest of the manuscript, as well as some reading guidelines. This thesis is presented as a *collection of published papers* [*compendio de publicaciones*], sometimes called an *article-based thesis*, meaning that its main contributions are contained in the reprint of the following list of papers, placed in Chapter 4:

1. (Proietti and Yuste-Ginel, 2020) (to which we will refer as Paper I, from now on). Full reference: Proietti, C. and Yuste-Ginel, A. (2020). Persuasive argumentation and epistemic attitudes. In Soares Barbosa, L. and Baltag, A., editors, *Dynamic Logic. New Trends and Applications*, volume 12005 of LNCS, pages 104–123. Springer. DOI: 10.1007/978-3-030-38808-9_7. A preprint version is available at https://eprints.illc.uva.nl/id/eprint/1736/1/Proietti_Yuste_PAEP_preprint.pdf.

2. (Proietti and Yuste-Ginel, 2021) (abbreviated as Paper II). Full reference: Proietti, C. and Yuste-Ginel, A. (2021). Dynamic epistemic logics for abstract argumentation. *Synthese*, 199(3): 8641–8700. DOI: 10.1007/s11229-021-03178-5. Available at: https://link.springer.com/content/pdf/10.1007/s11229-021-03178-5.pdf.

3. (Herzig and Yuste-Ginel, 2021c) (abbreviated as Paper III). Full reference: Herzig, A. and Yuste-Ginel, A. (2021c). On the epistemic logic of incomplete argumentation frameworks. In M. Bienvenu, G. Lakemeyer, and E. Erdem, editors, *Proceedings of the 18th International Conference on Principles of Knowledge Representation and Reasoning*, pages 681–685. IJCAI Organization. DOI: 10.24963/kr.2021/69. Available at: https://proceedings.kr.org/2021/69/.

4. (Herzig and Yuste-Ginel, 2021b) (abbreviated as Paper IV). Full reference: Herzig, A. and Yuste-Ginel, A. (2021b). Multi-agent abstract argumentation frameworks

with incomplete knowledge of attacks. In Zhou, Z.-H., editor, *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pages 1922–1928. IJCAI Organization. DOI: 10.24963/ijcai.2021/265. Available at: https://doi.org/10.24963/ijcai.2021/265.

5. (Burrieza and Yuste-Ginel, 2020) (abbreviated as Paper V). Full reference: Burrieza, A. and Yuste-Ginel, A. (2020). Basic beliefs and argument-based beliefs in awareness epistemic logic with structured arguments. In Prakken et al., editors, *Proceedings of the COMMA 2020*, pages 123–134. IOS Press. DOI: 10.3233/FAIA200498. Available at: https://ebooks.iospress.nl/volumearticle/55364.

6. (Burrieza and Yuste-Ginel, 2021) (abbreviated as Paper VI). Full reference: Burrieza, A. and Yuste-Ginel, A. (2021). An awareness epistemic framework for belief, argumentation and their dynamics. In Halpern and Perea, editors, *Proceedings TARK 2021*, EPTCS 335, pp. 69–83. Open Publishing Association. DOI: 10.4204/EPTCS.335.6. Available at: https://doi.org/10.4204/EPTCS.335.6.

Moreover, these contributions are divided into two main research tracks, that are contained respectively in Section 4.1 and Section 4.2. The purpose of the precedent chapters is depicting a thematic unity for all these contributions. More precisely, this unity is pursued in a series of steps: describing the research problem and goals that are common to all papers (as we have just done); introducing the used methodology, in the form of technical preliminaries (Chapter 2); explaining how the different contributions aim to approach the research problem (Chapter 3); discussing collectively all their results and comparing it with closely related work (Chapter 5); and extracting some general conclusions and paths for future work (Chapter 6). Moreover, in the form of appendices, we provide proofs that were missing in some of the original contributions due to space reasons, and we also correct typos and minor errors of the contributions that were spotted after publication.

Before kicking off with more precise, technical content, we provide several guidelines to different kinds of readers. As reading a whole PhD dissertation is a rather uncommon phenomenon among researchers, we start with an obvious but important advice: each paper listed in the main contributions of this thesis (Chapter 4) can be read independently. They are technically self-contained and they are also independently motivated.

The second advice is directed to readers coming from more informal approaches to either epistemic attitudes (e.g., non-formal epistemology) or argumentation (e.g., general argumentation theory), or from any of the two traditions exclusively (EL or FA), who are however attracted by the topic of this dissertation. If you are unfamiliar with either EL or FA, you can use Chapter 2, as a brief introduction to both fields, with a special focus on the treatment given here, and you can also use it as a guide to further reading.

As to readers looking for a general perspective and a brief summary of the topics investigated through this dissertation (probably as a tip for deciding if going on with the reading), they are referred to the precedent content of the current chapter as well as to Chapter 3. Moreover, an extended-abstract length presentation containing an overview of the thesis can be found in (Yuste-Ginel, 2021).

Finally, for readers interested in a rather detailed but selective survey on existing works dealing with the combination of EL and FA, we refer them to Chapter 5. However, for understanding its comparison with the contributions of this dissertation, the previous reading of Chapter 4 is needed, as we avoid tedious repetition of some of the technical details.

Chapter 2

Technical tools

Besides some degree of mathematical maturity, each of the contributions that form the core of this thesis (those reprinted in Chapter 4) is technically self-contained. The aim of this chapter is to briefly introduce and discuss the mathematical tools used through all of them, focusing on common features and on the way we have treated them. We start by recalling again that the methodological approach adopted in each publication emerges from adopting a hybrid perspective that connects techniques and ideas borrowed from two relatively disconnected research topics: *epistemic logic* (Meyer and van der Hoek, 1995; Fagin et al., 2004) –and its further dynamic extensions (van Ditmarsch et al., 2007; van Benthem, 2011)– on the one side; and *formal argumentation* (Baroni et al., 2018b; Gabbay et al., 2021) on the other side. The notion of *awareness*, as initially introduced by Fagin and Halpern (1987), provides a natural conceptual bridge among both families of tools throughout most of the works.

The main semantic construct used in this thesis are directed (multi-)graphs. A *directed graph* (or digraph, for short) is just an ordered pair (S, R) where S is a set, and $\mathsf{R} \subseteq \mathsf{S} \times \mathsf{S}$ is a binary relation on S. A *directed multi-graph* is obtained by changing the relation R by a finite set of them $\{\mathsf{R}_1, ..., \mathsf{R}_n\}$. Directed (multi-)graphs are widely used in theoretical computer science, and their modelling power is also well-known within the field of formal philosophy (Hansson and Hendricks, 2019). As for formal languages, *modal logic* (Blackburn et al., 2002) is one of the most celebrated candidates to reason about directed (multi-)graphs. This dissertation is built upon two informal interpretations of directed (multi-)graphs: *epistemic structures* (also known as Kripke frames) (Fagin et al., 2004; Meyer and van der Hoek, 1995; van Ditmarsch et al., 2007) and *argumentation frameworks* (Dung, 1995).

2.1 Epistemic modal logic and its dynamic extensions

2.1.1 Basic epistemic logic

When interpreted epistemically,[1] directed graphs are usually called *Kripke frames*, as they were first proposed by Kripke (1959) as a conceptually clear semantic for modal logic. Under its epistemic interpretation, S is understood as a set of *states*, *situations*, or *possible worlds* (and usually noted W instead of S). In its multi-agent version, Kripke frames are directed multi-graphs in which each relation (noted $\mathcal{R}_i$ instead of R_i) is indexed with an agent i, and it is informally interpreted as a *epistemic/doxastic* accessibility relation. More precisely, whenever $w\mathcal{R}_i w'$ in a Kripke frame, this informally means that agent i, situated at w, consider w' as a candidate for the actual state of affairs. For the rest of this dissertation Ag denotes a finite non-empty set of *agents* (whose elements are named $i, j, k, ...$ or $1, 2, 3, ...$), and At denotes a denumerable set of *atoms*, also called *atomic propositions* or *propositional variables* (usually named $p, q, r, ...$). In an epistemic reading, we are not only interested in frames, but also in models, as we need a richer description of states to express what are the beliefs/knowledge of the involved agents. A model extends a frame with a *propositional valuation*, that is, a function used to identify which atomic propositions hold at each world.

Definition 1 (Multi-agent Kripke frames and models)**.** *A* multi-agent Kripke frame *(*frame*, for short) is a pair* $\mathcal{K} = (\mathsf{W}, \mathcal{R})$ *where* $\mathsf{W} \neq \varnothing$*, and* $\mathcal{R} : \mathsf{Ag} \to \wp(\mathsf{W} \times \mathsf{W})$ *(we abbreviate* $\mathcal{R}(i)$ *as* $\mathcal{R}_i$ *and adopt the infix notation for denoting elements of* $\mathcal{R}_i$*). A* multi-agent Kripke model *(*epistemic model*, or just* model*, for short) is a triple* $M = (\mathsf{W}, \mathcal{R}, \mathsf{V})$ *where* $(\mathsf{W}, \mathcal{R})$ *is a frame and* V *is a propositional valuation, that is, a function* $\mathsf{V} : \mathsf{At} \to \wp(\mathsf{W})$*. A* pointed model*, is a pair* $(M, w) = ((\mathsf{W}, \mathcal{R}, \mathsf{V}), w)$*, where* $w \in \mathsf{W}$ *is a world intended to represent the actual state of affairs.*

Informally, $\mathsf{V}(p)$ represents the set of worlds where p holds. Multi-agent frames and models are omnipresent in the contributions that constitute this thesis, usually as the skeleton of richer epistemic structures. They are moreover described using the following multi-modal language:

Definition 2 (Multi-agent modal language)**.** *The language* $\mathcal{L}_{\Box}(\mathsf{Ag}, \mathsf{At})$ *is given by the following BNF*

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Box_i\varphi \qquad p \in \mathsf{At}, i \in \mathsf{Ag}.$$

The rest of Boolean connectors ($\vee$, $\to$, $\leftrightarrow$), and constants ($\top$, $\bot$) are defined as usual. Formulas $\Box_i\varphi$ are to be read (depending on the context) as "agent i believes/knows that φ". The *dual* of $\Box_i$, noted $\Diamond_i$, is defined as $\neg\Box_i\neg\varphi$ and is to be read as "agent i considers (epistemically/doxastically) possible that...". When assumed to stand for knowledge (resp. belief), $\Box_i$ is sometimes replaced by K_i (resp. B_i).

[1] In this chapter, we tend to use *epistemic* (and derived forms) in its broad sense, referring to both belief and knowledge. However, we sometimes use *epistemic/doxastic* to differentiate between belief and knowledge. The context will make each use clear enough.

Definition 3 (Truth, validity and consequence)**.** *Formulas of the multi-agent epistemic language are interpreted in pointed models recursively as follows:*

$$\begin{array}{rll} M, w \models p & \textit{iff} & w \in \mathsf{V}(p) \\ M, w \models \neg\varphi & \textit{iff} & M, w \not\models \varphi \\ M, w \models (\varphi \wedge \psi) & \textit{iff} & M, w \models \varphi \textit{ and } M, w \models \psi \\ M, w \models \Box_i\varphi & \textit{iff} & \textit{for all } v \in \mathsf{W}, w\mathcal{R}_i v \textit{ implies } M, v \models \varphi \end{array}$$

A formula φ is said to be valid in *M (noted $M \models \varphi$) iff it is true at every state. A formula φ is* valid *(noted $\models \varphi$) iff it is true in every pointed model. The notion of truth can be lifted to sets of formulas by setting $M, w \models \Gamma$ iff $M, w \models \varphi$ for every $\varphi \in \Gamma$. Given a set of formulas Γ and a class a of epistemic models C, we say that Γ is* C-consequence *of φ (in symbols, $\Gamma \models_{\mathsf{C}} \varphi$) iff for all pointed models pointed models (M, w) such that $M \in \mathsf{C}$ we have that $M, w \models \Gamma$ implies $M, w \models \varphi$.*

The popularity of epistemic models lies in the fact that they provide a compact representation not only of what each agent from a group knows/believes about the world, but also of what he knows/believes about the other agents, and about what the other agents know/believe of other agents, etc. In other words, epistemic models enable a simple representation of arbitrarily higher-order epistemic attitudes. Let us illustrate this idea with a simple example.

Example 3. *Figure 2.1 depicts an epistemic model for agents 1 and 2. Note that, at w_0, agent 1 knows that p is the case, while she does not consider epistemically possible that q, while the opposite happens for agent 2, in symbols $M, w_0 \models \Box_1 p \wedge \neg\Diamond_1 q \wedge \neg\Box_2 p \wedge \Diamond_2 q$. Moreover, note that the knowledge of 1 is not only accurate about the relevant objective facts (the atomic propositions p and q), but also about 2's epistemic attitudes, that is $M, w_0 \models \Box_1(\neg\Box_2 p \wedge \Diamond_2 q)$.*

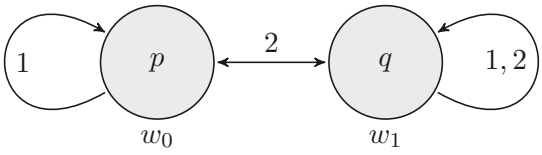


Figure 2.1: An epistemic model

On the properties of knowledge and belief. The debate on what are the appropriate formal properties that we should attribute to knowledge and belief is as old as the own field of epistemic logic, whose birth is typically situated in the works of Von Wright (1953) and Hintikka (1962). Semantically, the problem consists in deciding which constraints (if any) should satisfy each $\mathcal{R}_i$ in epistemic (doxastic) models. Axiomatically, this amounts to choosing a suitable set of axioms for capturing the notions of knowledge and belief.

Table 2.1 contains the foremost candidates for being principles of belief and knowledge. Although there is a noticeable lack of consensus in the literature, it seems fair to say that computer scientists tend to accept $S5$ (formed by axioms K, T, 4 and 5)[2] as *the* logic of knowledge, and $KD45$ (formed by axioms K, D, 4, and 5) as the one of belief (Fagin et al., 2004; Meyer and van der Hoek, 1995; van Ditmarsch et al., 2007). Philosophers, on their side, have questioned the negative introspection principle ($\neg\Box_i\varphi \to \Box_i\neg\Box_i\varphi$) as being a too strong formal property for knowledge (starting with Hintikka (1962) himself), and some of them have argued for weaker logics (e.g., $S4.1$ or $S4.2$) as more appropriate for capturing the notion of knowledge. The debate is extremely more intricate nowadays but it is somehow orthogonal to the questions that this dissertation deals with, as we have always tried to make our contributions independent from it. We refer to (Rendsvig and Symons, 2021, Section 2.6) and (Aucher, 2014) as introductions to the topic for the interested reader.

	Informal property	Axiom scheme	Property of $\mathcal{R}_i$
K	Distribution	$\Box_i(\varphi \to \psi) \to (\Box_i\varphi \to \Box_i\psi)$	
D	Consistency	$\neg\Box_i \bot$	Seriality
T	Factivity	$\Box_i\varphi \to \varphi$	Reflexivity
4	Positive introspection	$\Box_i\varphi \to \Box_i\Box_i\varphi$	Transitivity
5	Negative introspection	$\neg\Box_i\varphi \to \Box_i\neg\Box_i\varphi$	Euclideanity

Table 2.1: Candidates for formal properties of knowledge and belief

Canonical model technique and its transformations. Many of the technical results of the contributions that make up this dissertation consists in providing sound and complete axiomatizations for certain classes of epistemic models combined with argumentative structures (e.g., results in papers Paper I, Paper II, Paper III, and Paper VI). All our proofs are based on the standard technique of building *canonical models* for the targeted logics (Blackburn et al., 2002, Chpt. 4.2). However, in some cases our canonical construction required modifying the model to fit non-definable properties (Paper II and Paper VI). In particular, we had to take generated sub-models of the canonical model in Paper II, and use more subtle transformation functions in Paper VI.

2.1.2 Syntactic awareness logic

Epistemic logic has faced numerous objections as an adequate formalism to model human knowledge and belief (see (Solaki, 2021) for an overview of these criticisms), *logical omniscience* being probably the most discussed one (see e.g., (Fagin et al., 2004, Chapter 9) or (Stalnaker, 1991; Halpern and Pucella, 2011)).

Fact 1 (Logical omniscience)**.** Agents captured by epistemic models are logical omniscient: they know *all* the logical consequences of their own knowledge. *More formally,*

[2] Although this is the usual presentation, so as to make explicit all properties of knowledge, we recall that axiom 4 is redundant, that is $KT45 = KT5$.

for any class of models C*, any pointed model* (M, w) *with* $M \in \mathsf{C}$*, and any set of formulas* Γ *it holds that: if* $\Gamma \models_{\mathsf{C}} \varphi$*, then* $M, w \models \Box_i \psi$ *for every* $\psi \in \Gamma$ *implies* $M, w \models \Box_i \varphi$*.*

This is, of course, unrealistic. Logical omniscience implies that any agent captured by these models knows an infinite amount of sentences and she is, moreover, a perfect reasoner. Among the heterogeneous methods to solve this shortcoming (which was by the way already noticed by Hintikka (1962)), the *awareness approach*, initiated by Fagin and Halpern (1987), represents an important tradition in the literature, attracting the attention not only of computer scientists and philosophers, but also economists (see Schipper (2015) for a survey on formal models of awareness, and http://www.econ.ucdavis.edu/faculty/schipper/unaw.htm for a bibliography). We follow Fagin and Halpern (1987) in our presentation, as their *logic of general awareness* has been shown to be a very general approach to the notion, in which other options for representing awareness formally can be embedded (e.g., (Halpern, 2001; Halpern and Rêgo, 2008; Lorini and Song, 2020)).

Conceptually speaking, awareness logic avoids the problem of logical omniscience by differentiating between *implicit* and *explicit* knowledge, as previously suggested by Levesque (1984). More precisely, explicit knowledge that φ is defined as implicit knowledge that φ (the normal modal notion captured by epistemic models) plus awareness of φ. Awareness is formally modelled as a new component in epistemic models that adds a syntactic flavour to the picture, functioning as a 'filter' for computing explicit knowledge. Let us first introduce the formal language for awareness epistemic logic:

Definition 4 (Multi-agent awareness language)**.** *The language* $\mathcal{L}^A_{\Box}(\mathsf{Ag}, \mathsf{At})$ *is the one generated by the following BNF*

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Box_i \varphi \mid A_i \varphi \qquad p \in \mathsf{At}, i \in \mathsf{Ag},$$

where $A_i\varphi$ reads "agent i is aware of φ".

Combining the *implicit* epistemic operator $\Box_i$ and the awareness operator A_i, we can give the definition of explicit knowledge we are after[3]

$$\Box^e_i \varphi = \Box_i \varphi \wedge A_i \varphi.$$

We can now define the models where we interpret the new language.[4]

Definition 5 (Awareness epistemic models)**.** *An* awareness epistemic model *is a tuple* $M = (\mathsf{W}, \mathcal{R}, \mathsf{Aw}, \mathsf{V})$ *where* $(\mathsf{W}, \mathcal{R}, \mathsf{V})$ *is a multi-agent Kripke model (Definition 1) and* $\mathsf{Aw} : (\mathsf{Ag} \times \mathsf{W}) \to \wp(\mathcal{L}^A_{\Box}(\mathsf{Ag}, \mathsf{At}))$ *is an* awareness function *(we abbreviate* $\mathsf{Aw}(i, w)$ *as* $\mathsf{Aw}_i(w)$*).*

Informally, $\varphi \in \mathsf{Aw}_i(w)$ means that the agent is aware of φ at world w. So, we set the new truth clause as follows:

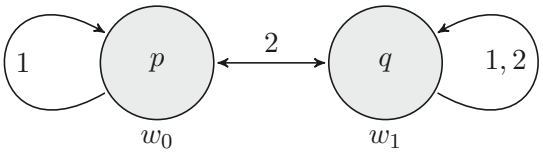
[3] The same analysis serves for differentiating between implicit and explicit beliefs.

[4] Note that the order of presentation of syntax and semantics is not optional now, since, as mentioned, awareness models are not syntax-free: they need the previous definition of the logical language to be well-defined.

$$M, w \models A_i\varphi \text{ iff } \varphi \in \mathsf{Aw}_i(w).$$

Several suitable readings have been proposed for interpreting formulas $A_i\varphi$ more precisely. For instance, but inexhaustibly, awareness can be understood *psychologically* (as an attentional skill), *computationally* (an agent is aware of φ iff she is able to compute the truth value of φ given some bounded resources[5]) or *linguistically* (an agent is aware of φ iff it belongs to her language).

Example 4. *Figure 2.2 represents an awareness epistemic model for agents* 1 *and* 2*. Note that, for instance,* $M, w_0 \models \Box_1^e p \wedge \neg\Box_1^e(p \vee q)$ *because agent* 1 *is not aware of* $p \vee q$*. This simple example shows how awareness blocks the problem of logical omniscience (since* $\{p\} \models p \vee q$*).*



$\mathsf{Aw}_i(w_0) = \mathsf{Aw}_i(w_1) = \{p, q\}$ with $i \in \{1,2\}$

Figure 2.2: An awareness epistemic model

Properties of awareness. General awareness epistemic models do not assume any restriction on awareness functions. Depending on the targeted informal notion or on the application that one has in mind, different properties can be arguably assumed. Table 2.2 shows some of the most discussed properties, where CSub-F abbreviates *closure under subformulas*; CNeg abbreviates *closure under negations*, CCon abbreviates *closure under conjunctions*; ARef abbreviates *awareness reflexivity*; and FIAw abbreviates *full awareness introspection*. See (Fagin and Halpern, 1987; Schipper, 2015) for more examples and discussion, and (Burrieza and Yuste-Ginel, 2021) for completeness and decidability theory.

The role of awareness in this thesis. The notion of awareness plays a central role in the contributions of this thesis, by serving as a bridge between EL and FA. The only exception is Paper IV, where we explore an alternative way of connecting both families of formalisms. As to the rest of contributions, the main conceptual switch, with respect to the literature, is to lift the range of awareness functions from the set of all formulas to set of all *arguments* (where the precise, formal meaning of this term is not homogeneous

[5] As suggested by Konolige (1986), this reading would better be understood as "an agent is aware of φ iff she is able to compute (within time T) whether φ follows from her set of initial premisses".

CSub-F	$\varphi \in \mathsf{Aw}_i(w)$ implies $\mathsf{sub}_F(\varphi) \subseteq \mathsf{Aw}_i(w)$	$A_i\neg\varphi \to A_i\varphi$ $A_i\Box_j\varphi \to A_i\varphi$ $A_iA_j\varphi \to A_i\varphi$ $A_i(\varphi \wedge \psi) \to (A_i\varphi \wedge A_i\psi)$
CNeg	$\varphi \in \mathsf{Aw}_i(w)$ iff $\neg\varphi \in \mathsf{Aw}_i(w)$	$A_i\varphi \leftrightarrow A_i\neg\varphi$
CCon	$\varphi \wedge \psi \in \mathsf{Aw}_i(w)$ iff $\varphi, \psi \in \mathsf{Aw}_i(w)$	$A_i(\varphi \wedge \psi) \leftrightarrow (A_i\varphi \wedge A_i\psi)$
ARef	$\varphi \in \mathsf{Aw}_i(w)$ implies $A_i\varphi \in \mathsf{Aw}_i(w)$	$A_i\varphi \to A_iA_i\varphi$
FIAw	$\varphi \in \mathsf{Aw}_i(w)$ and $w\mathcal{R}_iv$ implies $\mathsf{Aw}_i(w)\mathsf{Aw}_i(v)$	$A_i\varphi \to \Box_i\varphi$ $\neg A_i\varphi \to \Box_i\neg A_i\varphi$

Table 2.2: Usual studied restriction for awareness sets

through the different works). By letting agents be aware of arguments, we gain the power for modelling at least two interesting phenomena: *argument-based beliefs*, understood as a special case of *justified belief* (I explicitly believe that φ iff I have a strong enough argument concluding φ, just as in Example 2) and strategic reasoning (believing that my opponent is aware of certain arguments conditions what would I say in the next move of a dialogue, just as in Example 1).

2.1.3 Dynamic epistemic logic

Dynamic epistemic logic (DEL) is a logical approach to changes of information (van Ditmarsch et al., 2007; van Benthem, 2011), where the term *information* is precisely understood as the knowledge and beliefs that an agent (or a group of them) possesses. It is a well-established research topic, this being witnessed by the dedicated handbooks we have just quoted, but also by the monographic entry in the *Stanford Encyclopaedia of Philosophy* (Baltag and Renne, 2016). For a paper-length, critical introduction to the field, the reader is referred to Herzig (2017). Besides pointing out the seminal work of Plaza (1989), Gerbrandy and Groeneveld (1997), and Baltag et al. (1998), we omit any further historical reference. Our presentation will focus on the most general dynamic tool used in this dissertation (also known to be a very general approach to DEL), namely, *event models* (Baltag et al., 1998; Baltag and Moss, 2004) in their enriched version with *propositional change* (van Benthem et al., 2006; van Ditmarsch and Kooi, 2008).

From a technical perspective, the main idea of DEL consists in studying transformations of epistemic models, and how these transformations can in turn be described through the extension of the epistemic language with so-called *dynamic operators* (which are also modal operators in essence). Therefore, in the language of dynamic epistemic logic, we shall have expressions like

$$[\mathsf{action}]\Box_1\langle\mathsf{event}\rangle(p \vee q)$$

expressing that after all executions of action, agent 1 knows that there is a possible execution of event that makes either p or q true.

A preliminary notion, before defining event models with propositional change, is that of an *atomic substitution*.

Definition 6 (Substitutions)**.** *A* substitution *(or* assignment*) is a function* $\sigma : \mathsf{At} \to \{\bot, \top\}$*, where the number of elements in the domain of* σ *such that* $\sigma(p) \neq p$ *is assumed to be finite.*[6] *We use* SUB *to denote the set of all substitutions, and* λ *to denote the identity substitution.*

Substitutions can be understood as reassignments of atomic variables to either Truth or Falsity (or to its names, to be more precise). They are used as a meta-syntactic device to capture propositional change within event models. Curiously enough, the semantic structure underlying event models is, just as in the case of epistemic models (and of the forthcoming notion of argumentation framework), a directed multi-graph.

Definition 7 (Event model with propositional change)**.** *An* event model *for* $\mathcal{L}_{\Box}(\mathsf{Ag}, \mathsf{At})$ *is a tuple* $\mathcal{E} = (\mathsf{S}, \mathcal{T}, \mathsf{pre}, \mathsf{pos})$ *where:*

- S *is a finite set of* events*.*
- $\mathcal{T} : \mathsf{Ag} \to \wp(\mathsf{S} \times \mathsf{S})$*, where each* $\mathcal{T}(i) \subseteq \mathsf{S} \times \mathsf{S}$*, abbreviated* $\mathcal{T}_i$*, represents an* accessibility relation *for* i*, indicating how* i *perceives the effects of each event.*
- $\mathsf{pre} : \mathsf{S} \to \mathcal{L}_{\Box}(\mathsf{Ag}, \mathsf{At})$ *indicates the* precondition *of each event (when this is executable).*
- $\mathsf{pos} : \mathsf{S} \to \mathsf{SUB}$ *indicates the* postconditions *of each event, i.e., what are its effects on the semantic status of propositional variables.*

We note EVENT *the class of all event models. Let* $\mathcal{E}$ *be an event model, we note* $\mathcal{E}[S]$ *its set of events.*

The following definition makes formally clear what does it mean for an event to be executed in an epistemic model.

Definition 8 (Product update)**.** *Let* $M = (\mathsf{W}, \mathcal{R}, \mathsf{V})$ *be a model, let* $\mathcal{E} = (\mathsf{S}, \mathcal{T}, \mathsf{pre}, \mathsf{pos})$ *be an event model, and define* $M \otimes \mathcal{E} = (\mathsf{W}', \mathcal{R}', \mathsf{V}')$*, where:*

- $\mathsf{W}' = \{(w, s) \in \mathsf{W} \times \mathsf{S} \mid M, w \models \mathsf{pre}(s)\}$*;*
- *for every* $i \in \mathsf{Ag}$*,* $(w, s)\mathcal{R}'_i(v, t)$ *iff* $w\mathcal{R}_i v$ *and* $s\mathcal{T}_i t$*; and*
- $\mathsf{V}'(p) = \{(w, s) \in \mathsf{W}' \mid M, w \models \mathsf{pos}(s)(p)\}$.

Although event models are only used in one of the contributions of this thesis (Paper II), the rest of dynamic devices used in other contributions can be understood as special cases of these. For instance, the event known as the *public announcement of* φ, firstly introduced by Plaza (1989) and strongly used in Paper IV and Paper VI, is captured by the following event model:

[6] Alternatively, we could define σ as a partial function and then extend its domain by using the identity function for non-defined values. Moreover, and as it was shown by van Ditmarsch and Kooi (2008), the restriction of the co-domain of substitutions to $\{\top, \bot\}$ will not impose any modelling limitation. In other words, our choice it is equivalent to using partial functions $\sigma : \mathsf{At} \to \mathcal{L}_{\Box}(\mathsf{Ag}, \mathsf{At})$ with finite domain.

Example 5 (Public announcement)**.** *The event model for publicly announcing a formula φ is the tuple $\mathsf{Pub}^{\varphi} = (\mathsf{S}, \mathcal{T}, \mathsf{pre}, \mathsf{pos})$, where $\mathsf{S} = \{\triangle\}$, $\mathcal{T}_i = \{(\triangle, \triangle)\}$ for every $i \in \mathsf{Ag}$, $\mathsf{pre}(\triangle) = \varphi$, and $\mathsf{pos}(\triangle) = \lambda$ (see a graphical representation of Pub^{φ} for the case $\mathsf{Ag} = \{1, 2\}$ in the left-side part of Figure 2.3).*

Moreover, operations on epistemic models incorporating argumentative tools (such as the notion of *argument disclosure* of Paper I) can also be captured by refined forms of event models. We shall come back to this point in Chapter 5. Let us now look at a slightly more complicated example of event model:

Example 6 (Privately observing that p becomes true)**.** *The event model for agent 1 privately observing that p becomes true with $\mathsf{Ag} = \{1, 2\}$ is the tuple $\mathsf{Pri}_1^p = (\mathsf{S}, \mathcal{T}, \mathsf{pre}, \mathsf{pos})$ where $\mathsf{S} = \{\bullet, \circ\}$, $\mathcal{T}_1 = \{(\bullet, \bullet), (\circ, \circ)\}$ and $\mathcal{T}_2 = \{(\bullet, \circ), (\circ, \circ)\}$, $\mathsf{pre}(\bullet) = \mathsf{pre}(\circ) = \top$ and $\mathsf{pos}(\bullet) = \{p \mapsto \top\}$ and $\mathsf{pos}(\circ) = \lambda$ (see a graphical representation of Pri_1^p in the left-side part of Figure 2.3).*

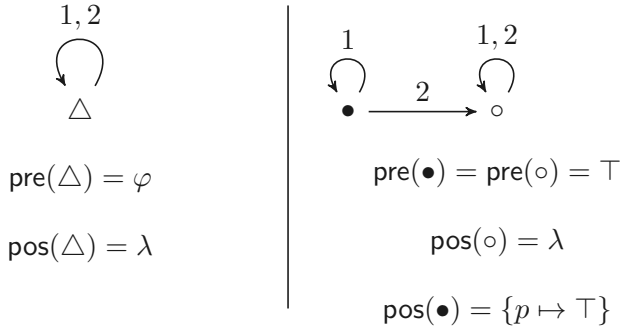


Figure 2.3: Representation of Pub^{φ} (left.) and Pri_1^{+p} (right.) for two agents $\{1, 2\}$

The effects of executing event models in epistemic models can be described through the use of the following dynamic languages:

Definition 9 (Dynamic language)**.** *Let $\star \subseteq \mathsf{EVENT}$, the dynamic language $\mathcal{L}^{\star}_{\Box}(\mathsf{At}, \mathsf{Ag})$ is given by the following BNF*

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Box_i\varphi \mid [\mathcal{E}, s]\varphi \qquad p \in \mathsf{At}, i \in \mathsf{Ag}, \mathcal{E} \in \star, s \in \mathcal{E}[S].$$

The notion of truth is extended to dynamic operators as follows:

$$M, w \models [\mathcal{E}, s]\varphi \text{ iff } M, w \models \mathsf{pre}(s) \text{ implies } (M \otimes \mathcal{E}), (w, s) \models \varphi$$

Example 7. *The bottom part of Figure 2.4 shows the execution of agent 1 privately observing that p becomes true –represented by the event model Pri_1^p at the top-right part of the figure– in the epistemic model M (top-left part of the figure). Intuitively, we could think of it as representing an initial situation (model M) where it does not rain ($\neg p$), and both Anne (1) and Bob (2) know this. After that, it begins to rain (event $\bullet$), but Anne is the only one looking through the windows, and hence realising the change in the weather. Note that $M, w \models [\mathsf{Pri}_1^p, \bullet](p \wedge \Box_1 p \wedge \Box_2 \neg p)$.*

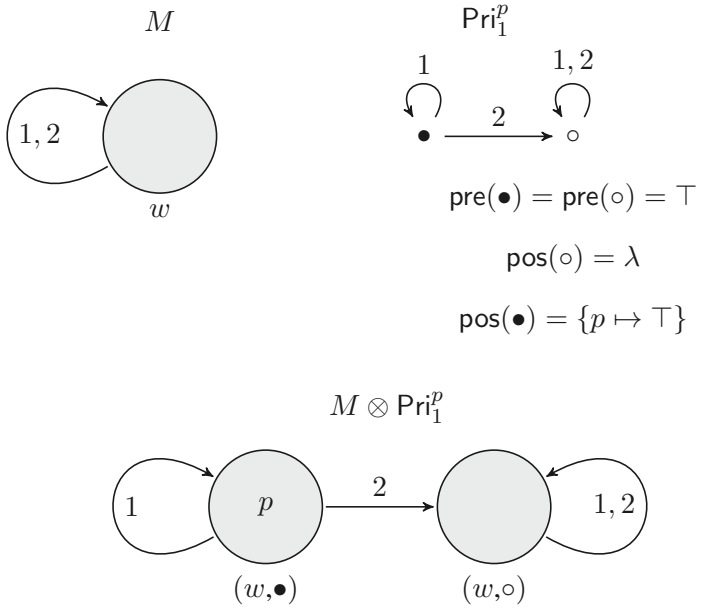


Figure 2.4: Example of event model execution

Completeness via reduction. In DEL, sound and complete axiomatisations are usually obtained through the *reduction axioms* technique (Kooi, 2007; Wang and Cao, 2013). This can be roughly described as follows. We first need to have a completeness result for the *static fragment* of the language we are working with (that is, the language without dynamic operators) with respect to the intended class of models. Second, we need to find a full set of sound *reduction axioms* that enables a recursive elimination of dynamic operators, by "pushing" them towards one type of formulas (usually, atoms), where they finally "melt". As examples of these "pushing" and "melting" axioms, consider the following ones, which are valid in the class of all epistemic models:

$$[\mathcal{E}, s]\Box_i\varphi \leftrightarrow (\mathsf{pre}(s) \rightarrow \bigwedge_{s\mathcal{T}_i t} \Box_i[\mathcal{E}, t]\varphi).$$

$$[\mathcal{E}, s]p \leftrightarrow (\mathsf{pre}(s) \rightarrow \mathsf{pos}(s)(p)).$$

Note that there is a straightforward complexity measure for formulas of $\mathcal{L}^{\star}_{\Box}(\mathsf{At}, \mathsf{Ag})$ such that the complexity of the formula(s) under the scope of $[\mathcal{E}, s]$ is smaller in the right-hand side than in the left-hand side of the above equivalences. This is crucial for guaranteeing that the reduction procedure will eventually terminate. We explicitly apply the reduction technique to obtain complete axiomatisations for the novel logics defined in Paper I, Paper II and Paper VI. Moreover, we make use of the validity of reduction axioms for public announcement logics when proving some of the results of Paper IV.

Dynamics of awareness. DEL and awareness logic have found a joint treatment in the *dynamics of awareness tradition* (Grossi and Velázquez-Quesada, 2009, 2015; van Benthem and Velázquez-Quesada, 2010; van Ditmarsch et al., 2012). The more general approach to the topic seems to be the work by van Benthem and Velázquez-Quesada (2010), where they extend the logic of general awareness of Fagin and Halpern (1987) with event models containing postconditions that modify the awareness sets of the original epistemic model. We follow this path in Paper II, but showing that awareness of atomic arguments can be reduced to atomic valuations for a special set of atomic propositions. Therefore, postconditions for modifying awareness of arguments are technically equivalent to those for propositional change (see Definition 7). This is interesting because, among other reasons, it enables transparent axiomatisations by reduction in the style of van Benthem et al. (2006), something that is still missing for the general awareness case of van Benthem and Velázquez-Quesada (2010). Moreover, we extend the approach of Grossi and Velázquez-Quesada (2009, 2015) to our structured argumentation setting in Paper VI.

On the problem of maintaining epistemic model restrictions after event model execution. One of the main conceptual pitfalls of DEL is that it does not give a clear account of belief revision (Herzig, 2017). In other words, when agents receive information that is inconsistent with their prior beliefs, they might end up believing everything, as their beliefs might become inconsistent. Technically speaking, the seriality of $\mathcal{R}_i$ in an epistemic model is not necessarily preserved by event model execution. More in general, this problem affects every restriction on the accessibility relations that involves existential quantification. Different solutions have been offered in the literature. For instance, Cao Son et al. (2015) found sufficient conditions (parametrised by M and E) for $M \otimes E$ to preserve serial relations. Moreover, Balbiani et al. (2012) provided sufficient and necessary conditions (only expressible in a language with the global or universal modality) for epistemic models to preserve doxastic relations after public announcements. This idea was previously introduced the work of Aucher (2008), where sufficient and necessary conditions for the preservation of seriality after event model execution are presented. We take advantage of these results to develop novel axiomatisations for our DELs in Paper II. The key idea is finding a formula that is satisfiable in the epistemic model (before the execution) iff the resulting model (after the execution) satisfies the target restriction (e.g., seriality). Once this is done, the preservation of the targeted property is added as a precondition of event execution, and the mentioned formula is consequently inserted within the reduction axioms for dynamic operators.

The problem of preservation of properties can be easily extrapolated to the context of awareness epistemic models, although it seems that it has not yet been the primary focus of any work on the dynamic of awareness literature. Preserving closure properties that only affect awareness sets, for instance (CSub) (see Table 2.2), looks relatively easy: it is enough to close again the updated model under the desired property. However, the properties that combine awareness and accessibility conditions are more involved. In Paper II, we found sufficient and necessary conditions that event models should satisfy in order to preserve FIAw (see Table 2.2), and related properties, for the special case of awareness

generated by atomic propositions.[7]

2.2 Formal argumentation

Formal argumentation is a mathematical approach to the study of argumentation. It has a dedicated handbook (currently formed of two volumes (Baroni et al., 2018b; Gabbay et al., 2021)), as well as several conferences and workshops (COMMA, CLAR, ICCMA or ArgStrength). It is an interdisciplinary area of research where computational, linguistic and cognitive perspectives on argumentative phenomena gather, with strong conceptual links to general argumentation theory (van Eemeren et al., 2014). Following Prakken (2017), the history of formal argumentation[8] can be understood as divided into two main branches of inquiry: the study of argument-based *inference* and the study of argument-based *dialogues* (which, by the way, seems to be connected to the distinction between *argument-as-a-product* and *argument-as-a-process* in general argumentation theory (see e.g., Reed and Walton (2003)). This section presents the main tools for modelling argument-based inference that we have used through the contributions of this thesis, although we shall also touch upon some topics related to dialogues.[9] Another common distinction that divides the literature on formal argumentation is the one between *abstract* models of argument and *structured* models of argument. We will adopt this distinction as a guideline for our presentation (just as we will do to divide the two tracks of Chapter 4). For the main intuitions underlying the tools that we are about to introduce, the reader is referred to Section 1.2.

2.2.1 Abstract argumentation frameworks

Once again, in abstract models of argumentation, the origin, nature and structure of arguments is left unspecified. The foremost of these models was introduced by Dung (1995) under the name of *argumentation frameworks*.

Definition 10 (Argumentation framework)**.** *An* argumentation framework *(AF) is a directed graph* $\mathsf{F} = (\mathsf{A}, \mathsf{R})$ *where* A *stands for a set of* arguments *and* R *stands for a* defeat *relation.*[10] *With* $B \subseteq \mathsf{A}$*, we let* $B^+ = \{y \in \mathsf{A} \mid (x, y) \in \mathsf{R}\}$ *be the* set of arguments defeated by B.

Let us illustrate the previous definition with a couple of examples.

[7] More precisely, we find such conditions for preserving FIAw formulated for awareness of atomic arguments, but it looks like they can be extrapolated to awareness of atomic propositions.

[8] Just as in the rest of the chapter, we will get by without many historical insights. The interested reader is referred to Prakken (2017) for an historical overview.

[9] Let us point out to this respect the works of Walton and Krabbe (1995); Amgoud et al. (2000) as classic references on argument-based dialogues, and the recent proposal of Barés Gómez and Fontaine (2017) for an integration of argumentation, abductive reasoning and dialogical logic.

[10] Usually, elements of R are said to represent *attacks* instead of defeats (starting with Dung (1995) himself). Nevertheless, it has been argued e.g., Modgil and Prakken (2013) that R is better understood in terms of a *defeat* relation. This distinction will be clearer later on, when we account for the structure of arguments and the nature of the dialectical relations.

Example 8. *All relevant arguments and defeats of Example 1 are depicted in Figure 2.5. a is an argument concluding that Anne stole the chocolate candy. The two possible alibis for Anne are represented by b (the alleged allergy), and c (she was at her office when the robbery took place). The two counter-alibis are represented by d (the access to her medical records) and e (the access to the security cameras).*

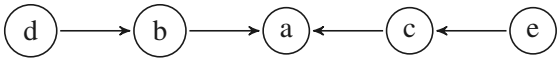


Figure 2.5: An AF for Example 1

Example 9. *All relevant arguments and defeats of Example 2 are depicted in Figure 2.6. Node a is the (enthymematic) argument concluding that it won't rain.*[11] *Node b stands for the argument constructed by Anne from the weather forecast. The fact that Anne considers b as strictly stronger than a is captured by the presence of (b, a) in R and the absence of (a, b).*

Figure 2.6: An AF for Example 2

As we exposed in Section 1.2, the notion of argument strength can be divided into three tiers or dimensions. The fundamental question regarding one of these dimensions, the *evaluative tier of argument strength*, can be now specified as follows: given a set of possibly conflicting arguments (an AF), how should a rational agent choose subsets of arguments that are justified? There exists an important variety of *argumentation semantics* nowadays, these are, precise, mathematical answers to the above-mentioned question. Moreover, there are two formal paradigms for introducing these semantics: the *extension-based* approach and the *labelling-based* approach. Although both in Paper I and in the discussion chapter, we make use of the labelling-based approach, and having into account that they have been proved to be equivalent for all the standard semantics (see e.g., (Caminada and Gabbay, 2009)), we only follow here the extension-based approach for sake of brevity. Moreover, we restrict our attention to the original Dung (1995)'s four semantics, as they are the one used through the contributions belonging to this thesis. For a much more detailed introduction to the topic, the reader is referred to Baroni et al. (2018a).

Definition 11 (AFs' semantics)**.** *Given an AF $\mathsf{F} = (\mathsf{A}, \mathsf{R})$, a set of arguments $B \subseteq \mathsf{A}$, and an argument $a \in \mathsf{A}$: we say that B* defends *a iff for every $c \in \mathsf{A}$: if $(c, a) \in \mathsf{R}$ then $c \in B^+$. Moreover, B is said to be*

[11] An *enthymeme* is an argument such that some of its components are left implicit. In our example, Anne's colleague only points out that "the sky looked cloudless" as a clear abbreviation of "the sky looked cloudless, so it won't rain".

- *a* stable extension *iff* $B \cap B^+ = \varnothing$ *(it is* conflict-free*), and* $B^+ = \mathsf{A} \setminus B$ *(it defeats every argument outside itself).*
- *a* complete extension $B \cap B^+ = \varnothing$ *(it is* conflict-free*), and for all* $b \in \mathsf{A}$*,* $b \in B$ *iff* B *defends* b *(it contains precisely the arguments that it defends).*
- *a* preferred extension *iff it is a maximal (w.r.t. set inclusion) complete extension.*
- *the* grounded extension *iff it is the smallest (w.r.t. set inclusion) complete extension.*[12]

Given a semantics $\sigma \in \{\mathsf{st}, \mathsf{co}, \mathsf{pr}, \mathsf{gr}\}$ *(standing respectively for* stable*,* complete*,* preferred *and* grounded*), and an AF* (A, R)*, we denote by* $\sigma(\mathsf{A}, \mathsf{R})$ *the set of all* σ*-extensions of* (A, R)*.*[13]

Note that every stable (resp. preferred, grounded) extension is also a complete extension. Moreover, these four semantics satisfy the so-called *admissibility principle*, meaning that any extension prescribed by them is (i) conflict-free, and (ii) self-defended (it defends all its elements). Interestingly, two families of semantics relaxing (ii) have been proposed in the literature under the name of *naivety-based* semantics (see e.g., Cramer and van der Torre (2019)), and *weak admissibility-based* semantics (Baumann et al., 2020).

The plurality of semantics for AFs, which is actually much larger nowadays, can be better understood if we have into account that they can be applied to different purposes. For instance, some semantics have been shown to capture transparently and intuitively different forms of non-monotonic inference (e.g., logic programming, default logic or defeasible logics). Moreover, when it comes to real-life reasoning, it has been argued that different semantics are more suitable for different contexts (see e.g., Prakken (2006)). While scientific reasoning (or more generally, epistemic reasoning) intuitively tends to favour more sceptic semantics (that is, semantics that sort out smaller sets of arguments, as the grounded semantics), deliberation might in contrast demand the maximization of accepted arguments (e.g., preferred semantics). Moreover, part of the formal argumentation community has recently been working on the empirical validation and comparison of existing semantics, through cognitive studies performed with human subjects (see Cerutti et al. (2021) for a recent survey).

Using the extensions of an AF, we can easily define precise notions of *argument acceptance*, also called *justification status*:

Definition 12 (Coarse-grained justification status)**.** *Given a semantics* $\sigma \in \{\mathsf{st}, \mathsf{co}, \mathsf{pr}, \mathsf{gr}\}$*, an AF* (A, R)*, and an argument* $a \in \mathsf{A}$*, we say that:*

[12] As it is known since (Dung, 1995), a grounded extension always exists and it is moreover guaranteed to be unique, so we are justified to use the article *the*.

[13] The unfamiliar reader is referred to *ConArg* (Bistarelli and Santini, 2011), and more specifically to its intuitive web interface https://conarg.dmi.unipg.it/web_interface.php, for testing her understanding of Dung's semantics.

- a *is* σ-sceptically accepted *iff for every* $E \in \sigma(\mathsf{A}, \mathsf{R})$, $a \in E$.[14]
- a *is* σ-credulously accepted *iff there is an* $E \in \sigma(\mathsf{A}, \mathsf{R})$ *such that* $a \in E$.

Example 10. *In the AF of Figure 2.5 argument* a *is both credulously and sceptically accepted with respect to all the four studied semantics. Similarly, argument* a *of Figure 2.6 is nor credulously nor sceptically accepted with respect to any of the studied semantics.*

The notions of justification status just discussed are somehow too extreme, since an argument is always either accepted or not. In order to account for more subtle scenarios, Wu and Caminada (2010) proposed a fine-grained counterpart of the notion of justification status. They did so for complete semantics through a labelling-based approach. For the sake of generality and uniformity, we reproduce the more general definition of Baroni et al. (2018a), given in extension-based terms.

Definition 13 (Fine-grained justification status)**.** *Let* (A, R) *be an AF, let* $a \in \mathsf{A}$*, let* σ *be a semantics, then* a *is always in one of the following (mutually exclusive)* justification status*:*

- $\sigma(\mathsf{A}, \mathsf{R}) \neq \varnothing$ *and* $\forall E \in \sigma(\mathsf{A}, \mathsf{R}), a \in E$ (strong acceptance (SA)).
- $\exists E \in \sigma(\mathsf{A}, \mathsf{R}) : a \in E$, $\exists E \in \sigma(\mathsf{A}, \mathsf{R}) : a \notin (E \cup E^+)$, $\nexists E \in \sigma(\mathsf{A}, \mathsf{R}) : a \in E^+$ (weak acceptance (WA)).
- $\sigma(\mathsf{A}, \mathsf{R}) \neq \varnothing$ *and* $\forall E \in \sigma(\mathsf{A}, \mathsf{R}), a \notin (E \cup E^+)$ (completely undecided (CU)).
- $\sigma(\mathsf{A}, \mathsf{R}) = \varnothing$ (borderline because no extension exists (NE)).
- $\exists E \in \sigma(\mathsf{A}, \mathsf{R}) : a \in E$, $\exists E \in \sigma(\mathsf{A}, \mathsf{R}) : a \in E^+$, $\nexists E \in \sigma(\mathsf{A}, \mathsf{R}) : a \notin (E \cup E^+)$ (in-out, borderline (IO)).
- $\exists E \in \sigma(\mathsf{A}, \mathsf{R}) : a \in E$, $\exists E \in \sigma(\mathsf{A}, \mathsf{R}) : a \in E^+$, $\exists E \in \sigma(\mathsf{A}, \mathsf{R}) : a \notin (E \cup E^+)$ (in-out-undec, borderline (IOU)).
- $\nexists E \in \sigma(\mathsf{A}, \mathsf{R}) : a \in E$, $\exists E \in \sigma(\mathsf{A}, \mathsf{R}) : a \in E^+$, $\exists E \in \sigma(\mathsf{A}, \mathsf{R}) : a \notin (E \cup E^+)$ (weak rejection (WR)).
- $\sigma(\mathsf{A}, \mathsf{R}) \neq \varnothing$ *and* $\forall E \in \sigma(\mathsf{A}, \mathsf{R}), a \in E^+$ (strong rejection (SR)).

Note that, for some semantics, some of the defined justification statuses are not possible. For instance, as the grounded semantics always exists, the status NE ('borderline because no extension exits') is never reached by an argument.

[14] For semantics that do not guarantee the existence of an extension (e.g., stable semantics), this notion is sometimes restricted to the class of AFs with a non-empty set of extensions (e.g., Baroni and Giacomin (2007)), as otherwise sceptical acceptance gets trivialised. We don't make the assumption formal here for the sake of simplicity.

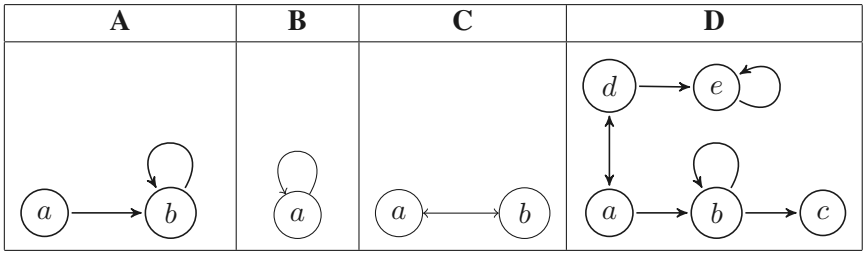


Table 2.3: Argumentation frameworks

	Stable	Complete	Grounded	Preferred
A				
a	SA	SA	SA	SA
b	SR	SR	SR	SR
B				
a	NE	CU	CU	CU
C				
a	IO	IOU	CU	IO
b	IO	IOU	CU	IO
D				
a	NE	IOU	CU	IO
b	NE	WR	CU	WR
c	NE	WA	CU	IU
d	NE	IOU	CU	IO
e	NE	WR	CU	WR

Table 2.4: Fine-grained justification status of arguments depicted in Table 2.3

Example 11 (Fine-grained JS)**.** *In the AF of Figure 2.5, arguments a, d and e are strongly accepted while arguments b and c are strongly rejected with respect to the four studied semantics. A richer example, borrowed from Dvořák (2012), is depicted in figures of Table 2.3. The figures represent four AFs (**A**, **B**, **C** and **D**) for which we list, in Table 2.4, the fine-grained justification status of each argument.*

Semantics used in this thesis. Let us briefly recall the semantics used in each contribution of this thesis. Complete semantics and fine-grained justification status are used in Paper I. In Paper II, we continue dealing with fine-grained acceptance but move instead to preferred semantics. Stable semantics is the one used in Paper IV and Paper III, not as an essential choice, but only because it is the one more transparently captured in propositional logic. In both works (Paper III and Paper IV), we focus on coarse-grained acceptability. Grounded semantics is chosen in Paper V and Paper VI, as it has been argued to be the most suitable one for epistemic reasoning (Caminada, 2006; Prakken, 2006), and this is what we aim to formalize there.

For the rest of this section, we briefly introduce some topics that were already present in the literature on abstract argumentation and that have been on the focus of the contributions that constitute the first track of this thesis. We do so in order to provide the reader with the main ideas and some references to the literature. We will come back in detail to all these topics when jointly discussing the results of the contributions and comparing it to related work (Chapter 5).

Abstraction and modelling power. AFs have won enormous popularity since their publication due, among other things, to their applicability and simplicity. They reduce a complex, natural phenomenon –argumentation– to an extremely simple and controllable mathematical structure –a directed graph. However, they are too abstract for some purposes. Consequently, AFs have been extended in multiple directions aimed, among other things, at incorporating further dialectical relations (e.g., supports (Cayrol and Lagasquie-Schiex, 2005), or recursive attacks (Baroni et al., 2009)), at handling preferences (Amgoud and Vesic, 2011), or at taking into account uncertainty about the existence of arguments and attacks (Baumeister et al., 2021).[15] Two of the essential limitations that AFs come equipped with are their lack of multi-agency and their static character. Both aspects seem indeed relevant, as many scenarios where we find real-life argumentation are irreducibly multi-agent and dynamic (think, for instance, of a debate). Hence, it is not surprising to find two traditions within the literature on AFs that try to overcome these limitations.

Bringing agents into the picture. The question of how to properly include the notion of *partial information* that each agent from a given group has with respect to an AF is subject to debate. There are at least two relevant sub-questions to answer regarding this matter. First, *which kind of epistemic attitude holds between agents and arguments*? As potential candidates for an answer we can think of *knowledge* (agent i knows argument a or agent i knows that a attacks b), *awareness* (agent i is aware of argument a), *familiarity* (agent i is familiar with argument a), or even some sort of *availability* or *ability to use* (e.g., agent i can use a as a counterargument to b during a debate). Second, *what is the primary content of this attitude*? Roughly speaking, we can choose either arguments or attacks/defeats as the object(s) of knowledge of agents (or any other epistemic attitude). It seems fair to say that the most popular option has been the former, with few exceptions focusing on the latter (including the work by Dyrkolbotn and Pedersen (2016) and our Paper IV). The formal way of including multi-agency in AFs strongly depends on the answers that we have in mind for the two questions above. For instance, a natural further assumption, when supposing that arguments are the primary (if not unique) source of partiality for agents in AFs consists in assuming that each agent has sound and complete knowledge of attacks modulo the knowledge of the involved arguments (this is what we call SCAA in paper Paper II). More formally, a multi-agent AF assuming SCAA is a triple $(\mathsf{A}, \mathsf{R}, \{\mathsf{A}_i\}_{i\in\mathsf{Ag}})$, where (A, R) is an AF, each $\mathsf{A}_i \subseteq \mathsf{A}$ represents the set of arguments that agent i knows (or is aware of, is familiar with, etc), and where the set of attacks

[15] Some of these formalisms have been recently criticised for performing *ad-hoc* modelling, where a structured argumentation approach would provide a more natural picture (Prakken and Winter, 2018).

that i knows, noted R_i, is defined as $\mathsf{R} \cap (\mathsf{A}_i \times \mathsf{A}_i)$. We give a (far from being exhaustive) list of works assuming SCAA in order to give evidence of the alleged popularity of this modelling choice (Rienstra et al., 2013; Thimm, 2014; Black et al., 2017; Sakama, 2012; Caminada and Sakama, 2017; Proietti, 2017; Schwarzentruber et al., 2012; Doutre et al., 2017; Rahwan and Larson, 2009) . For a detailed discussion on further assumptions when defining the notion of *multi-agent argumentation framework*, we refer the reader to Section 9 of Paper II. For a more general perspective, covering works on the interaction between argumentation and multi-agent systems, we recommend looking at the survey by Carrera and Iglesias (2015).

Making AFs dynamic. The second essential limitation we mentioned is the static character of AFs. There is a well-established tradition focusing on how this limitation can be meaningfully overcome. Some (but far from being all) of these approaches include: the study of adding and removing an argument or an attack (also called *elementary changes*) (Cayrol et al., 2010), ways of *enforcing* (making accepted) a set of arguments (Baumann and Brewka, 2010; Baumann, 2012; Baumann et al., 2021), as well as its relation to belief change (Booth et al., 2013), and the encodings of different types of changes in different logical languages (e.g., (Doutre et al., 2014; de Saint-Cyr et al., 2016)). We just point out the work by Doutre and Mailly (2018) as a recent survey on this topic. In Paper I, Paper II, and Paper IV we provide a combined approach of argumentative and epistemic dynamics, hence integrating the study of informational dynamics as developed within this tradition and DEL. Moreover, in Paper VI we extend these ideas to a system of structured argumentation.

Logical encodings of AFs. Dung's approach to argumentation has been encoded in different logical languages. There have been several motivations for these encodings: providing a systematic framework to reason about AFs and their semantics, using it as an intermediary step to develop a computational approach to Dung's theory, or plugging AFs on top of other formal settings. Just as it happens with dynamics, there is a well-established tradition of building logical theories for AFs, see (Besnard et al., 2014a) for a general approach and (Besnard et al., 2020) for a recent survey. Some of the most popular languages in this enterprise are:

- *classic propositional logic* (e.g., (Besnard and Doutre, 2004; Gabbay, 2011)) or equally expressive but more succinct languages as *quantified Boolean formulas* (QBFs) (Arieli and Caminada, 2013), or the *dynamic logic of propositional assignments* (DL-PA) (Doutre et al., 2014, 2017, 2019; Herzig and Yuste-Ginel, 2021a);
- *three-valued logic* (e.g., (Dyrkolbotn and Pedersen, 2016));
- *modal logic* (e.g., (Grossi, 2010b,a; Caminada and Gabbay, 2009));
- *first order logic* (e.g., (de Saint-Cyr et al., 2016; Caminada and Gabbay, 2009));

- and *second-order logic* (e.g., (Dvořák et al., 2012)) or *higher-order logic* (e.g., (Fuenmayor Pelaez and Steen, 2021)).

Through the contributions of the first track of this dissertation, we make use of two encodings of AFs in classic propositional logic. The first one, employed in Paper III and Paper IV, is essentially the one created by Besnard and Doutre (2004), but adapted to a multi-agent setting (as in (Doutre et al., 2017)), and taking care of the modal aspect our approach. The second one, designed by us specifically for Paper II, is a rather expressive propositional encoding, useful for theoretical purposes but not at all for implementations. We shall come back to both encodings on Chapter 5 (Section 5.1.1) for a detailed comparison and an explanation of how to replace one of them by the other.

2.2.2 Structured argumentation

For many purposes, abstract models of argument do not suffice (even if extended with multi-agent and dynamic features). For instance, they do not suffice if we want to model the *support dimension* of argument strength that we informally characterised in Section 1.2, or if we want to specify what is the set of sentences believed by an agent as the outcome of an argumentation process (so as to provide a formal account of C2). *Structured models of argumentation* represent an intermediate step of abstraction between real-life, natural language argumentation and abstract models as the one we have studied up to now. Just as in the abstract literature, there are many available options for formalizing structured argumentation, some of the most popular being *assumption-based argumentation* (ABA) (Cyras et al., 2018), *argumentation based on deductive logics* (Besnard and Hunter, 2018), *defeasible logic programming* (DeLP) (García and Simari, 2018), and ASPIC$^+$ (Modgil and Prakken, 2018). As general guides, and besides the Handbook's chapters that we have just quoted, the reader is referred to the tutorials of the monograph published in *Argument and Computation* (Besnard et al., 2014b).

The ASPIC$^+$ framework with symmetric negation

Among the approaches mentioned above, ASPIC$^+$ is a popular framework for structured argumentation developed by Modgil and Prakken (2013, 2014, 2018).[16] As some of its appealing features, one can highlight the fact that it gives freedom to users to make many designing choices when using it. This partially explains its success to simulate other approaches to non-monotonic inference. ASPIC$^+$ is the argumentative tool used in the contributions of the second track of this thesis (Paper V and Paper VI). We review here its main concepts, focusing on the special case of the framework defined with symmetric negation (as it is the one used in the mentioned contributions). For a more detailed explanation of our adaptation, the reader is referred to Section 5.2.2.

Definition 14 (Argumentation system)**.** *An* argumentation system (AS) *is a tuple* $\mathsf{AS} = (\mathcal{L}, \mathsf{Rules}, \mathfrak{n})$ *where:*

[16] We worked with the corrected version of the original paper (Modgil and Prakken, 2013), available at https://arxiv.org/abs/1804.06763.

- *$\mathcal{L}$ is a formal language closed under some sort of negation (denoted $\sim$).*

- $\mathsf{Rules} = \mathsf{Rules}_s \cup \mathsf{Rules}_d$ *where* Rules_s *(*strict inference rules*), and* Rules_d *(*defeasible inference rules*) are sets of finite sequences over $\mathcal{L}$ such that* $\mathsf{Rules}_s \cap \mathsf{Rules}_d = \varnothing$.

- $\mathfrak{n} : \mathsf{Rules}_d \to \mathcal{L}$ *is a possibly partial function, where $\mathfrak{n}(R)$ informally stands for the sentence "rule R is applicable".*

Example 12 (Anne and the weather, revisited)**.** *Let us design an AS for modelling Example 2. $\mathcal{L}$ is the language of classic propositional logic built from the set of atoms* $\mathsf{At} = \{\mathsf{CloudySky}, \mathsf{ForecastSaysRain}, \mathsf{Rain}\}$ *with the obvious intended meanings,* $\mathsf{Rules}_s = \varnothing$, $\mathsf{Rules}_d = \{(\neg\mathsf{CloudySky}, \neg\mathsf{Rain}), (\mathsf{ForecastSaysRain}, \mathsf{Rain})\}$, $\mathfrak{n} = \varnothing$.

Definition 15 (Knowledge base and argumentation theory)**.** *A* knowledge base (KB) *of an argumentation system $(\mathcal{L}, \mathsf{Rules}, \mathfrak{n})$ is a set $\mathcal{K} \subseteq \mathcal{L}$ partitioned into two subsets $\mathcal{K}_n$ (axioms) and $\mathcal{K}_p$ (ordinary premisses). Given $\mathsf{AS} = (\mathcal{L}, \mathsf{Rules}, \mathfrak{n})$, and a knowledge base of it $\mathcal{K}$, the pair $(\mathsf{AS}, \mathcal{K})$ is called an* argumentation theory *(AT).*

Example 13 (Anne and the weather, revisited)**.** *Let us design a KB for Example 2: $\mathcal{K}_n = \varnothing$, and $\mathcal{K}_p = \{\mathsf{CloudySky}, \mathsf{ForecastSaysRain}\}$.*

Definition 16 (Arguments and their conclusions)**.** *Given an argumentation theory* $\mathsf{AT} = (\mathsf{AS}, \mathcal{K})$*, an argument of* AT *is any syntactic chain α obtained by a finite number of applications of the following rules:*

- *$\langle\varphi\rangle$ is an argument whenever $\varphi \in \mathcal{K}$, with* $\mathsf{Conc}(\langle\varphi\rangle) = \varphi$.[17]

- *$\langle\alpha_1, ..., \alpha_n \twoheadrightarrow \varphi\rangle$ if $\alpha_1, ..., \alpha_n$ are arguments such that* $(\mathsf{Conc}(\alpha_1), ..., \mathsf{Conc}(\alpha_n), \varphi) \in \mathsf{Rules}_s$*. We let* $\mathsf{Conc}(\langle\alpha_1, ..., \alpha_n \twoheadrightarrow \varphi\rangle) = \varphi$.

- *$\langle\alpha_1, ..., \alpha_n \Rightarrow \varphi\rangle$ if $\alpha_1, ..., \alpha_n$ are arguments such that* $(\mathsf{Conc}(\alpha_1), ..., \mathsf{Conc}(\alpha_n), \varphi) \in \mathsf{Rules}_d$*. We let* $\mathsf{Conc}(\langle\alpha_1, ..., \alpha_n \Rightarrow \varphi\rangle) = \varphi$.

Given AT we denote by $\mathsf{Ar}^{\mathsf{AT}}$ the set of all arguments of AT. We define some other meta-syntactic functions, useful for analysing the structure of arguments.

- $\mathsf{Prem}(\alpha)$ returns the *premisses* of α and it is defined as follows: $\mathsf{Prem}(\langle\varphi\rangle) = \{\varphi\}$, $\mathsf{Prem}(\langle\alpha_1, ..., \alpha_n \hookrightarrow \varphi\rangle) = \mathsf{Prem}(\alpha_1) \cup ... \cup \mathsf{Prem}(\alpha_n)$ where $\hookrightarrow \in \{\twoheadrightarrow, \Rightarrow\}$.

- $\mathsf{sub}(\alpha)$ returns the *subarguments* of α and it is defined as follows: $\mathsf{sub}(\langle\varphi\rangle) = \{\langle\varphi\rangle\}$ and $\mathsf{sub}(\langle\alpha_1, ..., \alpha_n \hookrightarrow \varphi\rangle) = \{\langle\alpha_1, ..., \alpha_n \hookrightarrow \varphi\rangle\} \cup \mathsf{sub}(\alpha_1) \cup ... \cup \mathsf{sub}(\alpha_n)$ where $\hookrightarrow \in \{\twoheadrightarrow, \Rightarrow\}$.

[17] Conc is a function that, when applied to an argument α, returns its conclusion. Note that the notion of argument and the one of conclusion need to be defined by mutual recursion in ASPIC$^+$.

- $\mathsf{TopRule}(\alpha)$ returns the *top-rule* of α, i.e., the last one applied in the formation of α. It is defined as follows: $\mathsf{TopRule}(\langle\varphi\rangle)$ is left undefined, $\mathsf{TopRule}(\langle\alpha_1, ..., \alpha_n \twoheadrightarrow \varphi\rangle) = \mathsf{TopRule}(\langle\alpha_1, ..., \alpha_n \Rightarrow \varphi\rangle) = (\mathsf{Conc}(\alpha_1), ..., \mathsf{Conc}(\alpha_n), \varphi)$.

- $\mathsf{DefRule}(\alpha)$ returns the set of *defeasible rules* of α and it is defined as $\mathsf{DefRule}(\langle\varphi\rangle) = \varnothing$, $\mathsf{DefRule}(\langle\alpha_1, ..., \alpha_n \twoheadrightarrow \varphi\rangle) = \mathsf{DefRule}(\alpha_1) \cup ... \cup \mathsf{DefRule}(\alpha_n)$ and $\mathsf{DefRule}(\langle\alpha_1, ..., \alpha_n \Rightarrow \varphi\rangle) = \{(\mathsf{Conc}(\alpha_1), ..., \mathsf{Conc}(\alpha_n), \varphi)\} \cup \mathsf{DefRule}(\alpha_1) \cup ... \cup \mathsf{DefRule}(\alpha_n)$.

An argument α is said *firm* iff $\mathsf{Prem}(\alpha) \cap \mathcal{K}_p = \varnothing$ and it is said *strict* iff $\mathsf{DefRule}(\alpha) = \varnothing$. Given a set of formulas $\Gamma \subseteq \mathcal{L}$, we say that φ is a *strict consequence* of Γ (relatively to AT), abbreviated $\Gamma \vdash \varphi$, iff there is a strict argument α such that $\mathsf{Conc}(\alpha) = \varphi$ and $\mathsf{Prem}(\alpha) \subseteq \Gamma$.

We are now able to give a clear notion of attack, that partially accounts for the *dialectical tier* of argument strength that we informally characterised in Section 1.2.

Definition 17 (Attack)**.** *Let α and β be two arguments of a given argumentation theory, we define the following forms of* attack *between them:*

- α undermines β *(on φ) iff* $\mathsf{Conc}(\alpha) = \sim \varphi$ *for some* $\varphi \in \mathsf{Prem}(\beta) \cap \mathcal{K}_p$.

- α undercuts β *(on β') iff for some* $\beta' \in \mathsf{sub}(\beta)$*:* $\mathfrak{n}(\mathsf{TopRule}(\beta')) = \varphi$ *and* $\mathsf{Conc}(\alpha) = \sim \varphi$.

- α restrictedly rebuts β *(on β') iff for some* $\beta' \in \mathsf{sub}(\beta)$*:* $\beta' = \langle\beta'_1, ..., \beta'_n \Rightarrow \varphi\rangle$ *and* $\mathsf{Conc}(\alpha) = \sim \varphi$.

The notion of *restricted rebut* requires the targeted sub-argument (β') to have a defeasible top rule. This has been deemed unnatural, based on empirical cognitive studies (Yu et al., 2018). The more realistic notion of *unrestricted rebut*, where it is only required that β' is not firm, has been also explored in the literature (e.g., in Caminada et al. (2014); Heyninck and Straßer (2017)). However, if one opts for unrestricted rebut, then she has to take care of other design choices, as they may have important consequences on the rationality of the formalised agent (see below).

From attack to defeats via preferences. The above-defined attack relation expresses a sort of argument incompatibility: one should never accept simultaneously α and β if α attacks β or vice versa. However, attacks do not take into account the relative (supportive) strength of the arguments in conflict, which is essential to determine their dialectical strength. To illustrate this idea, let us go back to Example 2. We can now analyse the abstract argument a as the ASPIC$^+$ argument $\alpha = \langle\neg\mathsf{CloudySky} \Rightarrow \neg\mathsf{Rain}\rangle$, and the abstract argument b as the ASPIC$^+$ argument $\beta = \langle\mathsf{ForecastSaysRain} \Rightarrow \mathsf{Rain}\rangle$. It is clear that, according to Definition 17, α and β rebut each other. But then, why did we model the example abstractly as the asymmetric conflict $b \to a$? (see Figure 2.6).

As we said before, in an abstract argumentation framework (A, R), R is better understood as *defeat* relation. Defeats do not only take into account the structure of the involved arguments (as it happens with attacks), but also some sort of evaluative content that we might qualified as *subjective* or *agent-based*. In Example 2, Anne considers that β defeats α, but not vice versa because Anne concedes a higher reliability degree to the weather forecast than to her colleague's opinion. In an epistemic context, as the one of the current example, agents might not only use the reliability of the involved premisses, but also of the involved inference rules, so as to determine the relative strength of arguments (what we called the *support dimension* of argument strength in Section 1.2). In order to take this tier into account, ASPIC$^+$ remains abstract, and let us incorporate any *preference ordering* $\preceq$ (and its strict counter-part $\prec$, where $\alpha \prec \beta$ is defined as $\alpha \preceq \beta$ and $\beta \not\preceq \alpha$) among the arguments generated from an argumentation theory. *A priori*, no assumption is made about the properties of $\preceq$, although some of these are necessary to show that the outcome of the argumentation process is well-behaved. Moreover, according to Modgil and Prakken (2013) (whose view to this matter is ultimately rooted in the work of Pollock (1987)), undercutting attacks always succeed, no matter what the involved preferences are.

Definition 18 (Defeat)**.** α defeats β *iff either* α *undercuts* β*; or* α *undermines or rebuts* β *(on* β'*) and* $\alpha \not\prec \beta'$*.*

All arguments generated from an argumentation theory together with the defeat relation among them can be seen as a structure that we have already studied in some detail in this chapter:

Definition 19 (Associated argumentation framework)**.** *Let* $\mathsf{AT} = (\mathsf{AS}, \mathcal{K})$ *be an argumentation theory, and let* $\preceq$ *be an ordering on the set* $\mathsf{Ar}^{\mathsf{AT}}$ *of all arguments that can be constructed in* AT*, then* the argumentation framework associated to AT *is the pair* $(\mathsf{Ar}^{\mathsf{AT}}, \mathsf{R}^{\mathsf{AT}})$*, where* $\mathsf{R}^{\mathsf{AT}} \subseteq \mathsf{Ar}^{\mathsf{AT}} \times \mathsf{Ar}^{\mathsf{AT}}$ *is the defeat relation restricted to* $\mathsf{Ar}^{\mathsf{AT}}$*.*

The previous definition serves as a bridge between structured and abstract models of arguments or, more concretely, between ASPIC$^+$ and AFs. We are now able to apply Dung's semantics to a set of ASPIC$^+$ arguments generated from an argumentation theory, so as to sort out the set of sentences believed by an agent as the outcome of a process of arguing. Once again, several alternative notions can be adopted here. For instance, one could say that a sentence φ is believed by the agent of an argumentation theory if and only if φ is the conclusion of a σ-sceptically justified argument. As we have already warned several times, the user should, however, take some care on how the components of her instantiation of ASPIC$^+$ are defined, so that they satisfy a number of intuitive properties (usually called *rationality postulates*). We just review here the ones introduced by Caminada and Amgoud (2007), and refer to Caminada (2017) for further postulates.

Definition 20 (Rationality postulates)**.** *Let* E *be a complete extension of the AF associated to an argumentation theory* AT*, we say that* E *satisfies:*

- *the* sub-argument closure postulate *iff for every* $\alpha \in \mathsf{Ar}^{\mathsf{AT}}$*, if* $\alpha \in E$*, then* $\mathsf{sub}(\alpha) \subseteq E$*;*

- *the* direct-consistency postulate *iff there is no* $\varphi \in \mathcal{L}$ *such that* $\varphi, \sim \varphi \in \mathsf{Conc}(E)$*;*
- *the* indirect-consistency postulate *iff* $\mathsf{Conc}(E) \nvdash \bot$*;*
- *the* strict-closure postulate *iff* $\mathsf{Conc}(E) \vdash \varphi$ *implies* $\varphi \in \mathsf{Conc}$.

These postulates close our brief presentation of the mathematical tools upon which the contributions of this thesis are built. We hope to have provided the reader with the flavour of what is going to be the technical apparatus used through the contributions of Chapter 4. For a detailed discussion of our papers in relation to the existing literature, the reader is referred to Chapter 5.

Chapter 3

How the contributions approach the research problem

The contributions that constitute the core of this dissertation are split into two big thematic blocks that mimic the distinction, exposed in the previous chapter, between abstract and structured models of argumentation. The first of these tracks, which deals with various combinations of epistemic models and abstract argumentation frameworks, is formed by Paper I, Paper II, Paper III and Paper IV. In all these works, we focus entirely on the rhetoric reading of C1 that we saw in the Introduction:

$\mathsf{C1}_{\mathsf{rhetoric}}$ Higher-order epistemic attitudes[1] *conditions rhetoric argument evaluation*.[2]

Since we treat $\mathsf{C1}_{\mathsf{rhetoric}}$ from the perspective of abstract models of argumentation, then the nature, structure and origin of arguments are left unspecified in our analysis. Therefore, all the works belonging to this track have a dialectical focus, meaning that the main source for determining the strength of arguments are their dialectical relations with other arguments.

This connection between beliefs and argumentation, $\mathsf{C1}_{\mathsf{rhetoric}}$, has recently been exploited in the research area of strategic argumentation (Thimm, 2014), through the use of so-called *opponent modelling* techniques (Oren and Norman, 2009; Rienstra et al., 2013). In Paper I, we develop it from the perspective of epistemic logic. In order to do so, we first define persuasion in terms of a match between the goal of the speaker and the fine-grained *justification status* (Wu and Caminada, 2010) of the hearer –that is, how strongly she accepts a target argument in terms of its dialectical relations with others–[3] after communication has taken place. After that, we show how to use standard Kripke-style epistemic models to express uncertainty about the arguments that other agents are aware of, as well as higher-order beliefs, following the work of Schwarzentruber et al. (2012). Technically, we extend one of their languages in order to capture argument communication, providing also a complete axiomatization via reduction axioms. Our technical machinery allows distinguishing between what is actually persuasive and what is perceived as persuasive by

[1] These are, for instance, what an agent believes/knows that other believes/knows that... etc.

[2] That is, the persuasive force that an agent attributes to her available arguments.

[3] See Definition 13 for the precise, formal characterisation of this notion.

a speaker. We close the paper by providing some sufficient conditions for both notions to collapse, that is, for the beliefs of the speaker to be good enough for guaranteeing the achievement of her goals.

In Paper II, we sensibly expand the logical tools of Paper I in three different directions. First, we extend the propositional language we worked with in our previous publication, so as to capture the notion of fine-grained justification status (which was only described in the metalanguage before). This enables, combined with other elements, characterizing what is perceived as persuasive by a speaker (according to her beliefs above the listener's argumentative situation) directly in our object language. Second, we thoroughly study possible restrictions when defining the notion of *multi-agent argumentation framework* (these are, multi-agent extensions of the famous formalism of Dung (1995)), and how these can be combined meaningfully with Kripke-style epistemic semantics, providing sound and complete axiomatizations for these combinations. Third, we jump from the simple *argument disclosure* operation of Paper I, to the expressive framework of event models (Baltag and Moss, 2004) enriched with factual change operators (van Ditmarsch et al., 2005; van Benthem et al., 2006), in order to capture much more refined, mixed forms of argumentative and epistemic dynamics. The resulting logic is shown to be expressive enough to subsume and generalize two existing formalisms for reasoning about qualitative uncertainty and dynamics within the field of abstract argumentation: *incomplete argumentation frameworks* (Baumeister et al., 2018c) and *control argumentation frameworks* (Dimopoulos et al., 2018), as well as for studying systematically different ways of instantiating the idea of opponent modelling (Thimm, 2014).

In Paper III, we solve a question that remained open in the previous contribution: what is *the* epistemic logic underlying incomplete argumentation frameworks? We provide an answer by establishing a strong formal connection between incomplete argumentation frameworks (IAFs) and the epistemic logic of visibility (ELV) presented by Herzig et al. (2018) (which is actually the epistemic fragment of the logic of knowledge and control introduced by van der Hoek et al. (2011)). We strengthen that connection in two different directions. First, it is shown how argument acceptance problems in IAFs can be naturally translated to model checking problems in ELV. Second, we provide a minimal epistemic logic for IAFs, so as to unravel the hidden epistemic assumptions that IAFs come equipped with, which curiously turn out to be consistency of beliefs and distribution of the epistemic operator over disjunction of consistent literals (i.e., atoms or their negations).

Paper IV works in the spirit of the previous three contributions, but moving from the notion of *awareness of arguments* to the one of *knowledge of attacks*. Hence, we change our way of representing the partial information of each agent with respect to the underlying AF. This move is arguably interesting from a theoretical point of view for at least three reasons. First, it is more parsimonious, since we can directly apply knowledge-that modalities (the ones thoroughly studied in epistemic logic) to attacks, but we cannot do the same for arguments (an agent can *know an argument* or *be aware of it*, but it cannot *know that an argument*). Second, it fits well some real-life scenarios, such as those where arguments are not fully or explicitly stated (*enthymemes*), or those where agents have bounded reasoning resources, and therefore they might fail to see some of the conflicts.

Third, it seems that this modelling choice is much less studied in the literature (with the exception of Dyrkolbotn and Pedersen (2016)), so we also do it for the mere sake of exploration. Moreover, in this paper we do not depart from the expressive framework of full epistemic logic, as we did before, but from simpler multi-agent argumentation frameworks whose definition is clearly inspired by it, so as to make our approach more compact and closer to implementation. We show that this novel kind of structure is enough to model some kind of strategic behaviour. The main technical result consists in proving that this version of multi-agent argumentation frameworks, as well as their semantics and their updates with new attacks communicated by agents inside the system, can be characterized using public announcement logic.

Moving to the second track of our contributions, formed by Paper V and Paper VI, let us recall that C1 roughly says that argument evaluation is conditioned by the formation of certain epistemic attitudes, typically knowledge and belief. In these works, we narrow the interpretation of such a general principle, by focusing on an interpretation of it with a clear epistemic flavour. Let us retrieve it from the Introduction:

$\mathsf{C1}_{\mathsf{epistemic}}$ *arguments with believed (respectively, known) premisses are to be preferred to arguments with premisses that are not believed (respectively, that are unknown).*

There are two papers co-authored by the author of this dissertation (Burrieza and Yuste-Ginel, 2019, 2021) that have been left out of its core (Chapter 4) for presentation purposes, as they use a rather non-standard tool for modelling argumentation: justification logic.[4] However, they carried out a formal treatment of $\mathsf{C1}_{\mathsf{epistemic}}$ that can be seen as the preliminary step needed for Paper V and Paper VI.

In the first contribution included in this track, Paper V, we look at the problematic relation among $\mathsf{C1}_{\mathsf{epistemic}}$ and C2.[5] If adopted without restrictions, these principles jointly lead to an infinite regress, when the formalised agent is asked "why do you believe a certain sentence φ?". Our solution to this problem consists in distinguishing between *basic* (non-inferred) beliefs and *argument-based* beliefs within a formal language that imports ASPIC$^+$-like arguments (Modgil and Prakken, 2014) into an awareness epistemic logic *à la* Fagin and Halpern (1987). Conceptually speaking, our analysis finds qualified versions of $\mathsf{C1}_{\mathsf{epistemic}}$ and C2 that can be adopted consistently.[6] Finally, we examine the proposed formalism under the view of well-known rationality postulates for argumentation systems (Caminada and Amgoud, 2007), showing that our proposal accounts for a kind of minimal rationality, in contrast to more ideal approaches in the literature.

In Paper VI, we extend the conceptual content of the previous paper by constructing more solid theoretical grounds. In particular, we show how it could be understood

[4] We take advantage to point out by passing a recent interesting line of research that digs deeper into the relation between argumentation theory and justification logic, developed by Pandžić (2019, 2021).

[5] We apologize for the switch in notation: $\mathsf{C1}_{\mathsf{epistemic}}$ and C2 are noted, respectively, P2 and P1 in both papers of this track.

[6] This analysis is clearly inspired by foundationalist theories of epistemic justification (Hasan and Fumerton, 2018).

as a first approximation to model the distinction between *intuitive* and *inferred* beliefs of Sperber (1997) (called basic beliefs and argument-based beliefs in our terminology), used as one of the basis of the influential argumentative theory of reason presented by Mercier and Sperber (2011). Besides, we built upon the technical machinery of Paper V in two directions. First, we provide a complete axiomatization of its basic fragment and extend such a fragment with several dynamic operators. Concretely, we study the actions of *becoming aware* and *forgetting* an argument, *learning a defeasible rule* and *publicly announcing a formula* (actions imported for the dynamics of awareness literature (Grossi and Velázquez-Quesada, 2009)). Second, we broaden our postulates' analysis, by finding out conditions under which our framework becomes an instantiation of ASPIC$^+$, and therefore satisfies all Caminada and Amgoud (2007)'s rationality principles.

Chapter 4

Reprint of the contributions

We finally arrived at the core of this dissertation. As announced, it is formed by the reprint of six published contributions that deal with the combination of epistemic logic and formal argumentation. Besides, they are thematically divided in two tracks or blocks. The first track is contained in Section 4.1, entitled *Epistemic logics for abstract argumentation*, and formed by Paper I, Paper II, Paper IV and Paper III. The second track is contained in Section 4.2, entitled *Epistemic logics for structured argumentation*, and formed by Paper V and Paper VI. In the current on-line version, the original papers have been replaced by their bibliographic references, abstracts and external links, due to copyright issues.

4.1 Epistemic logics for abstract argumentation

4.1.1 Paper I

Full reference. Proietti, C. and Yuste-Ginel, A. (2020). Persuasive argumentation and epistemic attitudes. In Soares Barbosa, L. and Baltag, A., editors, *Dynamic Logic. New Trends and Applications*, volume 12005 of LNCS, pages 104–123. Springer. DOI: 10.1007/978-3-030-38808-9_7.

External link. A preprint version is available at https://eprints.illc.uva.nl/id/eprint/1736/1/Proietti_Yuste_PAEP_preprint.pdf.

Abstract. This paper studies the relation between persuasive argumentation and the speaker's epistemic attitude. Dung-style abstract argumentation and dynamic epistemic logic provide the necessary tools to characterize the notion of persuasion. Within abstract argumentation, persuasive argumentation has been previously studied from a game-theoretic perspective. These approaches are blind to the fact that, in real-life situations, the epistemic attitude of the speaker determines which set of arguments will be disclosed by her in the context of a persuasive dialogue. This work is a first step to fill this gap.

For this purpose we extend one of the logics of Schwarzentruber et al. with dynamic operators, designed to capture communicative phenomena. A complete axiomatization for the new logic via reduction axioms is provided. Within the new framework, a distinction between actual persuasion and persuasion from the speaker's perspective is made. Finally, we explore the relationship between the two notions.

4.1.2 Paper II

Full reference. Proietti, C. and Yuste-Ginel, A. (2021). Dynamic epistemic logics for abstract argumentation. *Synthese*, 199(3): 8641–8700. DOI: 10.1007/s11229-021-03178-5.

External link. https://link.springer.com/content/pdf/10.1007/s11229-021-03178-5.pdf.

Abstract. This paper introduces a multi-agent dynamic epistemic logic for abstract argumentation. Its main motivation is to build a general framework for modelling the dynamics of a debate, which entails reasoning about goals, beliefs, as well as policies of communication and information update by the participants. After locating our proposal and introducing the relevant tools from abstract argumentation, we proceed to build a three-tiered logical approach. At the first level, we use the language of propositional logic to encode states of a multi-agent debate. This language allows to specify which arguments any agent is aware of, as well as their subjective justification status. We then extend our language and semantics to that of epistemic logic, in order to model individuals' beliefs about the state of the debate, which includes uncertainty about the information available to others. As a third step, we introduce a framework of dynamic epistemic logic and its semantics, which is essentially based on so-called event models with factual change. We provide completeness results for a number of systems and show how existing formalisms for argumentation dynamics and unquantified uncertainty can be reduced to their semantics. The resulting framework allows reasoning about subtle epistemic and argumentative updates –such as the effects of different levels of trust in a source– and more in general about the epistemic dimensions of strategic communication.

4.1.3 Paper III

Full reference. Herzig, A. and Yuste-Ginel, A. (2021c). On the epistemic logic of incomplete argumentation frameworks. In M. Bienvenu, G. Lakemeyer, and E. Erdem, editors, *Proceedings of the 18th International Conference on Principles of Knowledge Representation and Reasoning*, pages 681–685. IJCAI Organization. DOI: 10.24963/kr.2021/69.

External link. https://proceedings.kr.org/2021/69/.

Abstract. We study the relation between two existing formalisms: incomplete argumentation frameworks (IAFs) and epistemic logic of visibility (ELV). We show that the set of completions of a given IAF naturally corresponds to a specific equivalence class of possible worlds within the model of visibility. This connection is further strengthened in two directions. First, we show how to reduce argument acceptance problems of IAFs to ELV model-checking problems. Second, we highlight the epistemic assumptions that underlie IAFs by providing a minimal epistemic logic for IAFs.

4.1.4 Paper IV

Full reference. Herzig, A. and Yuste-Ginel, A. (2021b). Multi-agent abstract argumentation frameworks with incomplete knowledge of attacks. In Zhou, Z.-H., editor, *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pages 1922–1928. IJCAI Organization. DOI: 10.24963/ijcai.2021/265.

External link. https://doi.org/10.24963/ijcai.2021/265.

Abstract. We introduce a multi-agent, dynamic extension of abstract argumentation frameworks (AFs), strongly inspired by epistemic logic, where agents have only partial information about the conflicts between arguments. These frameworks can be used to model a variety of situations. For instance, those in which agents have bounded logical resources and therefore fail to spot some of the actual attacks, or those where some arguments are not explicitly and fully stated (enthymematic argumentation). Moreover, we include second-order knowledge and common knowledge of the attack relation in our structures (where the latter accounts for the state of the debate), so as to reason about different kinds of persuasion and about strategic features. This version of multi-agent AFs, as well as their updates with public announcements of attacks (more concretely, the effects of these updates on the acceptability of an argument) can be described using S5-PAL, a well-known dynamic-epistemic logic. We also discuss how to extend our proposal to capture arbitrary higher-order attitudes and uncertainty.

4.2 Epistemic logics for structured argumentation

4.2.1 Paper V

Full reference. Burrieza, A. and Yuste-Ginel, A. (2020). Basic beliefs and argument-based beliefs in awareness epistemic logic with structured arguments. In Prakken et al., editors, *Proceedings of the COMMA 2020*, pages 123–134. IOS Press. DOI: 10.3233/FAIA200498.

External link. https://ebooks.iospress.nl/volumearticle/55364.

Abstract. There are two intuitive principles governing belief formation and argument evaluation that can potentially clash. After arguing that adopting them unrestrictedly leads to an infinite regress, we propose a formal framework in which qualified versions of both principles can be subscribed without falling into such a regress. The proposal integrates tools from two different traditions: structured argumentation and awareness epistemic logic. We show that our formalism satisfies certain rationality postulates and argue that the rest of them can be seen as too ideal when modelling resource-bounded agents.

4.2.2 Paper VI

Full reference. Burrieza, A. and Yuste-Ginel, A. (2021). An awareness epistemic framework for belief, argumentation and their dynamics. In Halpern and Perea, editors, *Proceedings TARK 2021*, EPTCS 335, pp. 69–83. Open Publishing Association. DOI: 10.4204/EPTCS.335.6.

External link. https://doi.org/10.4204/EPTCS.335.6.

Abstract. The notion of argumentation and the one of belief stand in a problematic relation to one another. On the one hand, argumentation is crucial for belief formation: as the outcome of a process of arguing, an agent might come to (justifiably) believe that something is the case. On the other hand, beliefs are an input for argument evaluation: arguments with believed premisses are to be considered as strictly stronger by the agent to arguments whose premisses are not believed. An awareness epistemic logic that captures qualified versions of both principles was recently proposed in the literature. This paper extends that logic in three different directions. First, we try to improve its conceptual grounds, by depicting its philosophical foundations, critically discussing some of its design choices and exploring further possibilities. Second, we provide a (heretofore missing) completeness theorem for the basic fragment of the logic. Third, we study, using techniques from dynamic epistemic logic, how different forms of information change can be captured in the framework.

Chapter 5

Related work

In the previous chapter, we have reprinted the contributions that constitute the core of this thesis. These were divided into two big tracks, one dealing with epistemic logics for abstract argumentation frameworks, and the other one dealing with epistemic logics for structured argumentation. This chapter pursues two main objectives:

1. discussing collectively the approaches and results provided in each of the contributions of the precedent chapter, so as to build more solid bridges among them (an *internal comparison*); and

2. comparing these results with closely related literature (an *external comparison*).

Once again, we proceed by tracks, so that objective 1. is pursued in sections 5.1.1 and 5.2.1; and objective 2. is pursued in sections 5.1.2 and 5.2.2. We recall that the understanding of this chapter presupposes the reading of the previous one (of course, this can be done selectively). As an introduction, let us sketch a quick guide to the previous literature.

The idea of combining epistemic logic and formal argumentation was not born in this thesis. It instead continues a recent tradition within the epistemic logic community. This tradition can be split into two well differentiated branches. The first one uses epistemic models in order to provide a formal theory of qualitative uncertainty and/or multi-agency about argumentation frameworks, and some of its main antecedents are the work by Schwarzentruber et al. (2012), the one by Sakama and Cao Son (2019, 2020), and the one by Dyrkolbotn and Pedersen (2016). The second one imports tools from argumentation theory into epistemic models in order to provide an argumentatively inspired definition of the notion of *justified belief*. To the best of our knowledge, the second branch is best represented by the work of Grossi and van der Hoek (2014), the ones by Shi et al. (2017, 2018, 2021) and Shi (2021), and by the recent approach by Li and Wáng (2020); Wáng and Li (2021). Both branches of research can be understood, at least partially, as general formal approaches to C1 and C2 respectively (the general connections between argumentation theory and epistemology that we spotted in the introduction to this thesis). Concurrently, some works within the formal argumentation literature have aimed, independently from the epistemic logic perspective but sharing some of its main goals, to build

formalisms for representing different sorts of qualitative uncertainty about AFs, providing, at least partially, an account of $\mathsf{C1}_{\mathsf{rhetoric}}$. Among these studies, there have been recent increasing interest on so-called *incomplete argumentation frameworks* (Baumeister et al., 2021, 2018a,c,b; Fazzinga et al., 2020), as well as on *control argumentation frameworks* (Dimopoulos et al., 2018; Niskanen et al., 2020), which furthermore include a dynamic component.

5.1 Epistemic reasoning about argumentation frameworks

5.1.1 Relation among contributions

Relative incomparability. We start our internal comparison of the first track of contributions by noting that, while Paper II is clearly an extension of Paper I; it is, *a priori* and only to a technical extent, incomparable to Paper III and Paper IV. The reason for this incomparability is the heterogeneity of the encodings of argumentative notions used in those works (something that we already mentioned in Chapter 2). However, we can provisionally skip this inconvenience by abstracting away from both encodings, since, at the end of the day, they are orthogonal to many important aspects. Moreover, we will show how to adapt part of the work performed in Paper II to the more standard encoding used in Paper IV and Paper III, so as to give evidence of the alleged orthogonality. Let us provisionally assume that we have a set of propositional variables (parametrised by a set of arguments A and a set of agents Ag) where nor in-variables nor $x \epsilon E$-variables occur (these are the kind of variables used in the two different encodings), that is, the set of variables is defined as

$$\mathsf{At}(\mathsf{A}, \mathsf{Ag}) = \{\mathsf{aw}_i(x) \mid i \in \mathsf{Ag}, x \in \mathsf{A}\} \cup \{x \rightsquigarrow y \mid x, y \in \mathsf{A}\}.$$

Note that there are notational variants of $\mathsf{At}(\mathsf{A}, \mathsf{Ag})$ that are subsets of the sets of variables employed in Paper II, Paper IV and Paper III (in the latter case, with $\mathsf{Ag} = \{1\}$, and 1 being omitted from the subscript $\mathsf{aw}_1(x)$).

Assuming this simplification, it can be claimed that Paper II is the work of this thesis that provides the most general account of epistemic reasoning about AFs and their dynamics, since all the logical tools that we used in the rest of contributions of this track can be seen as special cases of it.[1]

Paper I as a fragment of Paper II. Let us start justifying our last assertion by examining the logical tools employed in Paper I. Its static language is a notational variant of the language of Paper II restricted to aw-variables. It is true that we treat owns_i as an operator in Paper I, while it is treated as a set of atoms ($\mathsf{aw}_i(a)$, $\mathsf{aw}_i(b)$, ...) in Paper II; but this has no effects since both options can be shown to be equivalent. Semantically, each static model M of Paper I (Definition 7) can be systematically transformed into an

[1] Just to be clear, *more general* does not necessarily mean *better*, since the rest of the contributions enable a more fine-grained analysis of different subtopics.

$\mathcal{A}o\mathcal{A}$-model of Paper II (Definition 10) restricted to the set of aw-variables, by means of a transformation function τ. More in detail, let $M = (\mathsf{W}, \mathcal{R}, \mathcal{D})$ be a Paper I's model, we define $\tau(\mathsf{W}, \mathcal{R}, \mathcal{D}) = (\mathsf{W}, \mathcal{R}, \tau(\mathcal{D}))$, where $\mathcal{D}$ gets transformed in the atomic evaluation $\tau(\mathcal{D})$ defined by

$$w \in \tau(\mathcal{D})(\mathsf{aw}_i(x)) \text{ iff } x \in \mathcal{D}(i, w)$$

for every world w, every argument x and every agent i. It is almost immediate to show that a formula of Paper I's language is true in a pointed model (M, w) iff it is true in its corresponding transformation $(\tau(M), w)$.

Regarding the dynamics of Paper I, one can check that, for any Paper I's model M, and any argument a, we have that

$$\tau(M^{a!}) = \tau(M) \otimes \mathsf{Pub}^a,$$

where Pub^a is the event model for publicly adding an argument (see Example 5 of Paper II). Therefore, since $(\cdot)^{a!}$ is the only primitive action in Paper I, the above equality enables a reduction of its dynamic aspects to the formal machinery of Paper II.

Paper IV as a fragment of Paper II. The reduction of the multi-agent AFs of Paper IV to the $\mathcal{S}5(\mathcal{EA})$-models of paper Paper II (restricted to attack variables) is explicitly performed in Section 6.1 of Paper IV, so we skip it for avoiding tedious repetition. We just make the observation that while each $\mathcal{S}5(\mathcal{EA})$-model induces a unique multi-agent AF, each multi-agent AF can be represented by infinitely many different $\mathcal{S}5(\mathcal{EA})$-models. This is caused by the fact that while multi-agent AFs only contain first-order, second-order and public information about the attack relation, $\mathcal{S}5(\mathcal{EA})$-models contain arbitrarily higher-order information. Regarding dynamics, public announcement is the only primitive action used in Paper IV and, as it was mentioned in Chapter 2 (Example 5), it has a straightforward equivalent formulation in terms of the event models used in Paper II.

Paper III as a fragment of Paper II. As we said in Paper III, it can be understood as the solution of an open problem mentioned in Section 8.1. of Paper II (more concretely, in footnote 36). Furthermore, Herzig et al. (2018) showed that the logic of visibility used in Paper III is just a well-behaved fragment of the general epistemic logic used in Paper II.

Adapting Paper II to Paper III and Paper IV's encoding. We now proceed to sketch the promised adaptation of Paper II to the encoding of argumentative notions used in Paper IV and Paper III, which is ultimately rooted in the propositional encoding proposed by Besnard and Doutre (2004). Technically, we are going to provide a method for transforming any $\mathcal{EA}$-model M into a model M'' satisfying the same formulas of the multi-agent epistemic language restricted to $\mathsf{At}(\mathsf{A}, \mathsf{Ag})$, where moreover the encoding of Paper IV and Paper III can be plugged in.[2]

[2]This adaptation can be seen as a generalization to standard epistemic logic of the result provided in Proposition 4 of Paper III, that was only formulated for the epistemic logics of visibility.

So, let $M = (\mathsf{W}, \mathcal{R}, \mathsf{V})$ be an $\mathcal{EA}$-model, we first restrict the domain of V to the set of variables $\mathsf{At}(\mathsf{A}, \mathsf{Ag})$. Trivially, $M' = (\mathsf{W}, \mathcal{R}, \mathsf{V}_{|\mathsf{At}(\mathsf{A},\mathsf{Ag})})$ satisfies the same $\mathsf{At}(\mathsf{A}, \mathsf{Ag})$-formulas as M.[3]

Second, let $\mathsf{IN}_\mathsf{A} = \{\mathsf{in}_x \mid x \in \mathsf{A}\}$, we define the $\wp(\mathsf{A})$-*cluster model* as the mono-modal Kripke model $M^{\wp(\mathsf{A})} = (\mathsf{W}^{\wp(\mathsf{A})}, \mathcal{R}^{\wp(\mathsf{A})}, \mathsf{V}^{\wp(\mathsf{A})})$ for the set of variables IN_A where:

- $\mathsf{W}^{\wp(\mathsf{A})} = \{w_E \mid E \subseteq \mathsf{A}\}$;
- $\mathcal{R}^{\wp(\mathsf{A})} = \mathsf{W}^{\wp(\mathsf{A})} \times \mathsf{W}^{\wp(\mathsf{A})}$;[4] and
- for all $x \in \mathsf{A}$, $w_E \in V(\mathsf{in}_x)$ iff $x \in E$.

Finally, let $M = (\mathsf{W}, \mathcal{R}, \mathsf{V})$ be an $\mathcal{EA}$-model for the set of variables $\mathsf{At}(\mathsf{A}, \mathsf{Ag})$, we define $M'' = (\mathsf{W}'', \mathcal{R}^h, \mathcal{R}^v, \mathsf{V}'')$, where:

- $\mathsf{W}'' = \mathsf{W} \times \mathsf{W}^{\wp(\mathsf{A})}$.
- $\mathcal{R}^h : \mathsf{Ag} \to \wp(\mathsf{W}'' \times \mathsf{W}'')$, where $\mathcal{R}^h_i = \{((v, w_E), (u, w_E)) \mid (v, u) \in \mathcal{R}_i\}$ for every $i \in \mathsf{Ag}$.
- $\mathcal{R}^v \subseteq \mathsf{W}'' \times \mathsf{W}''$ is the relation $\{((v, w_E), (v, w_{E'})) \mid (w_E, w_{E'}) \in \mathcal{R}^{\wp(\mathsf{A})}\}$.
- $\mathsf{V}(p) = \begin{cases} \{(v, w_E) \mid v \in \mathsf{V}'(p), w_E \in \mathsf{W}^{\wp(\mathsf{A})}\} & \text{if} \quad p \in \mathsf{At}(\mathsf{A}, \mathsf{Ag}), \\ \{(v, w_E) \mid v \in \mathsf{W}, w_E \in \mathsf{V}^{\wp(\mathsf{A})}(p)\} & \text{if} \quad p \in \mathsf{IN}_\mathsf{A}. \end{cases}$

Note that each $\mathcal{EA}$-model M unequivocally determines its modification M''. We use the following extended language to describe these modified models:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Box_i\varphi \mid \Box^\wp\varphi \qquad p \in \mathsf{At}(\mathsf{A}, \mathsf{Ag}) \cup \mathsf{IN}_\mathsf{A}.$$

The modal formulas of the above language are interpreted at pointed (modified) models as follows:

$$\begin{array}{rll} M'', w \models \Box_i\varphi & \text{iff} & \text{for all } w' \in \mathsf{W}'', w\mathcal{R}^h_i w' \text{ implies } M'', w' \models \varphi \\ M'', w \models \Box^\wp\varphi & \text{iff} & \text{for all } w' \in \mathsf{W}'', w\mathcal{R}^v w' \text{ implies } M'', w' \models \varphi \end{array}$$

As the familiar reader may have noticed, the frame $(\mathsf{W}'', \mathcal{R}^h, \mathcal{R}^v)$ is the *product frame* $(\mathsf{W}, \mathcal{R}) \times (\mathsf{W}^{\wp(\mathsf{A})}, \mathcal{R}^{\wp(\mathsf{A})})$ (Gabbay and Shehtman, 1998; Kurucz et al., 2003). Hence, in addition to the epistemic/doxastic axioms chosen for each $\Box_i$ and to the $S5$-axioms for $\Box^\wp$, we have that the following schemas are valid in our modified models:

$$\begin{array}{ll} \Diamond_i\Diamond^\wp\varphi \leftrightarrow \Diamond^\wp\Diamond_i\varphi & \text{(commutativity) and;} \\ \Diamond_i\Box^\wp\varphi \rightarrow \Box^\wp\Diamond_i\varphi & \text{(confluence).} \end{array}$$

[3] These are formulas whose sets of variables are subsets of $\mathsf{At}(\mathsf{A}, \mathsf{Ag})$.

[4] As in any other Kripke model with a single accessibility relation which is also total, we could alternatively leave $\mathcal{R}^{\wp(\mathsf{A})}$ implicit by setting the semantic interpretation of the modal box as ranging over the whole domain.

Moreover, due to the definition of the modified valuation V'', the following schemas are also valid (for any $a, b \in \mathsf{A}$):

$$
\begin{array}{l}
a \rightsquigarrow b \rightarrow \Box^{\wp} a \rightsquigarrow b \\
\neg a \rightsquigarrow b \rightarrow \Box^{\wp} \neg a \rightsquigarrow b \\
\mathsf{aw}_i(a) \rightarrow \Box^{\wp} \mathsf{aw}_i(a) \\
\neg \mathsf{aw}_i(a) \rightarrow \Box^{\wp} \neg \mathsf{aw}_i(a)
\end{array}
$$

$$
\Diamond^{\wp}(\textstyle\bigwedge_{x \in E} \mathsf{in}_x \wedge \bigwedge_{x \in \mathsf{A} \setminus E} \neg \mathsf{in}_x) \quad \text{for every } E \subseteq \mathsf{A}.
$$

How are these models related to the aim of adapting our work of Paper II to the propositional encoding of AFs provided by Besnard and Doutre (2004)? Note that the modality $\Box^{\wp}$ essentially goes through every subset of A (represented by true in-variables). Hence, when we check if $\Box^{\wp}\varphi$ is true at $M, (v, w_E)$, we basically check if φ is true at all possible worlds that differ from the current one just in the values given to in-variables. This enables an adaptation of the so-called *satisfiability approach* of Besnard and Doutre (2004) for encoding abstract argumentation semantics (more precisely, of its multi-agent version, as developed by Doutre et al. (2017)). As an example, we provide the encoding of the stable semantics. Remember that the propositional formula for capturing stable semantics of a multi-agent AF is

$$
\begin{aligned}
\mathsf{Stable}_i = \bigwedge_{x \in \mathsf{A}} \Big(& (\mathsf{in}_x \rightarrow \mathsf{aw}_i(x)) \wedge \\
& \big(\mathsf{aw}_i(x) \rightarrow (\mathsf{in}_x \leftrightarrow \neg \bigvee_{y \in \mathsf{A}} (\mathsf{in}_y \wedge (\mathsf{aw}_i(y) \wedge x \rightsquigarrow y)))\big)\Big).
\end{aligned}
$$

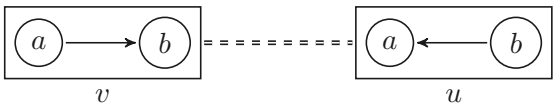
Proposition 1. *Let M be an $\mathcal{EA}$-model and let M'' its modified model with $w \in \mathsf{W}''$. We have that $M'', w \models \Diamond^{\wp}(\bigwedge_{x \in E} \mathsf{in}_x \wedge \bigwedge_{x \in \mathsf{A} \setminus E} \neg \mathsf{in}_x \wedge \mathsf{Stable}_i)$ iff E is stable with respect to $(\mathsf{A}_i(w), \mathsf{R}_i(w))$.*[5]

It can also be shown that the encoding is expressible enough to capture the notion of fine-grained justification status used in Paper II (see also Definition 13 of Chapter 2). As an example, consider the following result:

Proposition 2. *Let $a \in \mathsf{A}$, let M be an $\mathcal{EA}$-model, let M'' its modification, and let $w \in \mathsf{W}''$ be a world of M''. We have that $M'', w \models \Diamond^{\wp}(\mathsf{Stable}_i \wedge \mathsf{in}_a) \wedge \Diamond^{\wp}(\mathsf{Stable}_i \wedge \neg \mathsf{in}_a)$ iff a is borderline (IO) with respect to $(\mathsf{A}_i(w), \mathsf{R}_i(w))$.*

Let us illustrate both results with an example. Consider the following single-agent $\mathcal{EA}$-model:

[5] Let us retrieve from Paper II the meaning of this notation: $\mathsf{A}_i(w) = \{x \in \mathsf{A} \mid M, w \models \mathsf{aw}_i(x)\}$ and $\mathsf{R}_i(w) = \{(x, y) \in \mathsf{A} \times \mathsf{A} \mid M, w \models x \rightsquigarrow y\} \cap (\mathsf{A}_i(w) \times \mathsf{A}_i(w))$.



where arguments are represented as circles, defeats as directed arrows, possible worlds as rectangles, and the accessibility relation $\mathcal{R}$ for the single agent is depicted using a double, dashed, and undirected arrow (we assume symmetry and reflexivity of $\mathcal{R}$, hence we are working with an $\mathcal{S}5(\mathcal{EA})$-model). Moreover, we assume that the only agent is aware of both arguments a and b at each world. Note that this graphical representation abstracts away from the valuation of $x\epsilon E_k$-variables, so that it coincides with the draw assigned to M'. We can intuitively think of this model as representing uncertainty of the agent about her opponent's view of the underling AF. For instance, suppose that a and b are arguments with incompatible conclusions that were asserted, respectively, by the right-hand and the left-hand presidential candidate, and that our agent is not sure about the political choice of her opponent (but the agent knows that her opponent is already inclined towards one of the two positions). The fully modified model M'' is depicted in Figure 5.1, where the evaluation of in-variables is represented through dashed boxes inside each world, while the clusters generated by $\mathcal{R}^v$ are depicted as double lined rectangles. Let us suppose moreover that the actual world is v.[6] We have that

$$M'', (v, w_\varnothing) \models \Box\left(\Diamond^{\wp}\mathsf{Stable} \wedge \Box^{\wp}\Big(\mathsf{Stable} \rightarrow \big((\mathsf{in}_a \wedge a \rightsquigarrow b) \vee (\mathsf{in}_b \wedge b \rightsquigarrow a)\big)\Big)\right)$$

capturing the fact that the agent believes that either a is strongly accepted and b strongly rejected (by her opponent) or vice versa.[7]

We expect that the whole dynamic apparatus of Paper II can be adapted without excessive problems to this new setting, and that our axiomatisation results can also be extrapolated in a relatively simple manner (conserving all the work done for studying the preservation of awareness properties after event model execution). Details are left for future work.

To conclude this subsection, let us remark that the current adaptation throws some light on the difference among both encodings. On the one hand, Paper II's approach puts all the weight of the encoding process in the syntax. As we claimed in the paper, this technique provokes an exponential explosion on the size of formulas capturing argumentative notions. On the other hand, although the formulas used in the encoding adapted from Paper III and Paper IV are the original ones, and thus polynomially long on the size of the set A, the number of possible worlds grows exponentially. This exponential growth is "hidden" in the original approach, because instead of reducing the computation of argumentative notions to model checking problems (as we do here), it requires *satisfiability checking* of a propositional formula, which implies, in the worst case, going through $2^{||\mathsf{At}(\mathsf{A},\mathsf{Ag})\cup\mathsf{IN}_\mathsf{A}||}$ propositional valuations.

[6] Note that in modified $\mathcal{EA}$-models, epistemic alternatives (including the actual world) are captured by $\mathcal{R}^v$-clusters rather than by possible worlds, since each member of the cluster only differs from each other in the valuation of argumentative information (that is, in the value assigned to in-variables).

[7] Since we only have one agent, we have dropped subscripts from both $\Box$ and Stable.

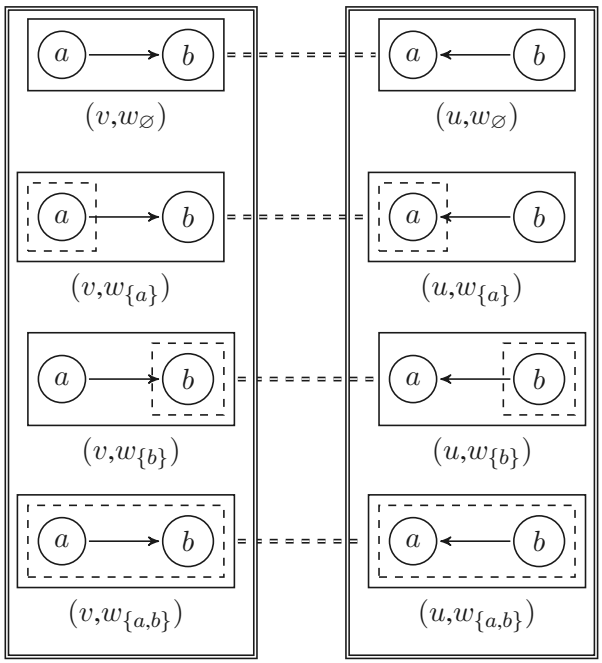


Figure 5.1: Example of a modified $\mathcal{EA}$-model

5.1.2 Closely related work

The approach by Schwarzentruber, Vesic and Rienstra

The work by Schwarzentruber et al. (2012) (SVR, for the rest of this section) is probably the clearest source of inspiration in our Paper I and Paper II. As general features differentiating both lines of work we point out two salient ones:

- Although SVR use dynamic notions as a strong intuition underlying their work, they leave these notions in a motivational/unformalised dimension, while we take their formalization as one of our main objectives.
- On the other hand, SVR provide a deep complexity analysis of reasoning problems associated to the studied logics; something that we leave aside in all our works.

We proceed to present more in detail their three logical frameworks, and to compare them with our contributions.

$\mathcal{L}_1$ and its models. The first language (noted $\mathcal{L}_1$) and models introduced by SVR are the ones that we used in Paper I. The only difference is that we assumed finiteness of the set of all arguments A, while this assumption is not present in their work. However, the difference is invisible to the language, as it cannot differentiate between models with infinite sets of arguments and models with finite sets. More precisely, each formula $\varphi \in$

$\mathcal{L}_1$ is satisfiable in a model for an infinite set of arguments iff it is satisfiable in a model for a finite subset of it.

$\mathcal{L}_2$ and its models. The second language defined by SVR, noted $\mathcal{L}_2$, is split into two layers and, in addition to the set of all arguments A, and the set of agents Ag, it is parametrised by a countable set of atomic propositional variables $\mathsf{At} = \{p, q, r, ..\}$. Both layers of the language are given by mutual recursion:

$\varphi ::= [\mathsf{U}]\psi \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid B_i\varphi \qquad i \in \mathsf{Ag},$

$\psi ::= p \mid \neg\psi \mid (\psi \wedge \psi) \mid \mathsf{isarg}(a) \mid \mathsf{ownedby}(i) \mid [\mathsf{attacks}]\psi \mid [\mathsf{is\_attacked}]\psi$
$p \in \mathsf{At}, a \in \mathsf{A}, i \in \mathsf{Ag}.$

An $\mathcal{L}_2$*-model* based on an AF (A, R) is a tuple $(\mathsf{W}, \mathcal{R}, \mathcal{A})$ where $(\mathsf{W}, \mathcal{R})$ is a *doxastic frame* (i.e., a multi-agent Kripke frame with serial, transitive, and euclidian accessibility relations), and $\mathcal{A}$ maps each member of W into a *labelled argumentation framework* $\mathcal{A}_w = (\mathsf{A}_w, \mathsf{R}_w, L_w)$, where:

- A_w is a finite subset of $\mathsf{A} \cup \{?_0, ?_1, ...\}$.
- $\mathsf{R}_w \subseteq \mathsf{A}_w \times \mathsf{A}_w$ is an attack relation such that $\mathsf{R}_w \cap (\mathsf{A} \times \mathsf{A}) = \mathsf{R} \cap (\mathsf{A}_w \times \mathsf{A}_w)$.
- $L_w : \mathsf{A}_w \to \wp(\mathsf{Ag} \cup \mathsf{At})$.

Note that the second condition is a reformulation of SCAA[8] for arguments in $\mathsf{A}_w \cap \mathsf{A}$. The intuition underlying the elements of $\{?_0, ?_1, ...\}$ is that they are arguments that an agent does not own (she is not able to use them in a debate), but about which she can hold beliefs. In other words, we can think of ?-arguments as devices used to express some kind of *de dicto* beliefs about arguments, as the one expressed in the sentence "I think that you have (you are aware of) a proof for theorem X, but I don't". Besides, the idea behind assigning subsets of At to each argument is to equip arguments with "atomic properties", so that $p \in L_w(a)$ can informally represent the assertion "argument a is about politics".

Moreover, two restrictions, resembling PIAw and GNIAw (see Paper II, Definition 10) are imposed on $\mathcal{L}_2$-models. For all agents i, j and for all worlds w, u:

1. $w\mathcal{R}_i u$ implies that $\{x \in \mathsf{A}_u \cap \mathsf{A} \mid i \in L_u(x)\} = \{x \in \mathsf{A}_w \cap \mathsf{A} \mid i \in L_w(x)\}$.
2. $w\mathcal{R}_i u$ implies that $\{x \in \mathsf{A}_u \cap \mathsf{A} \mid j \in L_u(x)\} \subseteq \{x \in \mathsf{A}_w \cap \mathsf{A} \mid i \in L_w(x)\}$.[9]

φ-formulas are interpreted in pointed models as follows (we omit the clauses for Boolean connectives):

$$\begin{array}{rll} M, w \models B_i\varphi & \text{iff} & \text{for all } v \in \mathsf{W},\ w\mathcal{R}_i v \text{ implies } M, v \models \varphi \\ M, w \models [\mathsf{U}]\psi & \text{iff} & \text{for all } x \in \mathsf{A}_w \text{ we have } \mathsf{A}_w, x \models \psi \end{array}$$

[8] See Section 3 of Paper II for the meaning of SCAA.
[9] We have corrected a typo from the original formulation of SVR.

so that the universal modality $[\mathsf{U}]$ ranges over all arguments of the labelled AF associated to the current world. ψ-formulas, on their side, are interpreted at arguments inside each possible world, by means of the following recursive clauses:

$$\begin{array}{rll} \mathsf{A}_w, a \models p & \text{iff} & p \in L_w(a) \\ \mathsf{A}_w, a \models \mathsf{isarg}(b) & \text{iff} & a = b \\ \mathsf{A}_w, a \models \mathsf{ownedby}(i) & \text{iff} & i \in L_w(a) \\ \mathsf{A}_w, a \models [\mathsf{attacks}]\psi & \text{iff} & \text{for all x such that } (a, x) \in \mathsf{R}_w \text{ we have } \mathsf{A}_w, x \models \psi \\ \mathsf{A}_w, a \models [\mathsf{is\_attacked}]\psi & \text{iff} & \text{for all x such that } (x, a) \in \mathsf{R}_w \text{ we have } \mathsf{A}_w, x \models \psi \end{array}$$

We now proceed to compare $\mathcal{L}_2$-models to our $\mathcal{A}o\mathcal{A}$-models (Paper II).[10] Clearly, they share the same modelling spirit, as the idea is to embed an epistemically alternative AF in each possible world of a Kripke model. The first obvious difference that we find is the presence of ?-arguments in $\mathcal{L}_2$-models, whose modelling role has already been explained. It seems that a slight adaptation of our $\mathcal{A}o\mathcal{A}$-models that includes a new subset of propositional variables $\{?_0, ..., ?_n\}$ could save the difference (as each A_w is assumed to be finite in SVR). Moreover, it is interesting to note that the role of ?-arguments can be simulated by weakening the set of assumptions about awareness of arguments and knowledge of attacks that are present in $\mathcal{L}_2$-models and $\mathcal{A}o\mathcal{A}$-models. More concretely, we are thinking of the subclass of $\mathcal{EA}$-models where PIAw and NIAw hold (so that agents are fully introspective regarding the arguments they are aware of), but where GNIAw can potentially fail (so that an agent i can consider epistemically possible that another agent j is aware of a argument a that i is not aware of), without postulating the existence of a new kind of arguments $\{?_0, ...\}$. In this fashion, PIAt and NIAt should be substituted by a weaker property that retains the informal spirit of being a modal version of SCAA. We propose the following property, expressing that the knowledge that each agent has of the attack relation is sound and complete whenever she is aware of the involved arguments

$$w\mathcal{R}_i u \text{ implies } \mathsf{R}(w) \cap (\mathsf{A}_i(w) \times \mathsf{A}_i(w)) = \mathsf{R}(u) \cap (\mathsf{A}_i(w) \times \mathsf{A}_i(w)),$$

[11] which is captured by the axiom scheme

$$(\mathsf{aw}_i(a) \wedge \mathsf{aw}_i(b)) \rightarrow \big((a \rightsquigarrow b \rightarrow \Box_i a \rightsquigarrow b) \wedge (\neg a \rightsquigarrow b \rightarrow \Box_i \neg a \rightsquigarrow b)\big).$$

In this new setting, the informal meaning of formulas that SVR give as the motivation for the introduction of $\mathcal{L}_2$ and their models can arguably be captured in our Paper II's static language. For instance, SVR's formula

$$\neg B_1(\langle \mathsf{U} \rangle(\mathsf{isarg}(b) \wedge \langle \mathsf{is\_attacked} \rangle \top) \wedge \neg B_1(\langle \mathsf{U} \rangle(\mathsf{isarg}(b) \wedge \neg \langle \mathsf{is\_attacked} \rangle \top)$$

informally expresses the same meaning as

[10] More precisely, we focus on $\mathcal{KD}45(\mathcal{A}o\mathcal{A})$-models, but we omit the $\mathcal{KD}45$ prefix in order to simplify notation.

[11] See the works by Arisaka et al. (2019b,a) for a similar property and for an interesting notion of multi-agent argumentation framework. We leave a detailed comparison to our approach for future work.

$$\neg\Box_1 \bigvee_{y\in\mathsf{A}} (y \rightsquigarrow b) \wedge \neg\Box_1\neg \bigvee_{y\in\mathsf{A}} (y \rightsquigarrow b).$$

The role of awareness in the usability of arguments, an intuition pointed out by SVR, could be thus formalised by establishing $\mathsf{aw}_i(a)$ as a global precondition for the event of agent i communicating a (just as we do in Paper I).

The second main difference that we find between $\mathcal{A}o\mathcal{A}$-models (o more in general $\mathcal{E}\mathcal{A}$-models) and $\mathcal{L}_2$-models is the possibility of assigning "atomic properties" to arguments, i.e., of assigning subsets of At to each $a \in \mathsf{A}_w$ through the function L_w. Differently to what happened before, we cannot find a simple way of overcoming this gap of expressivity.

Finally, the third main difference between $\mathcal{L}_2$ and the static language of Paper II is that the former uses several modalities to describe the structure of the underlying AF (following Grossi (2010b)), while we opted for a propositional encoding. As pointed out in Paper II, propositional logic does the job as long as we keep the set of all arguments being finite (something that is actually assumed by SVR for each A_w). Besides, the modal language present in ψ-formulas is only apt to capture some of the Dung (1995)'s semantics (stable and complete), as shown by Grossi (2010b), while our encoding captures them all.

$\mathcal{L}_3$ and its models. The language $\mathcal{L}_3$ of the third logic introduced by SVR is given by the following BNF:

$$\varphi ::= p \mid \mathsf{isarg}(a) \mid \mathsf{ownedby}(i) \mid B_i\varphi \mid [\mathsf{U}]\varphi \mid [\mathsf{attacks}]\varphi \mid [\mathsf{is\_attacked}]\varphi$$

where $p \in \mathsf{At}$, $a \in \mathsf{A}$ and $i \in \mathsf{Ag}$. Note that it has exactly the same operators as $\mathcal{L}_2$, but it is formulated without syntactic restrictions.

Formulas of $\mathcal{L}_3$ are interpreted at pointed world/argument models, which are nothing but an application of *product models* (Gabbay and Shehtman, 1998; Kurucz et al., 2003), that we have already used in the current chapter. Given a multi-agent doxastic frame $(\mathsf{W}, \mathcal{R})$, and an AF $(\mathsf{A}_r, \mathsf{R}_r)$ whose elements represent *real* or *existing* arguments and attacks among them, a *world/argument model* is a product model $(\mathsf{W}\times\mathsf{A}, \mathcal{R}, \mathsf{R}, L)$, where:

- (A, R) is an AF with $\mathsf{A} \subseteq \mathsf{A}_r \cup \{?_0, ?_1, ...\}$, and $\mathsf{R} \cap (\mathsf{A}_r \times \mathsf{A}_r) = \mathsf{R}_r \cap (\mathsf{A} \times \mathsf{A})$.
- $L : (\mathsf{W} \times \mathsf{A}) \to \wp(\mathsf{Ag} \cup \mathsf{At})$.

The truth conditions for formulas of $\mathcal{L}_3$ are:

$$\begin{array}{rll}
M,(w,a) \models p & \text{iff} & p \in L(w,a) \\
M,(w,a) \models \mathsf{isarg}(b) & \text{iff} & a = b \\
M,(w,a) \models \mathsf{ownedby}(i) & \text{iff} & i \in L(w,a) \\
M,(w,a) \models B_i\varphi & \text{iff} & \text{for all } u \in \mathsf{W},\ w\mathcal{R}_i u \text{ implies that } M,(u,a) \models \varphi \\
M,(w,a) \models [\mathsf{U}]\varphi & \text{iff} & \text{for all } x \in \mathsf{A}, \text{ we have } M,(w,x) \models \varphi \\
M,(w,a) \models [\mathsf{attacks}]\varphi & \text{iff} & \text{for all } x \in \mathsf{A},\ (a,x) \in \mathsf{R} \text{ implies } M,(w,x) \models \varphi \\
M,(w,a) \models [\mathsf{is_attacked}]\varphi & \text{iff} & \text{for all } x \in \mathsf{A},\ (x,a) \in \mathsf{R} \text{ implies } M,(w,x) \models \varphi
\end{array}$$

The notable semantic differences among our $\mathcal{EA}$-models and the structures that we have just defined make difficult a fine-grained comparison. However, we can point out that, just as in the previous case, the target example of SVR that motivates the introduction of $\mathcal{L}_3$ can be captured by Paper II's static language. More in detail, the $\mathcal{L}_3$'s formula

$$\langle \mathsf{U} \rangle \big(\mathsf{isarg}(a) \wedge \mathsf{ownedby}(i) \wedge B_i \langle \mathsf{attacks} \rangle (\mathsf{ownedby}(k) \wedge B_j \neg \mathsf{ownedby}(k)) \big)$$

informally expressing that "agent i owns a and believes that there exists an argument attacked by a which is owned by agent k, but agent j believes that this argument is not owned by k" can be rewritten in Paper II's language as

$$\mathsf{aw}_i(a) \wedge \Box_i \bigvee_{x \in \mathsf{A}} \big(a \rightsquigarrow x \wedge \mathsf{aw}_k(x) \wedge \Box_j \neg \mathsf{aw}_k(x) \big).$$

The approach by Sakama and Cao Son

In a couple of papers, Sakama and Cao Son (2019, 2020) (SCS, for the rest of this section) introduced and studied a novel formalism called *Epistemic Argumentation Frameworks* "as a means to integrate the beliefs of a reasoner with argumentation" or, more precisely, as a way of representing "different views of reasoners on the same argumentation framework". This aim is clearly related to the contributions of the first track of this thesis. Hence, let us first go through the main definitions of SCS, and then compare them with our papers. We adapt their notation in order to avoid excessive abuse and confusion with ours.

SCS build upon the *labelling-based* approach to argumentation semantics (Caminada and Gabbay, 2009), that we also used in Paper I. Let (A, R) be an AF, a *labelling* is a function $\mathfrak{L} : \mathsf{A} \to \{\mathsf{in}, \mathsf{out}, \mathsf{und}\}$. A labelling $\mathfrak{L}$ is *complete* iff for every $x \in \mathsf{A}$ we have:

(i) $\mathfrak{L}(x) = \mathsf{in}$ iff $\mathfrak{L}(y) = \mathsf{out}$ for every y that attacks x, and

(ii) $\mathfrak{L}(x) = \mathsf{out}$ iff there is an attacker of x, y such that $\mathfrak{L}(y) = \mathsf{in}$.

Furthermore, a complete labelling $\mathfrak{L}$ is said to be

- *stable* iff $\{x \in \mathsf{A} \mid \mathfrak{L}(x) = \mathsf{und}\} = \varnothing$;
- *grounded* iff $\{x \in \mathsf{A} \mid \mathfrak{L}(x) = \mathsf{in}\} \subseteq \{x \in \mathsf{A} \mid \mathfrak{L}'(x) = \mathsf{in}\}$ for any complete labelling $\mathfrak{L}'$;
- *preferred* iff there is no complete labelling $\mathfrak{L}'$ such that $\{x \in \mathsf{A} \mid \mathfrak{L}(x) = \mathsf{in}\} \subset \{x \in \mathsf{A} \mid \mathfrak{L}'(x) = \mathsf{in}\}$.

As we mentioned in Chapter 2, labelling-based and extension-based semantics can be proved to be equivalent for all the semantics considered in this thesis, so that each σ-labelling is associated with an σ-extension and vice versa.

Given an AF (A, R), its set of *labelled atoms* is defined as $\{\mathsf{in}(x), \mathsf{out}(x), \mathsf{und}(x) \mid x \in \mathsf{A}\}$. An *epistemic atom* is of the form $\Box\varphi$ or $\Diamond\varphi$, where φ is a Boolean formula over

the set of labelled atoms. An *epistemic literal* is an epistemic atom or its negation. An *epistemic formula*, is a propositional formula constructed over epistemic literals together with $\top$ and $\bot$. Hence, summing up, we have a modal language of labelled atoms with no nested epistemic operators.

Note that a labelling $\mathfrak{L}$ can be expressed as the set $\{\lambda(x) \mid x \in \mathsf{A}, \mathfrak{L}(x) = \lambda\}$, which is also a propositional valuation over the set of all labelled atoms. Therefore, Boolean formulas over the set of labelled atoms can be interpreted at labellings. Following this idea, epistemic formulas can be interpreted at *sets of labellings*. So, let SL be a set of labellings, SCS define the truth relation as usual for Boolean connectives, and give the following clauses for epistemic atoms:

$$\begin{array}{lll} SL \models \Box\varphi & \text{iff} & \text{for every } \mathfrak{L} \in SL,\ \mathfrak{L} \models \varphi, \\ SL \models \Diamond\varphi & \text{iff} & \text{for some } \mathfrak{L} \in SL,\ \mathfrak{L} \models \varphi. \end{array}$$

An *epistemic argumentation framework* (EAF) is a triple $(\mathsf{A}, \mathsf{R}, \varphi)$ where (A, R) is an AF and φ is an epistemic formula. In an EAF, (A, R) is meant to represent "objective evidence" while φ is a sort of "subjective belief", so that we can build different EAFs $(\mathsf{A}, \mathsf{R}, \varphi_1)$,..., $(\mathsf{A}, \mathsf{R}, \varphi_n)$ over the same AF for representing the views of n different agents. Finally, let σ denote any of the four Dung's semantincs, a *σ-epistemic labelling set* of an EAF $(\mathsf{A}, \mathsf{R}, \varphi)$ is a maximal (with respect to set inclusion) set of σ labellings of (A, R) that satisfies φ. σ-epistemic labelling sets provide a natural semantics for EAFs, as the epistemic constraint φ serves as a device to shrink the set of original possible labellings. The rest of the work of SCS –that we do not reproduce here– is aimed, among other things, to find out necessary and sufficient conditions for a σ-epistemic labelling set to be unique or non-empty; to find complexity results about the problem of finding a non-empty σ-epistemic labelling set for a given EAF; and to explain how EAFs can be extended to capture other relevant notions in argumentation, such as the accommodation of preferences among arguments.

We close this subsection by pointing out some of the main (non completely obvious) differences between SCS and our works Paper I, Paper II and Paper IV:

- Just as it happened with the approach of SVR, SCS include a complexity analysis that is absent throughout our contributions.
- Moreover, EAFs are essentially a static class of structures, while the inclusion of a theory of dynamics of information was one of our main desiderata when relating DEL and FA tools.
- The approach of SCS lacks an account of higher-order epistemic attitudes. This component of multi-agency is, as we have argued in several occasions, central to the study strategic behaviour in persuasive communication, that is one of our main interests.[12]

[12] We take advantage to correct the assertion made in Section 7 of Paper IV, where we said that SCS "do no take into account multi-agent scenarios", as they actually do face multi-agency in its first-order version (Sakama and Cao Son, 2019, Section 3.3).

- Finally, we point out what seems to be the main modelling assumption that differentiates both lines of work. SCS suppose that agents have total awareness/knowledge both of the set of all arguments and all attacks, and they add a subjective component through the inclusion of the epistemic constraint φ, whose final role is to filter out the set of labellings (and hence of arguments) accepted by the agent. On the other hand, our approach to multi-agency is rooted precisely in the different views that agents within a group may have of the set of all arguments (Paper I), attacks (Paper IV), or both (Paper II and Paper III). Roughly speaking, SCS's picture of multi-agency in AFs seems natural under the assumptions that (i) R represents an *objective* attack relation;[13] and (ii) agents are extremely good reasoners, as they do not only generate the set of all relevant arguments, but also spot correctly all the attacks among them. Under a set of assumptions questioning either (i) or (ii), we think that the option of modelling multi-agency in AFs through heterogeneous views of either arguments or attacks might be more appropriate.

The approach by Dyrkolbotn and Pedersen

Dyrkolbotn and Pedersen (2016) (DP, for the rest of this section) proposed to use a novel application of AFs to capture some of the essential features of Mercier and Sperber (2017)'s argumentative theory of reason.[14] This theory conceives reason as an irreducibly social skill that, seen from an individual perspective, tends to favour the ability of winning a debate, and not necessarily the one of reaching logically sound conclusions. DP is our main source of inspiration in Paper IV. Let us recall some of its main formal concepts and compare them to our work.

Multi-agency is represented by DP through the notion of an agent's view. Given a set of arguments A, *agent i's* view is an AF $(\mathsf{A}, \mathsf{R}_i)$. An *argumentative state* is a pair $\mathcal{S} = (\mathsf{A}, \{\mathsf{R}_i\}_{i\in\mathsf{Ag}})$, where each $(\mathsf{A}, \mathsf{R}_i)$ is an agent's view. A *deliberative state* over $\mathcal{S}$ is any $\mathsf{R}^d \subseteq \mathsf{A} \times \mathsf{A}$ such that

$$\bigcap_{i\in\mathsf{Ag}} \mathsf{R}_i \subseteq \mathsf{R}^d \subseteq \bigcup_{i\in\mathsf{Ag}} \mathsf{R}_i.$$

We note $\mathsf{d}(\mathcal{S})$ the set of all deliberative states over $\mathcal{S}$. The possible ways in which a debate may evolve through time are formally captured in DP's approach by the notion of argumentative model. An *argumentative model* over $\mathcal{S}$ is any $AM = (Q, R)$, where:

(i) $Q \subseteq \mathsf{d}(\mathcal{S})$, and

(ii) $R \subseteq Q \times Q$.

Intuitively, $(\mathsf{R}_1^d, \mathsf{R}_2^d) \in R$ means that there is an event that can transform the deliberative state R_1^d into R_2^d.

[13] Therefore excluding the interpretation of R as a *defeat* relation that naturally differs among different agents.

[14] DP use several modal extensions of Łukasiewizc's three-valued logic to encode their semantic constructs. But this is somehow orthogonal to our purposes, so we stay on a semantic dimension here.

We proceed to point out some salient differences between DP and our approach in Paper IV.

- Argumentative states can be seen as a type of multi-agent AF, where all arguments are commonly known and there is nothing like an objective attack relation (see Section 9 of Paper II for other possible modelling assumptions). This is an essential difference with our approach of Paper IV, where such an objective attack relation is used to account for the notion of *knowledge of attacks* (since known attacks are required to be true).

- Moreover, as it was pointed out in Section 3.3 of Paper IV, our notion of *public view* of an AF can be compared to that of a deliberative state. While the fact that an attack pair (a, b) is accepted by all agents is a sufficient condition to be part of any deliberative state, this is only a necessary condition to be part of a public view. Taking into account the differences among approaches, we wonder why it should be taken for granted that an attack perceived by everyone is always part of the *common ground* of a conversation (formalised in DP by the state of a debate). In other words, we may agree that a and b are incompatible, but this is not part of the state of the debate because no one has said it and, for instance, it can be the case that I'm not sure if you accept such incompatibility.

- Just as it happened with SCS, DP lacks an account of higher-order epistemic attitudes, which is one of our main focus in Paper IV.

- The final point concerns dynamics. DP remains abstract about the kind of events that, within an argumentative model, provokes the transition between one deliberative state and another (formally captured by the relation R). In Paper IV, we give a precise characterization of one suitable type of these events: publicly announcing parts of the attack relation. Nevertheless, DP sketch how to semantically and syntactically isolate classes of deliberative models that satisfy some desired properties, and how these can be studied from a more fine-grained perspective using more expressive modal languages. The combination of this idea with our framework (for instance, through the inclusion of modalities from the literature quantifying over public announcements (see e.g., (Balbiani et al., 2008)) seems an interesting path for future research.

Incomplete and control argumentation frameworks

We close this section by briefly recalling and highlighting the close relation of our contributions to *incomplete argumentation frameworks* (Baumeister et al., 2021, 2018a,c,b; Fazzinga et al., 2020), and *control argumentation frameworks* (Dimopoulos et al., 2018; Niskanen et al., 2020).[15]

[15]For a study of these and other formalisms for modelling qualitative uncertainty about AFs the reader is referred to Mailly (2021a); Herzig and Yuste-Ginel (2021a).

In Paper II, this relation is made explicit in sections 8.1 and 8.2 respectively, where our modal language and semantics are shown to be expressive enough to subsume both formalisms as special cases, as well as to point out some interesting assumptions underlying both kind of structures.

The question of revealing what is *the* precise epistemic logic underlying incomplete argumentation frameworks is left open in Paper II, and it constitutes the main research question of Paper III.

Finally, the idea of combining the multi-agent AFs with incomplete knowledge of attacks of Paper IV with the uncertainty captured by the partial AFs of Cayrol et al. (2007) (a studied subclass of incomplete AFs) is sketched in Section 6.2 of Paper IV. We leave a detailed development of this idea for future work.

5.2 Logics of argument-based beliefs

Introducing the missing ingredient in epistemic models. As exposed in the introduction to Paper VI, most formal models of belief (including both quantitative and qualitative approaches) lack an *evidence* or *justification* component. This lack seems indeed relevant, since those terms (*evidence* and *justification*) have played a central role along the history of epistemology.[16] The second track of research that constitutes the core of this thesis (Section 4.2) can be framed within the relatively recent trials of the epistemic logic community to include this missing ingredient into their formal models of epistemic attitudes. Although sharing the same starting point, these approaches are strongly heterogeneous with regard to the methodology that they use. A non-exhaustive classification is given by the following list:

- The first well-known formal approach to justified belief consists in the combination of *justification logic* (Artemov and Fitting, 2016) and epistemic logic, sometimes including awareness tools too. Examples of this enterprise are the works by Artemov and Nogina (2005); Baltag et al. (2012, 2014) and Artemov (2018, 2020).

- A second line of research, initiated by van Benthem et al. (2012, 2014), is based on the idea that pieces of evidence can be mathematically represented as sets of possible worlds. This equation naturally leads to the use of *neighbourhood semantics* for modal logic (Pacuit, 2017). Moreover, some technical and conceptual problems that arise in this approach (prominently, the lack of consistency in some scenarios where the agent is aware of an infinite set of pieces of evidence) motivated the introduction of topological tools on top of these neighbourhood models (Baltag et al., 2016; Özgün, 2017), originating another fruitful sub-branch in the literature.

- It is worth mentioning by passing that, maybe more marginally, evidence has also been introduced in epistemic models through the use of multi-valued semantics (e.g., in Santos (2017)).

[16] See Kelly (2016) and Pappas (2017) for philosophical surveys on both terms.

- Yet a fourth line of research can be categorised as the introduction of argumentative tools within epistemic models in order to provide an *argument-based* version of the notion of justified belief. Within this line, we can highlight the works by Shi et al. (2021), the one by Grossi and van der Hoek (2014), the one by Wáng and Li (2021), and our contributions in Paper V and Paper VI.

In this section, we will analyse in some detail the last line of research in order to compare it with Paper V and Paper VI, devoting moreover some additional words to our adaptations of ASPIC$^+$ notions within both works. But, before that, let us briefly comment on the relation among both contributions.

5.2.1 Relation among contributions

The task of comparing the contributions of this second track is strongly simplified with respect to the previous one, as we only have two papers this time, and one of them (Paper VI) is the direct continuation of the other one (Paper V). Besides the differences that are obvious after reading both works, and the ones mentioned in the introduction of Paper VI, we make a couple of additional points that were left out of the final version due to lack of space.

First, in Paper V, basic beliefs are used to establish a three-layer preference order among arguments. In contrast, in Paper VI, we decided to simplify such an order, by splitting the awareness set of the agent into two parts: one containing doxastically accepted arguments (those with all premisses being believed), and another one that fails to have the previous property. Furthermore, only doxastically accepted arguments are included in the structured AF that later on is used to define the notion of argument-based beliefs. Both options are different formal instantiations of $\mathsf{C1}_{\mathsf{epistemic}}$, but the main difference is that, with the more coarse-grained ordering of Paper VI, undermining attacks do not make sense, as they always fail due to the fact that the attacker is never going to be at least as preferred as the targeted premiss. In a similar informal setting (that is, not considering stratified sets of premisses), Pandžić (2021) has recently argued for a form of undermining that one may qualify as 'non-inferential', based on the importation of tools from the belief change literature. In our setting, this form of undermining is partially captured through the public announcement operator introduced in Paper VI.

The second main difference is rather technical. While the $\mathcal{L}_{\mathsf{BA}}$-models of Paper V were assumed to have *finite* awareness sets, as well as *finite* sets of accepted defeasible rules, these assumptions are dropped in Paper VI. The rationale for doing so is double: we obtain a more general logical framework, and we simplify the completeness proof.

5.2.2 Closely related work

Let us proceed to present some of the details of the closely related works that were introduced above, so as to compare each of them with Paper V and Paper VI.

The approach by Shi, Smets and Velázquez-Quesada

In a series of papers, Shi et al. (2017, 2018, 2021) and Shi (2021) undertook the task of combining the work on the notion of *topological (justified) belief* (Baltag et al., 2016; Özgün, 2017) with the AFs of Dung (1995), in order to provide an argumentatively informed refinement of the former.[17] Let us first present the main semantic construct introduced in these papers: *topological argumentation models*, as well as two different kinds of argument-based belief that can be defined over these models: *grounded belief* and *fully grounded belief*. We follow the definitions of Shi et al. (2021), adapting them to our notation and terminology, and refer to the whole approach as SSV-Q for the rest of this section.

A preliminary needed step is to have in mind what a topology is. A *topology* τ over a non-empty set W is a collection of subsets of W, in symbols $\tau \subseteq \wp(W)$, such that: (i) it contains the empty set and the unit ($\varnothing, \mathsf{W} \in \tau$); (ii) it is closed under finite intersections (if $A, B \in \tau$, then $A \cap B \in \tau$); and (iii) it is closed under arbitrary unions (for any –possibly infinite– family $\{A_x\}_{x \in X} \subseteq \tau$, we have that $\bigcup_{x \in X} A_x \in \tau$). Moreover, the topology generated by a family of sets $B \subseteq \wp(\mathsf{W})$ is the smallest topology τ_B such that $B \subseteq \tau_B$. Elements of a topology τ are usually called *opens*.

A *topological argumentation (TA) model* for a countable set of atomic variables At is a tuple $M = (\mathsf{W}, E_0, \tau_{E_0}, \mathsf{R}, \mathsf{V})$, where

- $\mathsf{W} \neq \varnothing$ is a set of *possible worlds*.
- $E_0 \subseteq \wp(\mathsf{W}) \setminus \{\varnothing\}$ is a collection of *basic pieces of evidence*.
- τ_{E_0} is the topology generated by E_0.
- $\mathsf{R} \subseteq \tau_{E_0} \times \tau_{E_0}$ is an *attack relation* satisfying:
  - for every $A \in \tau_{E_0} \setminus \{\varnothing\}$, we have $(A, \varnothing) \in \mathsf{R}$ and $(\varnothing, A) \notin \mathsf{R}$.
  - for every $A, B \in \tau_{E_0}$, we have $A \cap B = \varnothing$ iff either $(A, B) \in \mathsf{R}$ or $(B, A) \in \mathsf{R}$.
- $\mathsf{V} : \mathsf{At} \to \wp(\mathsf{W})$.

Let us look at some intuitions underlying TA models. The first novel component with respect to the epistemic models that we have been using so far is the collection of basic pieces of evidence $E_0 \subseteq \wp(\mathsf{W}) \setminus \{\varnothing\}$. The idea of semantic modelling pieces of evidence as sets of possible worlds can be traced back to van Benthem et al. (2012, 2014). Following Shi et al. (2021), the main assumption captured by this way of modelling evidence is that *evidence is understood as information-as-range*, so that if W represents all the epistemic alternatives of the formalised agent, a piece of evidence $A \subseteq \mathsf{W}$ is telling the agent that the actual world is in A (and hence $\mathsf{W} \setminus A$ should be disregarded according

[17] See also the PhD thesis of Shi (2018), where this line of work is originated.

to A). Note, however, how $A \in E_0$ does not informally mean that the agent accepts A, but she rather takes it as a starting point for reasoning.

The next component of TA models, τ_{E_0}, is imported from the work on topological justified belief (Baltag et al., 2016; Özgün, 2017). Following Shi et al. (2021), this amounts to assuming that evidence is *affirmative* information, in the sense that it is verifiable when it is true, and it is true precisely when it can be affirmed. Perhaps more intuitively, τ_{E_0} represents the possible ways in which the agent can logically combine her basic pieces of evidence. Importantly, the elements of τ play the role of *arguments* in SSV-Q (an idea already anticipated by Özgün (2017)).

Finally, R represents an attack relation among the elements of τ_{E_0} or, according to our terminology, a *defeat* relation (since, the fact that $(A, B) \in \mathsf{R}$ and $(B, A) \notin \mathsf{R}$, if considered in isolation, means both that A and B are incompatible and that A is strictly stronger than B).[18] Hence, following Shi et al. (2021), the introduction of R (which is actually the new component of TA models with respect to the topological models of Baltag et al. (2016); Özgün (2017)) functions as a way of modelling how contradictory pieces of evidence are weighted. This process of evaluation –and this is a central point– is *not* modelled in an order theory fashion, by simply establishing a preference relation among arguments, but rather through the conflict calculus introduced by Dung (1995).

We recall the two forms of argument-based belief that are defined over TA models (Shi et al., 2021).[19] Let $M = (\mathsf{W}, E_0, \tau_{E_0}, \mathsf{R}, \mathsf{V})$ be a TA model, and let $P \subseteq \mathsf{W}$ be a proposition, then:

- the agent has a *grounded belief* on P iff there is an $A \in \mathsf{gr}(\tau_{E_0}, \mathsf{R})$ such that $A \subseteq P$.
- the agent has a *fully grounded belief* on P iff for every $A \in \mathsf{gr}(\tau_{E_0}, \mathsf{R})$, there is an $A' \in \mathsf{gr}(\tau_{E_0}, \mathsf{R})$, such that $A' \subseteq A$ and $A' \subseteq P$.[20]

Curiously, while fully grounded beliefs satisfy KD45 axioms, grounded beliefs fail to be closed under conjunction (and hence do not satisfy the K axiom). Moreover, pairwise consistency among groundly believed propositions is guaranteed, but this is not the case when we consider sets of groundly believed propositions with more than two elements. In terms of the literature about rationality postulates for argumentation systems, grounded belief is directly consistent but not indirectly consistent. In contrast, fully grounded belief is indirectly (i.e., totally) consistent. Finally, grounded belief is strictly stronger than fully grounded belief, so that the former implies the latter but not vice versa. This brief comparison among both notions makes explicit the existing tension between *believing more* (or more informatively) and believing *more consistently* (see Shi (2018) and Shi et al. (2021) for a detailed discussion on such a tension).

[18] See Chapter 2 for more discussion on the difference between attacks and defeats.

[19] There are, certainly, other interesting epistemic attitudes that are definable over TA models, but we decided to leave them out from our analysis either because they are characterised in terms of knowledge (and our focus is more on belief), or because they are not argument-based notions.

[20] Recall from Chapter 2, that $\mathsf{gr}(\mathsf{A}, \mathsf{R})$ denotes the *grounded extension* of (A, R). We also remark that both notions of argument-based beliefs are originally defined using the alternative characterization of the grounded extension as the least fixpoint of the defence function.

When it comes to comparison, there is a central line of contrast between the SSV-Q approach and our contributions of this track (Paper V and Paper VI): their approach is essentially semantic, while ours is strongly oriented towards syntactic representations. This difference is prominently made explicit in the two diverse ways of modelling arguments: while they are members of a topology in SSV-Q, ours are syntactic objects of a given formal language. The semantic approach enables a clean and informative axiomatisation of the grounded belief operator, while it is far for being clear if any of our two argument-based belief operators is axiomatisable at all.[21] On the other hand, both grounded belief and fully grounded belief suffer from a weak form of omniscience: closure under equivalent formulas. In other words, if an agent (fully) groundly believes that φ, then she also believes every sentence ψ that is logically equivalent to φ. This fails to capture that, as far as real agents are concerned, the step from believing that φ to believing that ψ is usually done through (argumentative) reasoning.[22] Finally, the semantic nature of SSV-Q's arguments makes non-trivial the question of determining what is the structure of a given $A \in \tau_{E_0}$ (so as to give an account, for instance, of different kind of attacks among arguments). More in particular, it is legitimate to ask, what are the premisses of A? And what are its conclusions? As plausible answers, one may set $\mathsf{Prem}(A) = \{e \in E_0 \mid e \subseteq A\}$; and $\mathsf{Conc}(A) = \{P \subseteq \mathsf{W} \mid A \subseteq P\}$, but alternative definitions seem also defensible.

Another important difference among both approaches is that although SSV-Q consider different and subtle forms of justified belief (either in its topological or argumentative version), they do not take into account basic beliefs. This implies that, while SSV-Q provide a rich account of $\mathsf{C2}$, they abstract away from $\mathsf{C1}_{\mathsf{epistemic}}$. That partially explains the fact that the defeat relation among arguments in TA models is primitive (manually designed by the modeller), while it emerges from the combination of argumentative attacks and basic doxastic attitudes towards premisses in our approach.

Finally, although SSV-Q develop a deep analysis of the mentioned notions (and more), it is a *static* account. In contrast, we took the first steps towards the dynamization of our framework in Paper VI. A pending task for both lines of work is jumping from the single-agent case to multi-agent scenarios.

The approach by Wáng and Li

In a couple of works, Li and Wáng (2020); Wáng and Li (2021) (WL, for the rest of this section) introduced a logic in which knowledge is defined as a kind of "true belief that has an appropriate argument". This aim is clearly inserted within the tradition that formally studies JTB theories of knowledge, i.e., theories that define knowledge as *Justified True Belief*.[23] In this case, the notion of justification is instantiated as having "an appropriate argument". Although the overall objective of this enterprise differs sensibly with ours, we

[21] The axiomatisation of the SSV-Q's operator capturing fully grounded belief is still an open problem too.

[22] Think for instance of relatively complex instances of φ and ψ.

[23] As it is well-known, Gettier (1963) and his famous counter-examples questioned the appropriateness of this group of theories (whose origin can be traced back to Plato), originating an enormous literature about the topic. Although central to contemporary epistemology, the question is just tangentially related to our interests here, so we refer interested readers to Ichikawa and Steup (2018).

can find some interesting similarities. Let us first present the basics of their work.

Let At be a countable set of propositional variables, the language $\mathcal{L}_{LK}$ is the bimodal language generated by the following BNF:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid B\varphi \mid K\varphi \quad p \in \mathsf{At},$$

where B stands for belief, and K stands for knowledge. We use $\mathcal{L}$ to denote the propositional fragment of $\mathcal{L}_{LK}$.

$\mathcal{L}_{LK}$-models are tuples of the form $M = (\mathsf{W}, \mathcal{R}, S, \mathsf{V})$, where W, $\mathcal{R}$ and V are just as in standard doxastic models (thus $\mathcal{R}$ is serial, transitive and euclidean), and

$$S : \mathsf{W} \to \wp(\wp(\wp(\mathsf{W})))$$

is an *argumentation function*, assigning a set of sets of sets of possible worlds to each world. Intuitively, S assigns a set of *semantic arguments* to each world (in our terminology, the set of arguments that the agent is aware of). This representation of arguments is an extrapolation of ideas that are present in the neighbourhood semantics literature (Pacuit, 2017). In the latter, it is usually assumed that a neighbourhood function

$$N : \mathsf{W} \to \wp(\wp(\mathsf{W}))$$

assigns a set of *propositions* or *statements* to each world.[24] In epistemic terms, N assigns to w the set of propositions that are known or believed by the agent at w. Therefore, the lift of the range of the argumentation function is natural if one understands that "an argument is in some sense a collection of statements" (Wáng and Li, 2021). Formulas of $\mathcal{L}_{LK}$ are interpreted in pointed $\mathcal{L}_{LK}$-models; we just reproduce the truth clauses for the modal operators:[25]

$$\begin{array}{lll} M, w \models B\varphi & \text{iff} & w\mathcal{R}u \text{ implies } M, u \models \varphi \\ M, w \models K\varphi & \text{iff} & M, w \models \varphi \wedge B\varphi \text{ and there is an } X \in S(w) \text{ such that} \\ & & [[\varphi]]_M \in X \text{ and for every } \psi \in \mathcal{L} \text{ such that} \\ & & [[\psi]]_M \in X, \text{ we have that } M, w \models B\psi. \end{array}$$

Hence, belief is treated as a normal modal operator here, while knowledge is defined according to the informal equation that we have already mentioned: knowledge is true belief that has an appropriate argument. Interestingly, based on this semantic interpretation, we can claim that:

(i) the truth clause of K is an instantiation of $\mathsf{C2}$ (in particular, *argument evaluation conditions knowledge acquisition*); and that

[24] See Stalnaker (1976) for a philosophical analysis of the understanding of propositions as sets of possible worlds.

[25] Just as in standard epistemic models, $[[\varphi]]_M$ denotes the *truth-set of* φ with respect to M, that is, the set of worlds of M where φ is true.

(ii) the part of the truth clause of K capturing the notion of "appropriate argument" is indeed a semantic instantiation of $\mathsf{C1}_{\mathsf{epistemic}}$, so that arguments with believed propositional premisses are the only appropriate ones (i.e., the ones that are preferred over the rest).

Hence, it is tempting to extrapolate the distinction between basic beliefs and argument-based beliefs to WL's approach, by dropping the factive requirement in the truth clause of $K\varphi$, so that K becomes an argument-based belief operator. Besides this point of similarity, that speaks in favour of the intuitive appeal of $\mathsf{C1}_{\mathsf{epistemic}}$ and $\mathsf{C2}$, there exist obvious differences with our works. Just as it happens with the approach of SSV-Q, WL is strongly semantic, so most of our comments there can be applied also here. Interestingly, there is a point that aligns our works with those of SSV-Q and put them jointly in contrast with WL: the latter approach lacks the conflict calculus that, since the work of Dung (1995), is almost ubiquitous in the field of formal argumentation. Thus, although the elements of the range of S are called *arguments*, they could well be interpreted as other informal, more general notions (e.g., as *evidence*).

The approach by Grossi and van der Hoek

Grossi and van der Hoek (2014) (GvdH, for the rest of this section) authored one of the pioneer works on combining EL and FA in order to study the interactions between beliefs and argumentation. The main idea consists in using *product models* (Gabbay and Shehtman, 1998; Kurucz et al., 2003), where one of the relations represents doxastic accessibility and the other one represents argumentative attacks (just as in the third logic introduced by Schwarzentruber et al. (2012)). Let us introduce the more refined class of models they studied, as well as three different notions of argument-based beliefs that can be captured by these models.

A *doxastic structure* is a pair $(\mathsf{W}, \mathcal{B})$ with $\mathcal{B} \neq \varnothing$ and $\mathcal{B} \subseteq \mathsf{W}$. $\mathcal{B}$ represents a set of doxastically indistingisible worlds for the agent. We recall that, for the single agent case, these structures are modally equivalent to doxastic Kripke frames (see also our completeness proof in Paper VI). An *enriched AF*, is a tuple $(\mathsf{A}, \mathsf{E}, \mathsf{R})$ where (A, R) is an AF, and $\mathsf{E} \subseteq A$ is a non-empty set of *endorsed arguments*, that is, arguments that the agent accepts in some sense. Models of the logic DA (doxastic-argumentative logic) are tuples of the form $(\mathsf{A} \times \mathsf{W}, \mathsf{E}, \mathcal{B}, \mathsf{R}, \mathsf{V})$, where $(\mathsf{W}, \mathcal{B})$ is a doxastic structre, $(\mathsf{A}, \mathsf{E}, \mathsf{R})$ is an enriched AF, and $\mathsf{V} : \mathsf{At} \to \wp(\mathsf{W} \times \mathsf{A})$ is an atomic valuation. These models are described with the following multi-modal language $\mathcal{L}_{DA}$:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Box_B\varphi \mid \Box_A\varphi \mid \Box_E\varphi \mid [\mathsf{U}]_B\varphi \mid [\mathsf{U}]_A\varphi \quad p \in \mathsf{At}.$$

The duals of the modal operators above are noted $\Diamond_*$ with $* \in \{B, A, E\}$, and $\langle \mathsf{U} \rangle_\star$ with $\star \in \{B, A\}$, and defined as usual. Formulas are interpreted in pointed models, with the following truth clauses for the modal operators:

$$
\begin{array}{rll}
M,(w,a) \models \Box_B \varphi & \text{iff} & \text{for all } v \in \mathcal{B},\ M,(v,a) \models \varphi \\
M,(w,a) \models \Box_A \varphi & \text{iff} & \text{for all } b \in \mathsf{A},\ (b,a) \in \mathsf{R} \text{ implies } M,(w,b) \models \varphi \\
M,(w,a) \models \Box_E \varphi & \text{iff} & \text{for all } b \in \mathsf{E},\ M,(w,b) \models \varphi \\
M,(w,a) \models [\mathsf{U}]_B \varphi & \text{iff} & \text{for all } v \in \mathsf{W},\ M,(v,a) \models \varphi \\
M,(w,a) \models [\mathsf{U}]_A \varphi & \text{iff} & \text{for all } b \in \mathsf{A},\ M,(w,b) \models \varphi
\end{array}
$$

Moreover, it is assumed that there is a distinguished atom $\sigma \in \mathsf{At}$ such that $M,(w,a) \models \sigma$ informally means that argument a supports the world w, in the sense that Anne's colleague argument's "the sky looked cloudless when I came here [therefore it is not raining]" supports all worlds where the atomic sentence "it is raining" is false.

GvdH define three different types of argument-based beliefs in $\mathcal{L}_{DA}$:

$$
\begin{aligned}
SB\varphi &= \Box_B([\mathsf{U}]_A \varphi \wedge \langle \mathsf{U} \rangle_A \sigma), \\
EB\varphi &= \Box_B([\mathsf{U}]_A \varphi \wedge \Diamond_E \sigma), \\
JB(\varphi,\psi) &= \Box_B([\mathsf{U}]_A \varphi \wedge \Diamond_E(\sigma \wedge [\mathsf{U}]_B \psi)).
\end{aligned}
$$

$SB\varphi$ represents *supported belief* on φ, that is, φ is true in all doxastic alternatives, no matter what argument is entertained, and there are (possibly different) arguments supporting each of these doxastic alternatives. $EB\varphi$ represents *endorsed supported belief*, that is, φ is true in all doxastic alternatives, no matter what argument is entertained, and there are (possibly different) *endorsed* arguments supporting each of these doxastic alternatives. Finally, $JB(\varphi,\psi)$ represents *belief on φ justified by ψ*, that is, φ is true in all doxastic alternatives, no matter what argument is entertained, and there are (possibly different) *endorsed* arguments satisfying ψ that moreover support each of these doxastic alternatives. Typically, one may want ψ to be some argumentation-theoretic concept, such as "belonging to a complete extension". As we have already mentioned several times, Grossi (2010b) showed that the modal fragment containing the operators $\Box_A$ and $[\mathsf{U}]_A$ suffices to capture many of these concepts.

Clearly, justified belief is strictly stronger than endorsed belief, and this is, in turn, strictly stronger than supported belief. Moreover, all of these argument-based beliefs notions are $KD4$ operators but, curiously, they fail to satisfy the negative introspection principle (axiom 5) in the general case (a property that has sometimes been criticised for being too ideal for belief).

Regarding the comparison to the approach presented in Paper V and Paper VI, we should first remark again the semantic character of all the central notions of GvdH, in contrast with our syntactically flavoured system. This provokes, just as in the case of SSV-Q, the reintroduction of omniscience for all argument-based belief operators. Finally, as pointed out by Shi et al. (2021) (Footnote 3), in GvdH there are no direct dependencies between belief and argumentation as they are both taken to be independent, primitive dimensions, whose relations are studied *a posteriori*; while the study of some of these dependencies (C1 and C2) has been the main focus of the current thesis.

The approach by Modgil and Prakken

Although external to the epistemic logic tradition that articulates our discussion, we believe appropriate to sum up here the main features of our use of the structured argumen-

tation formalism ASPIC$^+$, developed by Modgil and Prakken (2013).[26]

- In contrast to ASPIC$^+$, preference among arguments is not a primitive notion in Paper V and Paper VI. Instead, it is a concept that is derived from the agent's doxastic attitudes toward the premisses of the involved arguments as well as from the reliability of the involved inference links. This is technically done by using an adaptation of awareness epistemic models (Definition 5) to our purpose. Conceptually, the partial reduction of preference to doxastic acceptability amounts to formally capturing $\mathsf{C1}_{\mathsf{epistemic}}$.

- In Paper V and Paper VI, we somehow build a logical meta-theory of an instantiation of ASPIC$^+$, where arguments are first-class citizens of our object language. As usual in logic, we let any string of a certain syntactic shape be an argument, as the definition of the language cannot take into account model-based features (e.g., accepted defeasible rules). This requires introducing the distinction between arguments in general (the objects of our language) and *well-shaped* arguments (a subset of the former, those that simulate ASPIC$^+$ arguments and whose existence is model-dependent). Such a distinction is formally captured in the semantics of our logic and recursively axiomatized in Paper VI.

- As we think of agents with bounded resources, we drop a central assumption in ASPIC$^+$: agents do not always build all arguments from a given argumentation theory. This is indeed not an innocent assumption since, as far as the set of inference rules contains, for instance, classical logic, the set of arguments is infinite. As one might expect, dropping this assumption has important effects on the rationality of the formalised agent. We analyse these effects under several conditions and provide some others in order to increase it gradually.

- As initially formulated by Modgil and Prakken (2013), ASPIC$^+$ is a static formalism. In Paper VI we provide a dynamic account of our adaptation of it, importing tools from dynamic epistemic logic and the dynamic of awareness literature.

We close this chapter by mentioning two lines of work that have been fundamental sources of inspiration for our approach, but that we believe to have commented enough within each contribution. First, the two papers on epistemic justification logic by Baltag et al. (2012, 2014). Our way of defining doxastic acceptance in Paper V and Paper VI is a direct extrapolation and generalization of the definition given by them. Moreover, their work can be understood as a joint approach to $\mathsf{C1}_{\mathsf{epistemic}}$ and $\mathsf{C2}$ from a justification logic perspective. The main conceptual difference is that they restrict their attention to deductive arguments with doxastically accepted premisses, and thus lack the need of argumentation theoretic concepts.[27] The second one are the papers on dynamic of syntactic

[26] For the unfamiliar reader, the basics of this system were introduced in Section 2.2.2.

[27] In this sense, our notion of *deductive belief* (Paper VI) is very close to their notion of justified belief (Baltag et al., 2012).

awareness of Grossi and Velázquez-Quesada (2009, 2015). As we pointed out several times in Paper VI, these works are the main antecedent used to develop the dynamization of our framework.

Chapter 6

Conclusion

Let us recapitulate what has been done so far, as a way to start closing this dissertation. In few words, we have aimed at the formal study of some of the existing relations between epistemic attitudes and argumentative phenomena –with a special focus on belief formation and argument evaluation, respectively. This enterprise has been developed through the joint use of mathematical and conceptual tools imported from two relatively disconnected research areas: (dynamic) epistemic logic and formal argumentation (abbreviated, respectively, as (D)EL and FA).

Conceptually speaking, we have focused on the examination of two very general principles governing belief formation and argument evaluation, namely, that the latter is somehow conditioned by the former (C1), and vice versa (C2). Both principles have been informally analysed in Chapter 1. There, C2 was unequivocally interpreted: in order to believe that φ, to believe that not φ, or none of them, a rational agent must take into account her arguments pro and against φ, and how strong these arguments are. However, C1 (*argument evaluation is affected by belief formation*) has been analysed according to two different intuitive readings. On the one hand, it is clear that agents assess arguments with believed premisses as being strictly stronger than arguments with premisses that are not believed ($\mathsf{C1}_{\mathsf{epistemic}}$), at least regarding epistemic reasoning. On the other hand, it also seems fair to claim that higher-order epistemic attitudes (for instance, "I believe that my opponent is aware of a counterargument to a, but not to b") may condition the persuasive strength that I concede to certain arguments ($\mathsf{C1}_{\mathsf{rhetoric}}$). The technical tools used in our analysis of C1 and C2 have been reviewed in Chapter 2, where we also gave some hints about further reading. Before reprinting the contributions that constitute the core of this thesis (Chapter 4), we have explained how they jointly represent an approach to C1 and C2 (Chapter 3). Finally, in the previous chapter, we have (i) analysed in detail the technical and conceptual connections and differences among the different contributions, and (ii) compared our approaches to closely related ones.

In the current chapter, we aim at pointing out paths for future work, since any research project is ultimately open. But, before that, we would like to highlight two of our central findings.

Our first remark concerns the protagonist role that the notion of awareness has played

in our enterprise of building an epistemic formal theory of argumentation, with the notable exception of Paper IV. In order to make possible a clear connection among EL and FA, we have "jumped" from the notion of *awareness of sentences à la* Fagin and Halpern (1987) to the one of *awareness of arguments*. All in all, we have shown that awareness preserves one of its most desirable properties when applied to arguments: its extreme flexibility. This appealing aspect seems to be due to its syntactic nature: the range of awareness functions in our epistemic argumentative models is still the power set of a set of *syntactic entities* (arguments); or, at least, of atomic arguments which are syntactically represented. This permits requiring further constraints that gradually directs the merely attentional reading of awareness towards more involved epistemic attitudes that we might qualified as *knowledge of arguments*. Moreover, we have also shown how the idea of *awareness of arguments* can be applied to quite different purposes in its abstract version (Paper I, Paper II, and Paper III) and in its structured refinement (Paper V and Paper VI). As for the first one, it enables, complemented with the notion of *knowledge of attacks* (Paper IV), a very general theory of multi-agency and qualitative uncertainty about argumentation frameworks, making possible in turn an explanation of the strategic behaviour of epistemic agents involved in a debate (and thus digging deeper in the study of $\mathsf{C1}_{\mathsf{rhetoric}}$). An important witness of this achievement is the reduction of existing formalisms for arguing with qualitative uncertainty to our more general framework (Section 8 of Paper II and Paper III). As to the notion of *awareness of structured arguments*, it serves to draw a syntactic alternative to the existing semantic approaches to argument-based beliefs. As expected, the more fine-grained nature of our syntactic approach came at the price of a relative loss of ideal rationality in the formalised agents, at least in the basic setting.

Our second point is articulated around the analysis of the very notion *argument strength* as performed by Beirlaen et al. (2018), that split this notion into three tiers or dimensions: the supportive one, the dialectical one and the evaluative one.[1] Our contributions can be seen as depicting a *fourth tier of argument strength*: the epistemic one. This dimension is, however, transverse to all the other three, so that we could maybe refer to it as the *epistemic aspects* of argument strength. More concretely, the epistemic aspect of the *support dimension*, can be synthesized in $\mathsf{C1}_{\mathsf{epistemic}}$ (developed in Paper V and Paper VI). Regarding the *dialectical dimension*, that is, how arguments attack, defeat and defend each other, epistemic attitudes of arguers about how other agents perceive these dialectical relations seem to play a fundamental role in the understanding of persuasive argumentation, as we have argued through the contributions of the first track of this thesis.[2] Finally, and regarding the *evaluative dimension*, we have not directly dealt with it from an epistemic point of view, but it is easy to imagine agents attributing different argumentation semantics to other agents, and how these attributions, in turn, partially determine their moves within an argumentative dialogue.

[1] See Section 1.2 for a brief exposition of this analysis.

[2] This is the same intuition motivating the use of *opponent models* in the field of strategic argumentation (Thimm, 2014). Along analogous lines, and within his analysis of philosophical problems concerning argumentation that can be clarified through the use of formal methods, Prakken (2011) claimed that "the persuasive force of arguments may depend on the listener".

It is now time to present some lines for future and ongoing research, besides the ones that we already pointed out at the end of each contribution.

The first open problem concerns the introspection of our versions of the notion of *argument-based belief*, something that we avoided discussing in both Paper V and Paper VI. First of all, note that although the enriched language of Paper V enables the expression of introspective properties, the argument-based belief operator is not introspective, since both $\mathsf{B}^{\mathsf{BA}}\varphi \wedge \neg \mathsf{B}^{\mathsf{BA}}\mathsf{B}^{\mathsf{BA}}\varphi$ and $\neg \mathsf{B}^{\mathsf{BA}}\varphi \wedge \neg \mathsf{B}^{\mathsf{BA}}\neg \mathsf{B}^{\mathsf{BA}}\varphi$ are satisfiable in our class of models. In Paper VI, where our argument-based belief operator is binary and explicitly talks about the argument in which the current belief is grounded, we do not allow expressions of the shape $\mathsf{B}(\alpha, \mathsf{B}(\beta, \varphi))$ to be formulas of the language, so that introspection is not even expressible. In contrast to semantic accounts of argumentative beliefs (e.g., Shi et al. (2021)), we have not yet found a technically sound and conceptually clear way of forcing our argument-based beliefs to be introspective.

The second pending challenge consists in working on the achievement of a meaningful integration of our two main tracks of research. A fairly appealing option is the development and study of a multi-agent version of our Paper V and Paper VI's models satisfying, in turn, some of the awareness constraints discussed in Paper II. It is open to discussion, however, how to define the set of accepted defeasible rules in a multi-agent environment. A reasonable, minimal assumption could be that agents have perfect knowledge about which defeasible rules they accept, but they are not sure about which rules are accepted by others. In more restricted contexts, one can also suppose that defeasible rules belong to the common ground of the debate. Something similar happens with the definition of the naming function for defeasible rules ($\mathfrak{n}$, in our notation) that formally enables the presence of undercutting attacks. It seems plausible to assume that $\mathfrak{n}$ does not depend on agents nor on possible worlds, that is, that all agents share a naming convention for defeasible rules, and that this is common knowledge. Although promising and interesting, it is important to point out that this kind of setting does not only inherit the advantages and appealing features of both of its components, but also their shortcomings and problems. Just to mention an example, regarding the structured part, we have still not solved the question of how to capture argumentative semantics in an object language with structured arguments and/or to axiomatise the argument-based belief operators from Paper V and Paper VI. A further interesting open question would be if these models could throw some light about the intuitions underlying the kind of uncertainty modelled by incomplete argumentation frameworks and related mathematical structures. More concretely, each of the doxastically accessible worlds of these models would be associated with a structured AF. This set of AFs resembles, in turn, an incomplete AF. The question is then whether the assumptions about these attributions underlying incomplete AFs (those spotted in Paper III) are natural enough or should be relaxed in order to get more general structures (see also (Herzig and Yuste-Ginel, 2021a; Mailly, 2021b)).

Third, as mentioned here and there, our contributions completely lack an analysis of computational issues. Although such an analysis was out of the scope of this thesis' project from the very beginning, it represents an urgent path to traverse, since research in artificial intelligence naturally requires it. In this direction, many of the argumenta-

tive and epistemic tools that we have used here have already been well studied from a computational point of view,[3] so that the results of both could be taken as a departure point.

Finally, there are also conceptual challenges for the future. For instance, it is obvious that $\mathsf{C1}_{\mathsf{epistemic}}$ and $\mathsf{C1}_{\mathsf{rhetoric}}$ do not constitute an exhaustive classification of the ways in which epistemic attitudes, and more concretely beliefs, may have an influence on argumentative processes. As an example, and along the lines of the work carried out by Mercier and Sperber (2017), decades of empirical findings talk in favour of the existence of an argumentative version of the so-called *myside* or *confirmation bias*: *arguing for statements that I believe to be true is easier than arguing for those I don't believe*.[4] However, we have decided to leave out this and other equally intuitive relations among argumentation and beliefs because it seems that an appropriate analysis of them requires the modelling of some motivational/emotional/intentional aspects in order to be relevant. Developing such an integration, e.g., by importing tools from the formal analysis of emotions and intentions, represents a promising future line of research.

[3] See e.g., (Aucher and Schwarzentruber, 2013) for DEL and (Dvorák and Dunne, 2018) for FA.

[4] I should thank my dear friend Filippo Donvito for calling my attention to this issue during a nice informal discussion.

Appendix

We devote the only appendix of this dissertation to depict the proofs omitted in Paper III and Paper IV for space reasons, as well as to correct typos and minor mistakes spotted in the contributions reprinted in Chapter 4.

Proofs of Paper III

Proposition 1. *Let* $\mathsf{IAF} = (\mathsf{A}, \mathsf{A}^?, \mathsf{R}, \mathsf{R}^?)$ *be an IAF such that* $\mathsf{A} \cup \mathsf{A}^? \subseteq \Sigma$*, let* v_{IAF} *be its propositional valuation, and let* $M_{vis} = (2^{\mathsf{ATM}^\Sigma}, \sim)$ *be the single-agent visibility model for the set of variables* ATM^Σ*, then:*

- *For each completion* $(\mathsf{A}^*, \mathsf{R}^*)$ *of* IAF *there is a* $u \in \sim [v_{\mathsf{IAF}}]$ *such that* $(\mathsf{A}^*, \mathsf{R}^*) = (\mathsf{A}_u, \mathsf{R}_u)$.[5]
- *For each* $u \in \sim [v_{\mathsf{IAF}}]$ *there is a completion* $(\mathsf{A}^*, \mathsf{R}^*)$ *of* IAF *such that* $(\mathsf{A}^*, \mathsf{R}^*) = (\mathsf{A}_u, \mathsf{R}_u)$.

Proof. For the first item, suppose that $(\mathsf{A}^*, \mathsf{R}^*)$ is a completion of IAF, which by definition of completion is equivalent to

$$\mathsf{A} \subseteq \mathsf{A}^* \subseteq \mathsf{A} \cup \mathsf{A}^?; \text{ and} \tag{1}$$

$$\mathsf{R}_{|\mathsf{A}^*} \subseteq \mathsf{R}^* \subseteq (\mathsf{R} \cup \mathsf{R}^?)_{|\mathsf{A}^*}. \tag{2}$$

Now, let $u^* = v_{\mathsf{IAF}} \cup \{\mathsf{aw}_x \mid x \in \mathsf{A}^*\} \cup \{\mathsf{r}_{x,y} \mid (x,y) \in \mathsf{R}^*\}$. By construction, we have that $(\mathsf{A}^*, \mathsf{R}^*) = (\mathsf{A}_{u^*}, \mathsf{R}_{u^*})$,[6] so we just have to show that $v_{\mathsf{IAF}} \sim u^*$, which amounts to showing that both conditions of the definition of $\sim$ are satisfied. The condition that u^* contains the same visibility atoms as v_{IAF} is immediate by definition of u^*. Hence, we just have to show that the other condition ($\mathsf{S}p \in v_{\mathsf{IAF}}$ implies ($p \in v_{\mathsf{IAF}}$ iff $p \in u^*$)) holds for every $p \in \mathsf{Prop}^\Sigma$. Suppose that $\mathsf{S}p \in v_{\mathsf{IAF}}$. By definition of v_{IAF}, the latter is equivalent to one of the following four cases, for which we show that $p \in v_{\mathsf{IAF}}$ iff $p \in u^*$.

[5] Where $\sim [v_{\mathsf{IAF}}]$ denotes the $\sim$-equivalence class of v_{IAF}.

[6] The definition of v_{IAF} together with 1 guarantee that we have not added any aw_x such that $x \notin \mathsf{A}^*$. The same happens for attacks.

Case A: $p = \mathsf{aw}_x$ and $x \in \mathsf{A}$. The latter implies $\mathsf{aw}_x \in v_{\mathsf{IAF}}$ (by definition of v_{IAF}), and this, in turn, that $\mathsf{aw}_x \in u^*$ (by definition of u^*), and we are done.

Case B: $p = \mathsf{aw}_x$ and $x \in \Sigma \setminus (\mathsf{A} \cup \mathsf{A}^?)$. The latter implies $\mathsf{aw}_x \notin v_{\mathsf{IAF}}$ (by definition of v_{IAF}). For the sake of contradiction, suppose $\mathsf{aw}_x \in u^*$, which implies $x \in \mathsf{A}^*$ (by definition of u^*), which in turn implies $x \in \mathsf{A} \cup \mathsf{A}^?$ (by (1)), which contradicts our hypothesis.

Case C: $p = \mathsf{r}_{x,y}$ and $(x, y) \in \mathsf{R}$. The latter implies $\mathsf{r}_{x,y} \in v_{\mathsf{IAF}}$ (by definition of v_{IAF}), which in turn implies $\mathsf{r}_{x,y} \in u^*$ (by definition of u^*), and we are done.

Case D: $p = \mathsf{r}_{x,y}$ and $(x, y) \in (\Sigma \times \Sigma) \setminus (\mathsf{R} \cup \mathsf{R}^?)$. The latter implies $\mathsf{r}_{x,y} \notin v_{\mathsf{IAF}}$ (by definition of v_{IAF}). Suppose, for the sake of contradiction, that $\mathsf{r}_{x,y} \in u^*$, which implies that $(x, y) \in \mathsf{R}^*$ (by definition of u^*), which implies that $(x, y) \in \mathsf{R} \cup \mathsf{R}^?$ (by (2)), which contradicts our hypothesis.

As for the second item, suppose that $v_{\mathsf{IAF}} \sim u$. We have to show that $(\mathsf{A}_u, \mathsf{R}_u)$ is a completion or, equivalently, that

$$\mathsf{A} \subseteq \mathsf{A}_u \subseteq \mathsf{A} \cup \mathsf{A}^?; \text{ and} \tag{3}$$

$$\mathsf{R}_{|\mathsf{A}_u} \subseteq \mathsf{R}_u \subseteq (\mathsf{R} \cup \mathsf{R}^?)_{|\mathsf{A}_u}. \tag{4}$$

For (3), suppose that $x \in \mathsf{A}$. The latter implies that $\mathsf{Saw}_x, \mathsf{aw}_x \in v_{\mathsf{IAF}}$ (be definition of v_{IAF}), which implies $\mathsf{aw}_x \in u$ (because $v_{\mathsf{IAF}} \sim v$), which implies $x \in \mathsf{A}_u$ (by definition of A_u), and we are done with the first inclusion. Now, suppose that $x \in \mathsf{A}_u$, which implies $\mathsf{aw}_x \in u$ (by definition of A_u). We continue by cases on $\mathsf{Saw}_x \in v_{\mathsf{IAF}}$. Suppose that $\mathsf{Saw}_x \in v_{\mathsf{IAF}}$, which yields (using the previous information) to $x \in \mathsf{A}$, and hence $x \in \mathsf{A} \cup \mathsf{A}^?$. Suppose that $\mathsf{Saw}_x \notin v_{\mathsf{IAF}}$, which yields $x \in \mathsf{A}^?$ (by definition of v_{IAF}), and hence $x \in \mathsf{A} \cup \mathsf{A}^?$, and we are done with the second inclusion.

As for (4), suppose that $(x, y) \in \mathsf{R}_{|\mathsf{A}_u}$. The latter implies $\mathsf{Sr}_{x,y}, \mathsf{r}_{x,y} \in v_{\mathsf{IAF}}$ (by definition of v_{IAF}), which in turn implies $\mathsf{r}_{x,y} \in u$ (because $v_{\mathsf{IAF}} \sim u$). Moreover, from $(x, y) \in \mathsf{R}_{|\mathsf{A}_u}$ we can also deduce that $x, y \in \mathsf{A}_u$. Both facts ($\mathsf{r}_{x,y} \in u$ and $x, y \in \mathsf{A}_u$) imply that $(x, y) \in \mathsf{R}_u$ (by definition of R_u), so we are done with the first inclusion. For the other one, suppose that $(x, y) \in \mathsf{R}_u$, which implies $x, y \in \mathsf{A}_u$ and $\mathsf{r}_{x,y} \in u$ (by definition of R_u). We continue by cases on $\mathsf{Sr}_{x,y} \in v_{\mathsf{IAF}}$. Suppose that $\mathsf{Sr}_{x,y} \in v_{\mathsf{IAF}}$, that yields $(x, y) \in \mathsf{R}$ (because $v_{\mathsf{IAF}} \sim u$ and $\mathsf{r}_{x,y} \in u$), and hence $(x, y) \in \mathsf{R} \cup \mathsf{R}^?$. For the other case, suppose that $\mathsf{Sr}_{x,y} \notin v_{\mathsf{IAF}}$, which implies $(x, y) \in \mathsf{R}^?$ (by definition of v_{IAF}), and hence $(x, y) \in \mathsf{R} \cup \mathsf{R}^?$, so we are done with the second inclusion. □

Proposition 2. *Let Σ be a signature, let $M_{vis} = (2^{\mathsf{ATM}^\Sigma}, \sim)$ be the single-agent visibility model for ATM^Σ. Then, we have that for every $\mathsf{IAF} = (\mathsf{A}, \mathsf{A}^?, \mathsf{R}, \mathsf{R}^?)$ built over Σ there is a $u \in 2^{\mathsf{ATM}^\Sigma}$ such that $u = v_{\mathsf{IAF}}$.*

Proof. It follows immediately from the fact that 2^{ATM^Σ} contains all possible valuations over Σ.

□

Proposition 3. *Let* $\mathsf{IAF} = (\mathsf{A}, \mathsf{A}^?, \mathsf{R}, \mathsf{R}^?)$ *be an IAF built over* Σ*, let* v_{IAF} *be its propositional valuation, let* $a \in A$ *and let* $M_{vis} = (2^{\mathsf{ATM}^\Sigma}, \sim)$ *be the single-agent visibility model for the set of variables* ATM^Σ*, then:*

- *The answer to the* st*-NSA (stable-necessary-sceptical-acceptance) problem with input* IAF *and* $a \in A$ *is yes iff* $M_{vis}, v_{\mathsf{IAF}} \models \mathbf{K}(\mathsf{Stable} \rightarrow \mathsf{in}_a)$.
- *The answer to the* st*-PCA (stable-possible-credulous-acceptance) problem with input* IAF *and* $a \in A$ *is yes iff* $M_{vis}, v_{\mathsf{IAF}} \models \hat{\mathbf{K}}(\mathsf{Stable} \wedge \mathsf{in}_a)$.

Proof. For the first item, we show that the needed chain of equivalences holds. We use $\Longleftrightarrow$ as an abbreviation of "the latter is true if and only if".

Suppose that the answer to the $\mathsf{st}{-}NSA$ (stable-necessary-sceptical-acceptance) problem with input IAF and $a \in A$ is yes.

$\Longleftrightarrow$

for every completion $(\mathsf{A}^*, \mathsf{R}^*)$ of IAF, every stable extension E of $(\mathsf{A}^*, \mathsf{R}^*)$ we have that $a \in E$ (by definition of $\mathsf{st}{-}NSA$)

$\Longleftrightarrow$

for every $u \in{\sim}\, [v_{\mathsf{IAF}}]$, every stable extension E of $(\mathsf{A}_u, \mathsf{R}_u)$ we have that $a \in E$ (by Proposition 1)

$\Longleftrightarrow$

for every $u \in{\sim}\, [v_{\mathsf{IAF}}]$, $u \models \mathsf{Stable} \rightarrow \mathsf{in}_a$ (from left to right, by (Doutre et al., 2017, Proposition 1). From right to left, by the same proposition plus the fact that for any valuation v' that only differs from v_{IAF} in the value assigned to in_x-variables, we have that $v' \in{\sim}\, [v_{\mathsf{IAF}}]$ (note that, by definition of v_{IAF}, the agent does not see the value of any in_x-variable at v_{IAF}))

$\Longleftrightarrow$

for every $v_{\mathsf{IAF}} \models \mathbf{K}(\mathsf{Stable} \rightarrow \mathsf{in}_a)$ (truth clause for $\mathbf{K}$).

The proof of the second item runs along analogous lines. □

Proposition 4. *Let* $\mathsf{IAF} = (\mathsf{A}, \mathsf{A}^?, \mathsf{R}, \mathsf{R}^?)$ *be an IAF built over* Σ*, let* v_{IAF} *be its propositional valuation, let* M_{vis} *be the visibility model for* $\mathsf{ATM}^\Sigma_{\{1,2\}}$*, and let* $a \in A$*, then:*

- *The answer to the* $\mathsf{st}{-}NCA$ *(stable-necessary-credulous-acceptance) problem with input* IAF *and* $a \in A$ *is yes iff* $M_{vis}, v_{\mathsf{IAF}} \models \mathbf{K}_1 \Diamond_2 (\mathsf{Stable} \wedge \mathsf{in}_a)$.
- *The answer to the* st*-PCA (stable-possible-sceptical-acceptance) problem with input* IAF *and* $a \in A$ *is yes iff* $M_{vis}, v_{\mathsf{IAF}} \models \hat{\mathbf{K}}_1 \Box_2 (\mathsf{Stable} \rightarrow \mathsf{in}_a)$.

Proof (Sketched). We just formulate incremental lemmas from which the proposition follows easily. First of all, we have to adapt Proposition 1 to the two-agent setting, showing that the epistemic part of the formalised agent (the one captured by the relation $\sim_1$) still describes properly the different completions of IAF.

Lemma 1. *Let* $\mathsf{IAF} = (\mathsf{A}, \mathsf{A}^?, \mathsf{R}, \mathsf{R}^?)$ *be an IAF, let* v_{IAF} *be its associated propositional valuation in the visibility model for* $\mathsf{ATM}^{\Sigma}_{\{1,2\}}$ *(with* $\mathsf{A} \cup \mathsf{A}^? \subseteq \Sigma$*), then:*

- *For each completion* $(\mathsf{A}^*, \mathsf{R}^*)$ *of* IAF*, there is a* $u \in{\sim_1}\, [v_{\mathsf{IAF}}]$ *such that* $(\mathsf{A}^*, \mathsf{R}^*) = (\mathsf{A}_u, \mathsf{R}_u)$.[7]
- *For each* $u \in{\sim_1}\, [v_{\mathsf{IAF}}]$*, there is a completion* $(\mathsf{A}^*, \mathsf{R}^*)$ *of* IAF *such that* $(\mathsf{A}^*, \mathsf{R}^*) = (\mathsf{A}_u, \mathsf{R}_u)$.

Then, we can show that:

Lemma 2. *Let* $\mathsf{IAF} = (\mathsf{A}, \mathsf{A}^?, \mathsf{R}, \mathsf{R}^?)$*,* Σ*,* $\mathsf{ATM}^{\Sigma}_{\{1,2\}}$*, and* v_{IAF} *be as in the previous lemma. Let* $u \in{\sim_1}\, [v_{\mathsf{IAF}}]$*, then:*

- *For every* $u' \in{\sim_2}\, [u]$*,* $(\mathsf{A}_u, \mathsf{R}_u) = (\mathsf{A}_{u'}, \mathsf{R}_{u'})$.
- 1. *For every* $E \subseteq \mathsf{A}_u$*, there is a* $v \in{\sim_2}\, [u]$*:* $E = E_v$.
  2. *For every* $v \in{\sim_2}\, [u]$*, there is an* $E \subseteq \mathsf{A}_u$*:* $E = E_v$.
- 1. *For every* $E \in \mathsf{st}(\mathsf{A}_u, \mathsf{R}_u), a \in E$ *iff* $u \models \Box_2(\mathsf{Stable} \to \mathsf{in}_a)$.
  2. *There is an* $E \in \mathsf{st}(\mathsf{A}_u, \mathsf{R}_u) : a \in E$ *iff* $u \models \Diamond_2(\mathsf{Stable} \wedge \mathsf{in}_a)$.

Finally, from the previous lemma, Proposition 4 follows easily.

□

Proposition 5. *Let* $M = (W, \mathsf{R}, V)$ *be a Kripke model for* Prop^{Σ}*. Then, we have that* M *is an IAF-friendly model iff* R_i *is serial for every* $i \in \mathtt{Agt}$*, and all instances of (awar) and (comp) are valid in* M.

Proof. We first recall the two conditions of the definition of IAF-friendly Kripke models:

(AWAR) if $\mathsf{R}_i[w] \cap V(\mathsf{r}_{x,y}) \neq \varnothing$, then $\mathsf{R}_i[w] \cap V(\mathsf{aw}_x) \neq \varnothing$ and $\mathsf{R}_i[w] \cap V(\mathsf{aw}_y) \neq \varnothing$; and

(COMP) there is an incomplete AF, IAF, built over Σ, such that: $\mathsf{completions}(\mathsf{IAF}) = \{(\mathsf{A}_v, \mathsf{R}_v) \mid v \in \mathsf{R}_i[w]\}$;

as well as the special corresponding axioms:

(awar) $\hat{\mathbf{K}}_i \mathsf{r}_{x,y} \to (\hat{\mathbf{K}}_i \mathsf{aw}_x \wedge \hat{\mathbf{K}}_i \mathsf{aw}_y)$; and

(comp) $(\hat{\mathbf{K}}_i l_1 \wedge \cdots \wedge \hat{\mathbf{K}}_i l_n) \to \hat{\mathbf{K}}_i (l_1 \wedge \cdots \wedge l_n)$.

[7] Where $\sim_i\, [v_{\mathsf{IAF}}]$ denotes the equivalence class of v_{IAF} w.r.t to $\sim_i$.

From left to right. Suppose that M is an IAF-friendly model, that is, for every $w \in W$, every $x, y \in \Sigma$, and every $i \in \mathtt{Agt}$, both (AWAR) and (COMP) holds.

To see that each R_i is serial follows from (COMP), note that every IAF has a non-empty set of completions by definition (even in the extreme case where $\mathsf{A} = \mathsf{A}^? = \mathsf{R} = \mathsf{R}^? = \varnothing$, then $(\varnothing, \varnothing)$ is the only completion), therefore $\{(\mathsf{A}_v, \mathsf{R}_v) \mid v \in \mathsf{R}_i[w]\}$ is not empty and hence $\mathsf{R}_i[w]$ is not empty either.

The model-validity of all instances of (awar) follows almost immediately from (AWAR).

As for the model-validity of all instances of (comp), take a world $u \in W$ and a consistent list of literals $l_1, ..., l_n$ from Prop^{Σ}, and suppose that $M, u \models \hat{\mathbf{K}}_i l_1 \wedge, ..., \wedge \hat{\mathbf{K}}_i l_n$. Recall that, since (COMP) holds, there is an IAF, let us call it IAF^u, such that $\mathsf{completions}(\mathsf{IAF}^u) = \{(\mathsf{A}_v, \mathsf{R}_v) \mid v \in \mathsf{R}_i[u]\}$. It is then easy to check, using the hypothesis that (AWAR) holds,that for each l in the list $l_1, ..., l_n$, if $l = \mathsf{aw}_x$, then x belongs to the set of arguments of a completion of IAF^u, and if $l = \mathsf{r}_{x,y}$, then (x, y) belongs to the set of attacks of a completion of IAF^u.[8] From the definition of completion and the previous assertion, it follows that there is a completion $(\mathsf{A}^*, \mathsf{R}^*)$ of IAF^u such that for every literal l in the list $l_1, ..., l_n$, if $l = \mathsf{aw}_x$, then $x \in \mathsf{A}^*$; and if $l = \mathsf{r}_{x,y}$, then $(x, y) \in \mathsf{R}^*$. This implies, together with the equality between completions of IAF^u and the set of AFs associated to i-successors of u stated above, that $M, u \models \hat{\mathbf{K}}_i(l_1 \wedge, ..., \wedge l_n)$.

From right to left. Suppose that R_i is serial for every $i \in \mathtt{Agt}$, and that all instances of (awar) and (comp) are valid in M. We need to show that (AWAR) and (COMP) holds for every $i \in \mathtt{Agt}$, $x, y \in \Sigma$, $w \in W$. The former is almost immediate from the model-validity of all instances of (awar). Therefore, the remaining of the proof consists in showing that for every $w \in W$ and every $i \in \mathtt{Agt}$ there is an IAF, let us call it IAF_i^w, such that $\mathsf{completions}(\mathsf{IAF}_i^w) = \{(\mathsf{A}_u, \mathsf{R}_u) \mid u \in \mathsf{R}_i[w]\}$. We show this by construction. Let us define $\mathsf{IAF} = (\mathsf{A}_i^w, \mathsf{A}_i^{?w}, \mathsf{R}_i^w, \mathsf{R}_i^{?w})$, where

- $\mathsf{A}_i^w = \{x \in \Sigma \mid M, w \models \mathbf{K}_i \mathsf{aw}_x\}$,
- $\mathsf{A}_i^{?w} = \{x \in \Sigma \mid M, w \models \hat{\mathbf{K}}_i \mathsf{aw}_x \wedge \hat{\mathbf{K}}_i \neg \mathsf{aw}_x\}$.
- $\mathsf{R}_i^w = \{(x, y) \in \Sigma \times \Sigma \mid M, w \models \hat{\mathbf{K}}_i \mathsf{aw}_x \wedge \hat{\mathbf{K}}_i \mathsf{aw}_y \wedge \mathbf{K}_i((\mathsf{aw}_x \wedge \mathsf{aw}_y) \rightarrow \mathsf{r}_{x,y})\}$.
- $\mathsf{R}_i^{?w} = \{(x, y) \in \Sigma \times \Sigma \mid M, w \models \hat{\mathbf{K}}_i(\mathsf{aw}_x \wedge \mathsf{aw}_y \wedge \mathsf{r}_{x,y}) \wedge \hat{\mathbf{K}}_i(\mathsf{aw}_x \wedge \mathsf{aw}_y \wedge \neg \mathsf{r}_{x,y})\}$.

Now we show that IAF_i^w is indeed an IAF, and that both inclusions of $\mathsf{completions}(\mathsf{IAF}_i^w) = \{(\mathsf{A}_u, \mathsf{R}_u) \mid u \in \mathsf{R}_i[w]\}$ hold.

To see that IAF_i^w is an IAF, note that all conditions from the definition of IAF, that are, $\mathsf{A}_i^w \cap \mathsf{A}_i^{?w} = \varnothing$, $\mathsf{R}_i^w \cap \mathsf{R}_i^{?w} = \varnothing$, and $\mathsf{R}_i^w, \mathsf{R}_i^{?w} \subseteq (\mathsf{A}_i^w \cup \mathsf{A}^{?w}) \times (\mathsf{A}_i^w \cup \mathsf{A}^{?w})$, follow from the definition of IAF_i^w and simple modal reasoning.

[8] For the second assertion, note that whenever $l = \mathsf{r}_{x,y}$ is in the list, then $M, w \models \hat{\mathbf{K}}_i \mathsf{aw}_x \wedge \hat{\mathbf{K}}_i \mathsf{aw}_y$ follows from (AWAR).

Regarding the right-to-left inclusion, i.e., $\{(\mathsf{A}_u, \mathsf{R}_u) \mid u \in \mathsf{R}_i[w]\} \subseteq \mathsf{completions}(\mathsf{IAF}_i^w)$, take $v \in \mathsf{R}_i[w]$. We need to show that both chains of inclusions of the definition of completion hold. For the first chain, suppose that $x \in \mathsf{A}_i^w$, which implies $M, w \models \mathbf{K}_i \mathsf{aw}_x$ (by definition of A_i^w), which implies $M, v \models \mathsf{aw}_x$ (because $v \in \mathsf{R}_i[w]$), which implies $x \in \mathsf{A}_v$ (by definition of A_v), which completes the first step of the chain. For the second step, suppose that $x \in \mathsf{A}_v$, which implies $M, v \models \mathsf{aw}_x$ (by definition of A_v), which implies $M, w \models \hat{\mathbf{K}}_i \mathsf{aw}_x$ (because $v \in \mathsf{R}_i[w]$), which implies $x \in \mathsf{A}_i^w \cup \mathsf{A}_i^{?w}$ (by definition of IAF_i^w and modal reasoning). As for the second chain, suppose that $(x, y) \in \mathsf{R}_i^w \cap (\mathsf{A}_v \times \mathsf{A}_v)$. This implies $(x, y) \in \mathsf{R}_i^w$ and $x, y \in \mathsf{A}_v$, which implies $M, v \models \mathsf{aw}_x \wedge \mathsf{aw}_y$ and $M, w \models \mathbf{K}_i(\mathsf{aw}_x \wedge \mathsf{aw}_y \rightarrow \mathsf{r}_{x,y})$ (by definition of A_v and R_i^w, respectively), which implies $M, v \models \mathsf{aw}_x \wedge \mathsf{aw}_y$ and $M, v \models \mathsf{aw}_x \wedge \mathsf{aw}_y \rightarrow \mathsf{r}_{x,y}$ (because $v \in \mathsf{R}_i[w]$), which implies $M, v \models \mathsf{aw}_x \wedge \mathsf{aw}_y$ and $M, v \models \mathsf{r}_{x,y}$ (by propositional reasoning), which implies $(x, y) \in \mathsf{R}_v$ (by definition of $(\mathsf{A}_v, \mathsf{R}_v)$), and this concludes the first step of the chain. For the second step of the chain, suppose that $(x, y) \in \mathsf{R}_v$, which implies $M, v \models \mathsf{aw}_x \wedge \mathsf{aw}_y \wedge \mathsf{r}_{x,y}$ (by definition of R_v), which implies $M, w \models \hat{\mathbf{K}}_i(\mathsf{aw}_x \wedge \mathsf{aw}_y \wedge \mathsf{r}_{x,y})$ (because $v \in \mathsf{R}_i[w]$). We continue by cases on the truth value of the formula $\mathbf{K}_i \neg(\mathsf{aw}_x \wedge \mathsf{aw}_y \wedge \neg \mathsf{r}_{x,y})$ at (M, w). For the first case, suppose that $(M, w) \models \mathbf{K}_i \neg(\mathsf{aw}_x \wedge \mathsf{aw}_y \wedge \neg \mathsf{r}_{x,y})$, which implies $M, w \models \mathbf{K}_i((\mathsf{aw}_x \wedge \mathsf{aw}_y) \rightarrow \mathsf{r}_{x,y}))$ (by propositional reasoning), which implies $(x, y) \in \mathsf{R}_i^w$ (by definition of R_i^w). For the second case, suppose that $(M, w) \models \neg \mathbf{K}_i \neg(\mathsf{aw}_x \wedge \mathsf{aw}_y \wedge \neg \mathsf{r}_{x,y})$, which is equivalent to $(M, w) \models \hat{\mathbf{K}}_i(\mathsf{aw}_x \wedge \mathsf{aw}_y \wedge \neg \mathsf{r}_{x,y})$ (by definition of $\hat{\mathbf{K}}_i$), which together with $M, w \models \hat{\mathbf{K}}_i(\mathsf{aw}_x \wedge \mathsf{aw}_y \wedge \mathsf{r}_{x,y})$ (that we knew) implies $(x, y) \in \mathsf{R}_i^{?w}$ (by definition of $\mathsf{R}_i^{?w}$). From our analysis of cases we can conclude $(x, y) \in (\mathsf{R}_i^w \cup \mathsf{R}_i^{?w})$ which, together with $x, y \in \mathsf{A}_v$ (that we knew), implies $(x, y) \in (\mathsf{R}_i^w \cup \mathsf{R}_i^{?w}) \cap (\mathsf{A}_v \times \mathsf{A}_v)$, which concludes the second step of the chain, and the proof that $\{(\mathsf{A}_u, \mathsf{R}_u) \mid u \in \mathsf{R}_i[w]\} \subseteq \mathsf{completions}(\mathsf{IAF}_i^w)$.

Regarding the left-to-right inclusion, i.e., $\mathsf{completions}(\mathsf{IAF}_i^w) \subseteq \{(\mathsf{A}_u, \mathsf{R}_u) \mid u \in \mathsf{R}_i[w]\}$, take $(\mathsf{A}^*, \mathsf{R}^*) \in \mathsf{completions}(\mathsf{IAF}_i^w)$, which is equivalent by definition of completion to

$$\mathsf{A}_i^w \subseteq \mathsf{A}^* \subseteq (\mathsf{A}_i^w \cup \mathsf{A}_i^{?w}); \text{ and} \tag{5}$$

$$(\mathsf{R}_i^w)_{|\mathsf{A}^*} \subseteq \mathsf{R}^* \subseteq (\mathsf{R}_i^w \cup \mathsf{R}_i^{?w})_{|\mathsf{A}^*}. \tag{6}$$

Now, we will prove the following assertion

$$M, w \models \bigwedge_{x \in \mathsf{A}^*} \hat{\mathbf{K}}_i \mathsf{aw}_x \wedge \bigwedge_{x \in \Sigma \setminus \mathsf{A}^*} \hat{\mathbf{K}}_i \neg \mathsf{aw}_x \wedge \bigwedge_{(x,y) \in \mathsf{R}^*} \hat{\mathbf{K}}_i \mathsf{r}_{x,y} \wedge \bigwedge_{(x,y) \in \mathsf{A}^* \times \mathsf{A}^* \setminus \mathsf{R}^*} \hat{\mathbf{K}}_i \neg \mathsf{r}_{x,y}. \tag{7}$$

We proceed by showing that each of the four big conjuncts holds at (M, w). For the first one, suppose that $x \in \mathsf{A}^*$, which implies $x \in \mathsf{A}_i^{?w}$ or $x \in \mathsf{A}_i^{?w}$ (by (5)), which implies $M, w \models \hat{K}_i \mathsf{aw}_x$ (by definition of $\mathsf{A}_i^{?w}$, $\mathsf{A}_i^{?w}$, and modal reasoning). Generalizing over x and applying the semantics of $\wedge$ we obtain the first big conjunct. For the second one,

suppose that $x \notin \mathsf{A}^*$, which implies $x \notin \mathsf{A}_i^w$ (by (5)), which implies $M, w \models \hat{\mathbf{K}}_i \mathsf{aw}_x$ (by definition of A_i^w and modal reasoning). Generalizing over x and applying the semantics of $\wedge$ we obtain that the second big conjunct is true at (M, w). For the third one, suppose $(x, y) \in \mathsf{R}^*$, which yields $x, y \in \mathsf{A}^*$ and (either $(x, y) \in \mathsf{R}_i^w$ or $(x, y) \in \mathsf{R}_i^{?w}$), by (6). An analysis of cases shows that both $(x, y) \in \mathsf{R}_i^w$ and $(x, y) \in \mathsf{R}_i^{?w}$ leads to $M, w \models \hat{\mathbf{K}}_i \mathsf{r}_{x,y}$ (by definition of R_i^w, $\mathsf{R}_i^{?w}$ and modal reasoning). Generalizing over x and y and applying the semantics of $\wedge$, we are done with the third conjunct. For the fourth conjunct, suppose that $(x, y) \in (\mathsf{A}^* \times \mathsf{A}^*) \setminus \mathsf{R}^*$ that yields $x, y \in \mathsf{A}^*$ and $(x, y) \notin \mathsf{R}_i^w$ (by (6)). The previous assertion implies $x, y \in \mathsf{A}^*$ and (either $M, w \models \neg\hat{\mathbf{K}}_i \mathsf{aw}_x$ or $M, w \models \neg\hat{\mathbf{K}}_i \mathsf{aw}_y$ or $M, w \models \hat{\mathbf{K}}_i(\mathsf{aw}_x \wedge \mathsf{aw}_y \wedge \neg \mathsf{r}_{x,y})$), by the definition of R_i^w and modal reasoning. The previous assertion implies The previous assertion implies $M, w \models \hat{K}_i \mathsf{aw}_x$, $M, w \models \hat{K}_i \mathsf{aw}_y$ and (either $M, w \models \neg\hat{\mathbf{K}}_i \mathsf{aw}_x$ or $M, w \models \neg\hat{\mathbf{K}}_i \mathsf{aw}_y$ or $M, w \models \hat{\mathbf{K}}_i(\mathsf{aw}_x \wedge \mathsf{aw}_y \wedge \neg \mathsf{r}_{x,y})$), because of the truth of the first big conjunct (i.e., $M, w \models \bigwedge_{x \in \mathsf{A}^*} \hat{\mathbf{K}}_i \mathsf{aw}_x$). The previous assertion implies by propositional reasoning $M, w \models \hat{\mathbf{K}}_i(\mathsf{aw}_x \wedge \mathsf{aw}_y \wedge \neg \mathsf{r}_{x,y})$, which implies, by modal reasoning that $M, w \models \hat{K}_i \neg \mathsf{r}_{x,y}$. Finally, generalizing over x, y and applying the semantics of $\wedge$, we obtain that the four big conjunct is true at (M, w).

From (7) and the hypothesis that all instances of (comp) are valid, we obtain the following[9]

$$M, w \models \hat{\mathbf{K}}_i\Big(\bigwedge_{x \in \mathsf{A}^*} \mathsf{aw}_x \wedge \bigwedge_{x \in \Sigma \setminus \mathsf{A}^*} \neg \mathsf{aw}_x \wedge \bigwedge_{(x,y) \in \mathsf{R}^*} \mathsf{r}_{x,y} \wedge \bigwedge_{(x,y) \in \mathsf{A}^* \times \mathsf{A}^* \setminus \mathsf{R}^*} \neg \mathsf{r}_{x,y} \Big).$$

which implies, by the semantics of $\hat{K}_i$, that there is a $u \in \mathsf{R}_i[w]$ such that

$$M, u \models \Big(\bigwedge_{x \in \mathsf{A}^*} \mathsf{aw}_x \wedge \bigwedge_{x \in \Sigma \setminus \mathsf{A}^*} \neg \mathsf{aw}_x \wedge \bigwedge_{(x,y) \in \mathsf{R}^*} \mathsf{r}_{x,y} \wedge \bigwedge_{(x,y) \in \mathsf{A}^* \times \mathsf{A}^* \setminus \mathsf{R}^*} \neg \mathsf{r}_{x,y} \Big).$$

which implies that there is a $u \in \mathsf{R}_i[w]$ such that $(\mathsf{A}^*, \mathsf{R}^*) = (\mathsf{A}_u, \mathsf{R}_u)$ (by propositional reasoning and the definition of $(\mathsf{A}_u, \mathsf{R}_u)$. This concludes the proof. □

Proofs of Paper IV

Proposition 1. *For every multi-agent AF* $\mathcal{M} = \langle A, R, \{R^i\}_{i \in \mathtt{Agt}}, \{(R^i)^j\}_{i,j \in \mathtt{Agt}}, R^{pub} \rangle$ *we have:*

- $(a, b) \in R$ *iff* $\theta(\mathcal{M}) \to \mathsf{r}_{a,b}$ *is S5-valid;*

[9] Note that, the subscripts of the big conjunctions in (7) guarantee that there are not inconsistent literals in the formula.

- $(a,b) \notin R$ *iff* $\theta(\mathcal{M}) \rightarrow \neg \mathsf{r}_{a,b}$ *is S5-valid;*
- $(a,b) \in R^i$ *iff* $\theta(\mathcal{M}) \rightarrow \mathbf{K}i\mathsf{r}_{a,b}$ *is S5-valid;*
- $(a,b) \notin R^i$ *iff* $\theta(\mathcal{M}) \rightarrow \neg\mathbf{K}i\mathsf{r}_{a,b}$ *is S5-valid;*
- $(a,b) \in (R^i)^j$ *iff* $\theta(\mathcal{M}) \rightarrow \mathbf{K}j\mathbf{K}i\mathsf{r}_{a,b}$ *is S5-valid;*
- $(a,b) \notin (R^i)^j$ *iff* $\theta(\mathcal{M}) \rightarrow \neg\mathbf{K}j\mathbf{K}i\mathsf{r}_{a,b}$ *is S5-valid;*
- $(a,b) \in R^{pub}$ *iff* $\theta(\mathcal{M}) \rightarrow \mathbf{C}_{\mathtt{Agt}}\mathsf{r}_{a,b}$ *is S5-valid;*
- $(a,b) \notin R^{pub}$ *iff* $\theta(\mathcal{M}) \rightarrow \neg\mathbf{C}_{\mathtt{Agt}}\mathsf{r}_{a,b}$ *is S5-valid.*

Proof (sketched). The proof of every item is analogous to each other. Every statement follows from the fact that epistemic theories are consistent and complete w.r.t. attack variables (resp. individual first-order variables, individual second-order variables and public variables) by simple propositional reasoning. □

Proposition 2. *Let $\mathcal{M}$ be a multi-agent AF.*

- *Then E^{pub} is a public stable extension of $\mathcal{M}$ if and only if*

$$\theta(\mathcal{M}) \wedge \mathsf{Stable}^{pub} \wedge \left(\bigwedge_{a \in E^{pub}} \mathsf{in}_a \right) \wedge \left(\bigwedge_{a \in A \setminus E^{pub}} \neg \mathsf{in}_a \right)$$

  is satisfiable in S5.

- *If $R^{pub} = R$, then $\theta(\mathcal{M}) \rightarrow (\mathsf{Stable}^{pub} \leftrightarrow \mathsf{Stable})$ is S5-valid.*
- *$a \in A$ is publicly-sceptically accepted (resp. publicly-credulously accepted) iff $(\theta(\mathcal{M}) \wedge \mathsf{Stable}^{pub}) \rightarrow \mathsf{in}_a$ is S5-valid (resp. $\theta(\mathcal{M}) \wedge \mathsf{Stable}^{pub} \wedge \mathsf{in}_a$ is S5-satisifiable).*

Proof. **First item. From right to left:** Suppose

$$\theta(\mathcal{M}) \wedge \mathsf{Stable}^{pub} \wedge \left(\bigwedge_{a \in E^{pub}} \mathsf{in}_a \right) \wedge \left(\bigwedge_{a \in A \setminus E^{pub}} \neg \mathsf{in}_a \right)$$

is satisfiable in S5. Let

$$M, w \models \theta(\mathcal{M}) \wedge \mathsf{Stable}^{pub} \wedge \left(\bigwedge_{a \in E^{pub}} \mathsf{in}_a \right) \wedge \left(\bigwedge_{a \in A \setminus E^{pub}} \neg \mathsf{in}_a \right)$$

$$\Longrightarrow \quad \text{(semantics of } \wedge\text{)}^{10}$$

$$M, w \models \theta(\mathcal{M}) \text{ AND } M, w \models \mathsf{Stable}^{pub} \text{ AND}$$
$$M, w \models \left(\bigwedge_{a \in E^{pub}} \mathsf{in}_a\right) \wedge \left(\bigwedge_{a \in A \setminus E^{pub}} \neg \mathsf{in}_a\right)$$

$$\Longrightarrow \quad \text{(definition of } \theta(\mathcal{M}) \text{ and propositional reasoning)}$$

$$M, w \models \bigwedge_{(x,y) \in R^{pub}} \mathbf{C}_{\mathtt{Agt}} \mathsf{r}_{x,y} \wedge \bigwedge_{(x,y) \in (A \times A) \setminus R^{pub}} \neg \mathbf{C}_{\mathtt{Agt}} \mathsf{r}_{x,y}$$

AND

$$M, w \models \mathsf{Stable}^{pub}$$

AND

$$M, w \models \left(\bigwedge_{a \in E^{pub}} \mathsf{in}_a\right) \wedge \left(\bigwedge_{a \in A \setminus E^{pub}} \neg \mathsf{in}_a\right)$$

$\Longrightarrow$ (propositional reasoning, set notation, and definition of Stable^{pub})

$$R^{pub} = \{(x, y) \in A \times A \mid M, w \models \mathbf{C}_{\mathtt{Agt}} \mathsf{r}_{x,y}\}$$

AND

$$M, w \models \bigwedge_{a \in A} \left(\mathsf{in}_a \leftrightarrow \bigwedge_{b \in A} (\mathbf{C}_{\mathtt{Agt}} \mathsf{r}_{b,a} \rightarrow \neg \mathsf{in}_b)\right)$$

AND

$$E^{pub} = \{x \in A \mid M, w \models \mathsf{in}_x\}$$

$\Longrightarrow$ (propositional reasoning and set notation)

$$R^{pub} = \{(x, y) \in A \times A \mid M, w \models \mathbf{C}_{\mathtt{Agt}} \mathsf{r}_{x,y}\}$$

AND

$$\forall a \Big(a \in \{x \in A \mid M, w \models \mathsf{in}_x\} \quad \text{iff} \quad \forall b \in A((b, a) \in \{(y, x) \in A \times A \mid M, w \models$$
$$\mathbf{C}_{\mathtt{Agt}} \mathsf{r}_{y,x}\} \quad \text{implies} \quad b \notin \{x \in A \mid M, w \models \mathsf{in}_x\})\Big)$$

AND

$$E^{pub} = \{x \in A \mid M, w \models \mathsf{in}_x\}$$

[10] Throughout this appendix the symbol $\Longrightarrow$ (resp. $\Longleftrightarrow$) abbreviates "the previous statement implies that" (resp. the previous statement holds if and only if) and the justification of this inference step is written between parentheses right after it (unless it is obvious).

$$\Longrightarrow \text{(replacement of identicals)}$$

$$\forall a\Big(a \in E^{pub} \quad \text{iff} \quad \forall b \in A((b,a) \in R^{pub} \quad \text{implies} \quad b \notin E^{pub})\Big)$$

$$\Longrightarrow \text{(definition of stable extension)}$$

$$E^{pub} \text{ is a stable extension of } \langle A, R^{pub}\rangle$$

First item. From left to right (sketched):

Let E^{pub} be a stable extension of $\langle A, R^{pub}\rangle$. Since $\mid \mathtt{Agt} \mid \geq 2$ by assumption, let $i_0, i_1 \in \mathtt{Agt}$. Now, we build a model where the target formula is true. Let $M^{\mathcal{M}} = \langle W, \{\sim_i\}_{i\in\mathtt{Agt}}, V\rangle$ where each of the components is defined as follows:

- $W = \{w_0\} \cup \{w_i \mid i \in \mathtt{Agt}\} \cup \{(w_i)^j \mid i, j \in \mathtt{Agt}, i \neq j\} \cup \{w_{pub}\}$;
- $\sim_i = \begin{cases} \{\langle w, w\rangle \mid w \in W\} \cup \{\langle w_0, w_i\rangle, \langle w_i, w_0\rangle\} \cup \{\langle w_i, (w_j)^i\rangle, \langle (w_j)^i, w_i\rangle \mid j \in \mathtt{Agt}, i \neq j\} \\ \cup\{\langle (w_{i_1})^{i_0}, w_{pub}\rangle, \langle w_{pub}, (w_{i_1})^{i_0}\rangle\} \quad \text{if} \quad i = i_0; \\ \{\langle w, w\rangle \mid w \in W\} \cup \{\langle w_0, w_i\rangle, \langle w_i, w_0\rangle\} \cup \{\langle w_i, (w_j)^i\rangle, \langle (w_j)^i, w_i\rangle \mid j \in \mathtt{Agt}, i \neq j\} \\ \quad \text{otherwise.} \end{cases}$
- $V(w_0) = \{\mathsf{r}_{x,y} \in \mathtt{Prp}_A \mid (x,y) \in R\} \cup \{\mathsf{in}_x \mid x \in E^{pub}\}$;
- $V(w_i) = \{\mathsf{r}_{x,y} \in \mathtt{Prp}_A \mid (x,y) \in R^i\} \cup \{\mathsf{in}_x \mid x \in E^{pub}\}$;
- $V((w_i)^j) = \{\mathsf{r}_{x,y} \in \mathtt{Prp}_A \mid (x,y) \in (R^i)^j\} \cup \{\mathsf{in}_x \mid x \in E^{pub}\}$; and
- $V(w_{pub}) = \{\mathsf{r}_{x,y} \in \mathtt{Prp}_A \mid (x,y) \in R^{pub}\} \cup \{\mathsf{in}_x \mid x \in E^{pub}\}$.

As an illustration of the previous definition, the Kripke frame, that is the tuple $\langle W, \{\sim_i \}_{i\in\mathtt{Agt}}\rangle$, for the special case where $\mathtt{Agt} = \{1,2,3\}$, $i_0 = 1$, and $i_1 = 2$ is depicted in Figure 1.

We first need to show that $M^{\mathcal{M}}$ is actually an $S5$-model. Note that reflexivity and symmetry of each $\sim_i$ is a direct consequence of the definition of $\sim_i$. For transitivity, one has to show that, due to the definition of $M^{\mathcal{M}}$ again, we have that whenever $w \sim_i w' \sim_i w''$ then either $w = w'$ or $w' = w''$. The $w \sim_i w''$ follows easily. Details are left to the reader.

Then, all we need to show is that $\theta(\mathcal{M}) \wedge \mathsf{Stable}^{pub} \wedge \left(\bigwedge_{a\in E^{pub}} \mathsf{in}_a\right) \wedge \left(\bigwedge_{a\in A\setminus E^{pub}} \neg\mathsf{in}_a\right)$ is true at $M^{\mathcal{M}}, w_0$. We just note that $M^{\mathcal{M}}, w_0 \models \theta(\mathcal{M}) \wedge \left(\bigwedge_{a\in E^{pub}} \mathsf{in}_a\right) \wedge \left(\bigwedge_{a\in A\setminus E^{pub}} \neg\mathsf{in}_a\right)$ follows from the definition of the model $M^{\mathcal{M}}$ and the shorthand $\theta(\mathcal{M})$, while $M^{\mathcal{M}}, w_0 \models \mathsf{Stable}^{pub}$ follows from the previous statement and the hypothesis that E^{pub} is a stable extension of $\langle A, R^{pub}\rangle$. □

As a simple corollary of what we just have proved, it holds that

Corollary 1. $M, w \models \theta(\mathcal{M}) \wedge \mathsf{Stable}^{pub}$ *implies that* $\{x \in A \mid M, w \models \mathsf{in}_x\}$ *is a stable extension of* $\langle A, R^{pub}\rangle$.

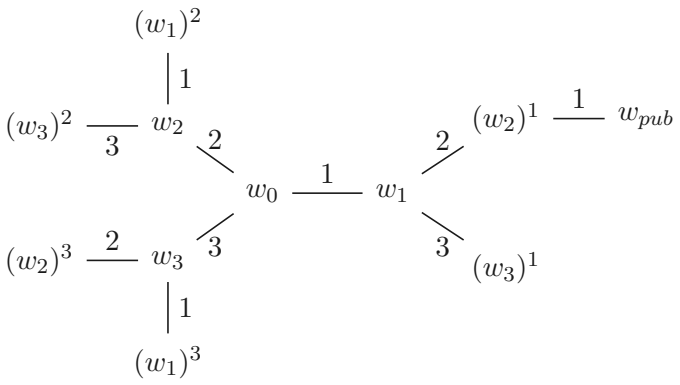


Figure 1: Kripke frame for the left-to-right direction of the first item of Proposition 2. We represent the case where $\mathtt{Agt} = \{1, 2, 3\}$, $i_0 = 1$, and $i_1 = 2$ (see the proof for clarification). Reflexive arrows as well as the direction of the remaining ones are omitted.

Second item. [Recall: If $R^{pub} = R$, then $\theta(\mathcal{M}) \to (\mathsf{Stable}^{pub} \leftrightarrow \mathsf{Stable})$ is S5-valid.]

Proof. Suppose that $R^{pub} = R$. We need to show that $\theta(\mathcal{M}) \to (\mathsf{Stable}^{pub} \leftrightarrow \mathsf{Stable})$ is S5-valid, which amounts to showing that $\theta(\mathcal{M}) \to (\mathsf{Stable}^{pub} \to \mathsf{Stable})$ and $\theta(\mathcal{M}) \to (\mathsf{Stable} \to \mathsf{Stable}^{pub})$ are both S5-valid. We just show the validity of the former formula (the other one is similar). Take an arbitrary pointed S5-model $\langle M, w\rangle$, and suppose that

$$M, w \models \theta(\mathcal{M}) \text{ AND } M, w \models \mathsf{Stable}^{pub}$$

$$\Longrightarrow \text{(Corollary 1)}$$

$$\{x \in A \mid M, w \models \mathsf{in}_x\} \text{ is a stable extension of } \langle A, R^{pub}\rangle$$

$$\Longleftrightarrow \quad \text{(substitution of identicals)}$$

$$\{x \in A \mid M, w \models \mathsf{in}_x\} \text{ is a stable extension of } \langle A, R\rangle$$

$$\Longleftrightarrow \quad \text{(semantics of atomic propositions)}$$

$$\{x \in A \mid \mathsf{in}_x \in V(w)\} \text{ is a stable extension of } \langle A, R\rangle \qquad (*)$$

From the hypothesis $M, w \models \theta(\mathcal{M})$ we can also deduce (by definition of $\theta(\mathcal{M})$ and propositional reasoning) that $M, w \models \theta(\langle A, R\rangle)$, which together with $(*)$ implies (by (**?**, Proposition 1)) that

$$V(w) \models \mathsf{Stable}^{11}$$

$$\iff \quad \text{(because Stable does not contain epistemic operators)}$$

$$M, w \models \mathsf{Stable}$$

By propositional reasoning we can conclude that $M, w \models \theta(\mathcal{M}) \rightarrow (\mathsf{Stable}^{pub} \leftrightarrow \mathsf{Stable})$. Since $\langle M, w\rangle$ was picked arbitrarily we can infer that $\theta(\mathcal{M}) \rightarrow (\mathsf{Stable}^{pub} \rightarrow \mathsf{Stable})$ is S5-valid. □

Third item. [Recall: $a \in A$ is publicly-sceptically accepted (resp. publicly-credulously accepted) iff $(\theta(\mathcal{M}) \wedge \mathsf{Stable}^{pub}) \rightarrow \mathsf{in}_a$ is S5-valid (resp. $\theta(\mathcal{M}) \wedge \mathsf{Stable}^{pub} \wedge \mathsf{in}_a$ is S5-satisfiable).]

Proof. **Public-sceptic acceptance. From left to right:** Suppose $a \in A$ is publicly-sceptically accepted, which means by definition that a belongs to every stable extension of $\langle A, R^{pub}\rangle$. Suppose, that $M, w \models \theta(\mathcal{M}) \wedge \mathsf{Stable}^{pub}$. The previous implies (by Corollary 1) that $\{x \in A \mid M, w \models \mathsf{in}_x\}$ is a stable extension of $\langle A, R^{pub}\rangle$. Both statements jointly implies $M, w \models \mathsf{in}_a$. The previous chain of reasoning, together with the meaning of $\rightarrow$, allows stating that $M, w \models (\theta(\mathcal{M}) \wedge \mathsf{Stable}^{pub}) \rightarrow \mathsf{in}_a$. Since $\langle M, w\rangle$ was arbitrarily taken, we can conclude that $(\theta(\mathcal{M}) \wedge \mathsf{Stable}^{pub}) \rightarrow \mathsf{in}_a$ is S5-valid.

From right to left: Suppose that $(\theta(\mathcal{M}) \wedge \mathsf{Stable}^{pub}) \rightarrow \mathsf{in}_a$ is S5-valid, and that E is a public stable extension of $\mathcal{M}$. We have, by the first item of this proposition, that $\theta(\mathcal{M}) \wedge \mathsf{Stable}^{pub} \wedge \left(\bigwedge_{x \in E} \mathsf{in}_x\right) \wedge \left(\bigwedge_{x \in A \setminus E} \neg \mathsf{in}_x\right)$ is satisfiable in S5. Let M, w be a pointed S5-model where the previous formula is true. We have that, in particular, $M, w \models \theta(\mathcal{M}) \wedge \mathsf{Stable}^{pub}$. From the previous statement and the validity of $(\theta(\mathcal{M}) \wedge \mathsf{Stable}^{pub}) \rightarrow \mathsf{in}_a$ (which is assumed by hypothesis), we have that $M, w \models \mathsf{in}_a$. It is then easy to deduce that $a \in E$ (recall that, $M, w \models \left(\bigwedge_{x \in E} \mathsf{in}_x\right) \wedge \left(\bigwedge_{x \in A \setminus E} \neg \mathsf{in}_x\right)$). Therefore, whenever E is a public stable extension of $\mathcal{M}$, $a \in E$, i.e., a is publicly-sceptically accepted.

Public-credulous acceptance. From left to right: Suppose $a \in A$ is publicly-credulously accepted, which means by definition that $a \in E$ for some public extension of $\mathcal{M}$. Let E be such an extension. By the first item of this proposition, we have that $\theta(\mathcal{M}) \wedge \mathsf{Stable}^{pub} \wedge \left(\bigwedge_{x \in E} \mathsf{in}_x\right) \wedge \left(\bigwedge_{x \in A \setminus E}\right)$ is S5-satisfiable. The latter implies (by the semantics of $\wedge$ and the fact that $a \in E$) that $\theta(\mathcal{M}) \wedge \mathsf{Stable}^{pub} \wedge \mathsf{in}_a$ is S5-satisfiable.

[11] Here '$\models$' denotes truth in propositional logic.

From right to left: Suppose $\theta(\mathcal{M}) \wedge \mathsf{Stable}^{pub} \wedge \mathsf{in}_a$ is S5-satisfiable. Let M, w be a pointed S5-model where the previous formula is true. By propositional reasoning, we have that $M, w \models \theta(\mathcal{M}) \wedge \mathsf{Stable}^{pub}$, which implies (by Corollary 1) that $\{x \in A \mid M, w \models \mathsf{in}_x\}$ is a public extension of $\mathcal{M}$. It is clear that $a \in \{x \in A \mid M, w \models \mathsf{in}_x\}$, therefore a belongs to at least one public extension of $\mathcal{M}$ which, by definition, means that a is publicly-credulously accepted. □

Proposition 3. *Let $\mathcal{M}$ be a multi-agent AF. Then $(a, b) \in R^i$ iff $\theta(\mathcal{M}) \rightarrow \langle \mathbf{K} i\mathsf{r}_{a,b}! \rangle \top$ is S5-PAL-valid.*

Proof. **From left to right:** Suppose $(a, b) \in R^i$. The latter implies, according to Proposition 1, that $\theta(\mathcal{M}) \rightarrow \mathbf{K} i\mathsf{r}_{a,b}$ is S5-valid, therefore it is S5-PAL-valid, too (because S5-PAL is an extension of S5). Note that the schema $\varphi \rightarrow \langle \varphi! \rangle \top$ is S5-PAL-valid (we omit the proof), so in particular $\mathbf{K} i\mathsf{r}_{a,b} \rightarrow \langle \mathbf{K} i\mathsf{r}_{a,b}! \rangle \top$ is S5-PAL-valid. From transitivity of $\rightarrow$ it follows that $\theta(\mathcal{M}) \rightarrow \langle \mathbf{K} i\mathsf{r}_{a,b}! \rangle \top$ is S5-PAL-valid.

From right to left: Suppose $\theta(\mathcal{M}) \rightarrow \langle \mathbf{K} i\mathsf{r}_{a,b}! \rangle \top$ is S5-PAL-valid. The latter implies that $\theta(\mathcal{M}) \rightarrow \langle \mathbf{K} i\mathsf{r}_{a,b}! \rangle \top$ and $\langle \mathbf{K} i\mathsf{r}_{a,b}! \rangle \top \rightarrow \mathbf{K} i\mathsf{r}_{a,b}$ are both S5-PAL-valid (because the schema $\langle \varphi! \rangle \top \rightarrow \varphi$ is S5-valid). The latter implies, by transitivity of $\rightarrow$, that $\theta(\mathcal{M}) \rightarrow \mathbf{K} i\mathsf{r}_{a,b}$ is S5-PAL-valid. Which in turn implies, since $\theta(\mathcal{M})$ does not contain dynamic operators by definition and S5-PAL is a conservative extension of S5, that $\theta(\mathcal{M}) \rightarrow \mathbf{K} i\mathsf{r}_{a,b}$ is also S5-valid. Which implies by Proposition 1 that $(a, b) \in R^i$. □

Proposition 4. *Let $\mathcal{M}$ be a multi-agent AF such that $(a, b) \in R^i$. Let*

$$\mathcal{M}\dot{+}i{:}(a,b) = \langle A, R, \{R^j \cup \{(a,b)\}\}_{j \in \mathtt{Agt}}, \{(R^j)^k \cup \{(a,b)\}\}_{j,k \in \mathtt{Agt}}, R^{pub} \cup \{(a,b)\} \rangle$$

be its update by $i{:}(a, b)$. For every $(c, d) \in A \times A$:

- *$(c, d) \in R^j \cup \{(a, b)\}$ iff $\theta(\mathcal{M}) \rightarrow [\mathbf{K} i\mathsf{r}_{a,b}!] \mathbf{K} j\mathsf{r}_{c,d}$ is S5-PAL-valid;*
- *$(c, d) \in (R^j)^k \cup \{(a, b)\}$ iff $\theta(\mathcal{M}) \rightarrow [\mathbf{K} i\mathsf{r}_{a,b}!] \mathbf{K} k \mathbf{K} j\mathsf{r}_{c,d}$ is S5-PAL-valid;*
- *$(c, d) \in R^{pub} \cup \{(a, b)\}$ iff $\theta(\mathcal{M}) \rightarrow [\mathbf{K} i\mathsf{r}_{a,b}!] \mathbf{C}_{\mathtt{Agt}} \mathsf{r}_{c,d}$ is S5-PAL-valid.*

Proof. **First item. From left to right:** Suppose $(c, d) \in R^j \cup \{(a, b)\}$ which is equivalent to $((c, d) \in R^j$ or $(c, d) = (a, b))$. Let us reason by cases. If $(c, d) \in R^j$, we have by Proposition 1 that $\theta(\mathcal{M}) \rightarrow \mathbf{K} j\mathsf{r}_{c,d}$ is S5-valid (and therefore also S5-PAL-valid). Note that $\mathbf{K} j\mathsf{r}_{c,d} \rightarrow [\mathbf{K} i\mathsf{r}_{a,b}!] \mathbf{K} j\mathsf{r}_{c,d}$ is S5-PAL-valid, too. (We omit the proof.) Hence by transitivity of $\rightarrow$ we have that $\theta(\mathcal{M}) \rightarrow [\mathbf{K} i\mathsf{r}_{a,b}!] \mathbf{K} j\mathsf{r}_{c,d}$ is S5-PAL-valid. On the other hand, if $(c, d) = (a, b)$ we have that $[\mathbf{K} i\mathsf{r}_{a,b}!] \mathbf{K} j\mathsf{r}_{c,d}$ amounts to $[\mathbf{K} i\mathsf{r}_{a,b}!] \mathbf{K} j\mathsf{r}_{a,b}$ which is S5-PAL-valid (we omit the proof) and, by propositional reasoning, we have that $\theta(\mathcal{M}) \rightarrow [\mathbf{K} i\mathsf{r}_{a,b}!] \mathbf{K} j\mathsf{r}_{c,d}$ is S5-valid, too.

From right to left: The case for $i = j$ follows easily using Proposition 1, axiom 4 and the S5-PAL-validity of the reduction axioms for public announcement **?**. We

show that the contrapositive holds for the case $i \neq j$. Suppose $(c,d) \notin R^j \cup \{(a,b)\}$, which clearly amounts to $(c,d) \notin R^j$ and $(c,d) \neq (a,b)$. We want to show that $\theta(\mathcal{M}) \rightarrow [\mathbf{K}i\mathsf{r}_{a,b}!]\mathbf{K}j\mathsf{r}_{c,d}$ is not S5-PAL-valid which amounts to showing that its negation is S5-PAL satisfiable, which in turn amounts to showing that there is a pointed S5-model such that $M, w \models \theta(\mathcal{M}) \wedge \neg[\mathbf{K}i\mathsf{r}_{a,b}!]\mathbf{K}j\mathsf{r}_{c,d}$. Using the validity of reduction axioms together with some modal and propositional reasoning we get the following chain of equivalences.
$\iff M, w \models \theta(\mathcal{M})$ AND $M, w \models \neg[\mathbf{K}i\mathsf{r}_{a,b}!]\mathbf{K}j\mathsf{r}_{c,d}$
$\iff M, w \models \theta(\mathcal{M})$ AND $M, w \models \neg(\mathbf{K}i\mathsf{r}_{a,b} \rightarrow \mathbf{K}j[\mathbf{K}i\mathsf{r}_{a,b}!]\mathsf{r}_{c,d})$
$\iff M, w \models \theta(\mathcal{M})$ AND $M, w \models \neg(\mathbf{K}i\mathsf{r}_{a,b} \rightarrow \mathbf{K}j(\mathbf{K}i\mathsf{r}_{a,b} \rightarrow \mathsf{r}_{c,d}))$
$\iff M, w \models \theta(\mathcal{M})$ AND $M, w \models \mathbf{K}i\mathsf{r}_{a,b}$ AND $M, w \models \neg\mathbf{K}j(\mathbf{K}i\mathsf{r}_{a,b} \rightarrow \mathsf{r}_{c,d})$
$\iff M, w \models \theta(\mathcal{M})$ AND $M, w \models \mathbf{K}i\mathsf{r}_{a,b}$ AND $M, w \models \hat{\mathbf{K}}_j\neg(\mathbf{K}i\mathsf{r}_{a,b} \rightarrow \mathsf{r}_{c,d})$
$\iff M, w \models \theta(\mathcal{M})$ AND $M, w \models \mathbf{K}i\mathsf{r}_{a,b}$ AND $M, w \models \hat{\mathbf{K}}_j(\mathbf{K}i\mathsf{r}_{a,b} \wedge \neg\mathsf{r}_{c,d})$

Note that, since $(a,b) \in R^i$, by Proposition 1 the formula $\theta(\mathcal{M}) \rightarrow \mathbf{K}i\mathsf{r}_{a,b}$ is S5-valid. Now, we build a pointed model in which $\theta(\mathcal{M}) \wedge \hat{\mathbf{K}}_j(\neg\mathsf{r}_{c,d} \wedge \mathbf{K}i\mathsf{r}_{a,b})$ is true. We just have to slightly modify the model used in the proof of the first item of Proposition 2. Let $M^{\mathcal{M}} = \langle W, \{\sim_k\}_{k\in\mathtt{Agt}}, V\rangle$ where each component is defined as follows (note that the references of agents' names i and j are now fixed):

- $W = \{w_0\} \cup \{w_k \mid k \in \mathtt{Agt}\} \cup \{(w_k)^m \mid k, m \in \mathtt{Agt}, k \neq m\} \cup \{w_{pub}\} \cup \{w'_j, (w'_i)^j\}$;
- $\sim_k = \begin{cases} \{\langle w,w\rangle \mid w \in W\} \cup \{\langle w_0,w_k\rangle, \langle w_k,w_0\rangle\} \cup \{\langle w_0,w'_j\rangle, \langle w'_j,w_0\rangle\} \\ \cup\{\langle w_k,w'_j\rangle, \langle w'_j,w_k\rangle\} \cup \{\langle w_m,(w_k)^m\rangle, \langle (w_k)^m,w_m\rangle \mid m \in \mathtt{Agt}, k \neq m\} \\ \cup\{\langle (w'_i)^j, w_{pub}\rangle, \langle w_{pub}, (w'_i)^j\rangle\} \quad \text{if} \quad k = j; \\ \\ \{\langle w,w\rangle \mid w \in W\} \cup \{\langle w_0,w_k\rangle, \langle w_k,w_0\rangle\} \\ \cup\{\langle w_m,(w_k)^m\rangle, \langle (w_k)^m,w_m\rangle \mid m \in \mathtt{Agt}, k \neq m\} \\ \cup\{\langle w'_j, (w'_i)^j\rangle, \langle (w'_i)^j, w'_j\rangle\} \quad \text{if} \quad k = i; \\ \\ \{\langle w,w\rangle \mid w \in W\} \cup \{\langle w_0,w_k\rangle, \langle w_k,w_0\rangle\} \\ \cup\{\langle w_m,(w_k)^m\rangle, \langle (w_k)^m,w_m\rangle \mid m \in \mathtt{Agt}, k \neq m\} \\ \quad \text{otherwise.} \end{cases}$
- $V(w_0) = \{\mathsf{r}_{x,y} \in \mathtt{Prp}_A \mid (x,y) \in R\}$;
- $V(w_k) = \{\mathsf{r}_{x,y} \in \mathtt{Prp}_A \mid (x,y) \in R^k\}$;
- $V(w'_j) = \{\mathsf{r}_{x,y} \in \mathtt{Prp}_A \mid (x,y) \in R^j\} \cup \{\mathsf{r}_{a,b}\}$;
- $V((w'_i)^j) = \{\mathsf{r}_{x,y} \in \mathtt{Prp}_A \mid (x,y) \in (R^i)^j\} \cup \{\mathsf{r}_{a,b}\}$;
- $V((w_k)^m) = \{\mathsf{r}_{x,y} \in \mathtt{Prp}_A \mid (x,y) \in (R^k)^m\}$; and

- $V(w_{pub}) = \{\mathsf{r}_{x,y} \in \mathtt{Prp}_A \mid (x,y) \in R^{pub}\}$.

As an example, the frame $\langle W, \{\sim_k\}_{k\in\mathtt{Agt}}\rangle$ for $\mathtt{Agt} = \{1,2\}$, $i = 1$ and $j = 2$ is depicted in Figure 2. Now, we have to check that M is indeed an S5-model. Reflexivity and symmetry are immediate consequence of the definition of $\sim_k$. As for transitivity, one can show that, due to the definition of W and $\sim_k$, we have that whenever $w \sim_k w' \sim_k w''$ then either $w = w'$ or $w' = w''$ or $w, w', w'' \in \sim_j [w_0]$ and $k = j$.[12] Then, $w \sim_i w''$ follows easily. Details are left to the reader.

Finally, we can check that, by definition of M, it holds that $M, w_0 \models \theta(\mathcal{M})$ and $M, w_0 \models \hat{\mathbf{K}}_j(\mathbf{K}i\mathsf{r}_{a,b} \wedge \neg\mathsf{r}_{c,d})$, because $w_0 \sim_j w'_j$ and $M, w'_j \models \mathbf{K}i\mathsf{r}_{a,b} \wedge \neg\mathsf{r}_{c,d}$. Note that for showing that the last assertion is true, it is necessary to use both parts of the initial hypothesis, namely $(c,d) \notin R^j$ and $(c,d) \neq (a,b)$.

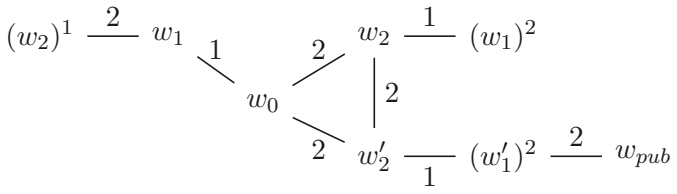


Figure 2: Multi-agent frame for the right-to-left direction of the first item of Proposition 3. We represent the case where $\mathtt{Agt} = \{1,2\}$, $i = 1$ and $j = 2$. Reflexive arrows as well as the direction of the remaining ones are omitted.

□

Second item. [Recall: $(c,d) \in (R^j)^k \cup \{(a,b)\}$ iff $\theta(\mathcal{M}) \to [\mathbf{K}i\mathsf{r}_{a,b}!]\mathbf{K}k\mathbf{K}j\mathsf{r}_{c,d}$ is S5-PAL-valid.]

From left to right: Suppose $(c,d) \in (R^j)^k \cup \{(a,b)\}$. We proceed by cases. If $(c,d) \in \{(a,b)\}$, then $(c,d) = (a,b)$. Since $[\mathbf{K}i\mathsf{r}_{a,b}!]\mathbf{K}k\mathbf{K}j\mathsf{r}_{a,b}$ is S5-PAL-valid (we omit the proof) we get, by propositional reasoning, that $\theta(\mathcal{M}) \to [\mathbf{K}i\mathsf{r}_{a,b}!]\mathbf{K}k\mathbf{K}j\mathsf{r}_{a,b}$ is S5-valid, too. On the other hand, if $(c,d) \in (R^j)^k$, then $\theta(\mathcal{M}) \to \mathbf{K}k\mathbf{K}j\mathsf{r}_{c,d}$ is S5-PAL-valid (by Proposition 1). Note that $\mathbf{K}k\mathbf{K}j\mathsf{r}_{c,d} \to [\psi!]\mathbf{K}k\mathbf{K}j\mathsf{r}_{c,d}$ is S5-valid (in words, second-order knowledge of propositional variables is preserved through public announcements; we omit the proof), so as a particular instance we have that $\mathbf{K}k\mathbf{K}j\mathsf{r}_{c,d} \to [\mathbf{K}i\mathsf{r}_{a,b}!]\mathbf{K}k\mathbf{K}j\mathsf{r}_{c,d}$ is S5-PAL-valid. Then, by propositional reasoning, we have that $\theta(\mathcal{M}) \to [\mathbf{K}i\mathsf{r}_{a,b}!]\mathbf{K}k\mathbf{K}j\mathsf{r}_{c,d}$ is S5-PAL-valid, too.

From right to left: We proceed by cases on possible identities among agents, these are: $i = j = k$, $i = j$ and $i \neq k$; $i = k$ and $i \neq j$; $j = k$ and $i \neq j$; and $i \neq j, i \neq k, j \neq k$.

For the case $i = j = k$, suppose that $\theta(\mathcal{M}) \to [\mathbf{K}i\mathsf{r}_{a,b}!]\mathbf{K}i\mathbf{K}i\mathsf{r}_{c,d}$ is S5-PAL-valid. This amounts to supposing that $\theta(\mathcal{M}) \to \Big(\mathbf{K}i\mathsf{r}_{a,b} \to \mathbf{K}i(\mathbf{K}i\mathsf{r}_{a,b} \to \mathbf{K}i(\mathbf{K}i\mathsf{r}_{a,b} \to \mathsf{r}_{c,d}))\Big)$ is S5-PAL-valid (by the S5-PAL-validity of the reduction axioms). Using the previous

[12] Where $\sim_j [w_0] = \{w \in W \mid w_0 \sim_j w\}$.

validity, the hypothesis $(a,b) \in R^i$, Proposition 1 and S5-axioms, we can deduce that $\theta(\mathcal{M}) \rightarrow \mathbf{K}i\mathbf{K}i\mathsf{r}_{c,d}$ is S5-PAL-valid, which by Proposition 1 is equivalent to $(c,d) \in (R^i)^i$. Since $i = j = k$ by hypothesis, we have that $(c,d) \in (R^j)^k$, which implies $(c,d) \in (R^j)^k \cup \{(a,b)\}$.

For the rest of the cases, we show that the contrapositive holds. Suppose $(c,d) \notin (R^j)^k \cup \{(a,b)\}$. We want to show that the negation of $\theta(\mathcal{M}) \rightarrow [\mathbf{K}i\mathsf{r}_{a,b}!]\mathbf{K}k\mathbf{K}j\mathsf{r}_{c,d}$ is S5-PAL-satisfiable or, equivalently (using reduction axioms and modal reasoning) that $\mathbf{K}i\mathsf{r}_{a,b} \wedge \hat{\mathbf{K}}_k(\mathbf{K}i\mathsf{r}_{a,b} \wedge \hat{\mathbf{K}}_j(\mathbf{K}i\mathsf{r}_{a,b} \wedge \neg\mathsf{r}_{c,d}))$ is S5-satisfiable.

When $i = j$ and $i \neq k$, we have to show that $\mathbf{K}i\mathsf{r}_{a,b} \wedge \hat{\mathbf{K}}_k(\mathbf{K}i\mathsf{r}_{a,b} \wedge \hat{\mathbf{K}}_i(\mathbf{K}i\mathsf{r}_{a,b} \wedge \neg\mathsf{r}_{c,d}))$ is S5-satisfiable. For this, we can use the frame of Figure 2, with $i = j = 1$ and $k = 2$. The valuation is the same than in the proof of the right-to-left direction of the first item (note that we need the hypothesis $(c,d) \notin (R^j)^k \cup \{(a,b)\}$ to show that the target formula is true).

When $i = k$ and $i \neq j$, we have to show that $\mathbf{K}i\mathsf{r}_{a,b} \wedge \hat{\mathbf{K}}_i(\mathbf{K}i\mathsf{r}_{a,b} \wedge \hat{\mathbf{K}}_j(\mathbf{K}i\mathsf{r}_{a,b} \wedge \neg\mathsf{r}_{c,d}))$ is S5-satisfiable. We use the same model than in the previous case, with $i = k = 1$ and $j = 2$. The same comments apply here.

When $j = k$ and $i \neq j$, we have to show that $\mathbf{K}i\mathsf{r}_{a,b} \wedge \hat{\mathbf{K}}_j(\mathbf{K}i\mathsf{r}_{a,b} \wedge \hat{\mathbf{K}}_j(\mathbf{K}i\mathsf{r}_{a,b} \wedge \neg\mathsf{r}_{c,d}))$ is S5-satisfiable. We use the same model than in the previous case, with $i = 1$ and $j = k = 2$. The same comments apply here.

Finally, for $i \neq j, i \neq k, j \neq k$, we have to show that $\mathbf{K}i\mathsf{r}_{a,b} \wedge \hat{\mathbf{K}}_k(\mathbf{K}i\mathsf{r}_{a,b} \wedge \hat{\mathbf{K}}_j(\mathbf{K}i\mathsf{r}_{a,b} \wedge \neg\mathsf{r}_{c,d}))$ is S5-satisfiable. We use the model of Figure 3, with $i = 1$, $j = 2$ and $k = 3$. Again, the hypothesis $(c,d) \notin (R^j)^k \cup \{(a,b)\}$ is crucial to show that the target formula holds at w_0.

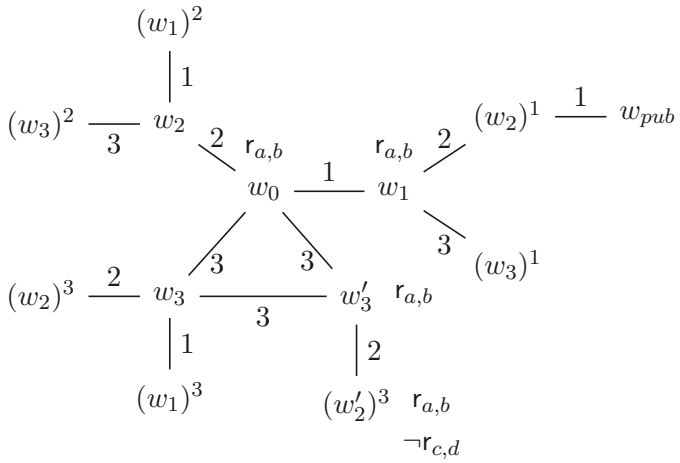


Figure 3: Kripke model for mutually different i, j, k of the right-to-left direction of the second item of Proposition 3. We represent the case where $i = 1$, $j = 2$ and $k = 3$ (see the proof for clarification). Reflexive arrows as well as the direction of the remaining ones are omitted. We only represent the evaluation of atoms that have a key role in the proof.

Third item. [Recall: $(c,d) \in R^{pub} \cup \{(a,b)\}$ iff $\theta(\mathcal{M}) \rightarrow [\mathbf{K}i\mathsf{r}_{a,b}!]\mathbf{C}_{\mathtt{Agt}}\mathsf{r}_{c,d}$ is S5-PAL-valid.]

From left to right: Suppose that $(c,d) \in R^{pub} \cup \{(a,b)\}$. We continue by cases. If $(c,d) \in R^{pub}$, then $\theta(\mathcal{M}) \rightarrow \mathbf{C}_{\mathtt{Agt}}\mathsf{r}_{c,d}$ is S5-valid (by Proposition 1), and hence it is S5-PAL-valid, too. Moreover, note that $\mathbf{C}_{\mathtt{Agt}}\mathsf{r}_{c,d} \rightarrow [\mathbf{K}i\mathsf{r}_{a,b}!]\mathbf{C}_{\mathtt{Agt}}\mathsf{r}_{c,d}$ is S5-PAL-valid (we omit the proof; intuitively, common knowledge of atoms cannot be lost by publicly announcing any formula). By propositional reasoning we have that $\theta(\mathcal{M}) \rightarrow [\mathbf{K}i\mathsf{r}_{a,b}!]\mathbf{C}_{\mathtt{Agt}}\mathsf{r}_{c,d}$ is S5-PAL-valid. As for the other case, suppose that $(c,d) \in \{(a,b)\}$, which is equivalent to $(c,d) = (a,b)$. Then, showing $\theta(\mathcal{M}) \rightarrow [\mathbf{K}i\mathsf{r}_{a,b}!]\mathbf{C}_{\mathtt{Agt}}\mathsf{r}_{c,d}$ is S5-PAL-valid amounts to showing that $\theta(\mathcal{M}) \rightarrow [\mathbf{K}i\mathsf{r}_{a,b}!]\mathbf{C}_{\mathtt{Agt}}\mathsf{r}_{a,b}$ is S5-PAL-valid. Note that $[\mathbf{K}i\mathsf{r}_{a,b}!]\mathbf{C}_{\mathtt{Agt}}\mathsf{r}_{a,b}$ is S5-PAL-valid (we omit the proof), so, by propositional reasoning, $\theta(\mathcal{M}) \rightarrow [\mathbf{K}i\mathsf{r}_{a,b}!]\mathbf{C}_{\mathtt{Agt}}\mathsf{r}_{a,b}$ is S5-PAL-valid, too.

From right to left: We prove the contrapositive statement, using a slight modification of the model employed in the proof of the right-to-left direction of the first item (whose underlying frame is depicted in Figure 2). The definition of the new model adds a world w'_{pub} to the domain of $M^{\mathcal{M}}$, the edges $\{\langle (w'_i)^j, w'_{pub}\rangle, \langle w'_{pub}, (w'_i)^j\rangle, \langle w_{pub}, w'_{pub}\rangle, \langle w'_{pub}, w_{pub}\rangle\}$ to $\sim_j$, and the valuation is extended by setting $V(w'_{pub}) = V(w_{pub}) \cup \{\mathsf{r}_{a,b}\}$. The reader can check that $\sim_j$ is still an equivalence relation. We represent the new model for the case $\mathtt{Agt} = \{i,j\}$ with $i = 1$ and $j = 2$ in Figure 4. So, suppose that $(c,d) \notin R^{pub} \cup \{(a,b)\}$. Note that then $\mathsf{r}_{c,d} \notin V(w_{pub})$ and $\mathsf{r}_{c,d} \notin V(w'_{pub})$ (by definition of the model), so $M^{\mathcal{M}}, w_0 \models \neg\mathbf{C}_{\mathtt{Agt}}\mathsf{r}_{c,d}$. Moreover, after announcing $\mathbf{K}i\mathsf{r}_{a,b}$, the chain $w_0 \sim_j w'_j \sim_i (w'_i)^j \sim_j w'_{pub}$ (in the particular case of Figure 4, the chain $w_0 \sim_2 w'_2 \sim_1 (w'_1)^2 \sim_2 w'_{pub}$) stays in the model, as every of its points verifies $\mathbf{K}i\mathsf{r}_{a,b}$ ($\mathbf{K}1\mathsf{r}_{a,b}$, in the figure) by construction, so $M^{\mathcal{M}}, w_0 \models \neg[\mathbf{K}1\mathsf{r}_{a,b}]\mathbf{C}_{\mathtt{Agt}}\mathsf{r}_{c,d}$, but $M^{\mathcal{M}}, w_0$ still satisfies $\theta(\mathcal{M})$, hence $\theta(\mathcal{M}) \rightarrow [\mathbf{K}i\mathsf{r}_{a,b}!]\mathbf{C}_{\mathtt{Agt}}\mathsf{r}_{c,d}$ is not S5-PAL-valid.

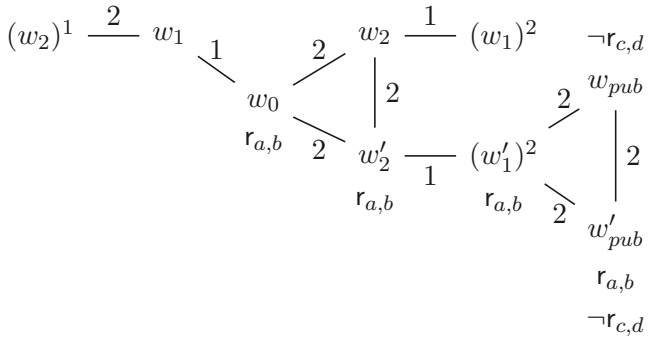


Figure 4: Multi-agent model for the right-to-left direction of the third item of Proposition 3, for the particular case where $\mathtt{Agt} = \{1,2\}$. Reflexive arrows as well as the direction of the remaining ones are omitted. We only depict the valuation of atoms that play a key role in the proof.

Erratum

In this part of the appendix, we correct some typos and minor mistakes spotted in the contributions reprinted in Chapter 4 after they were published.

Paper I

Typos:

- (p. 120, Proof of Proposition 2, first line) "on the construction of φ" should say "on the construction of δ".
- (p.120, Definition 12) "$d(\varphi \wedge \psi) = max(d(\varphi), d(\psi))$" should say "$d(\varphi \wedge \psi) = 1 + max(d(\varphi), d(\psi))$".

Minor mistakes:

- The structures depicted in figures 1 and 2 are not $\mathcal{L}(A)$-models. Actually, they both violate the second constraint of Definition 7. Nevertheless they can be substituted respectively with the models depicted in figures 6 and 7 (which are indeed $\mathcal{L}(A)$-models), without altering the informal description of the corresponding examples.
- (p.109 and p.116) We claimed that, in the Bloody Crime Example, disclosing either d or $\{c, d\}$ would represent an irreparable mistake for agent 1, since he will not be able to reach his goal any more (making argument a strongly accepted for agent 2), but this is false. If he disclosed d, he can still disclose c to reach his goal. If he disclosed $\{c, d\}$, then his goal is already achieved. We apologize for the error, and we also take advantage to point out hat the general phenomenon (irreparable argumentative mistakes due to wrong beliefs) can still be modelled in this framework. Let us see an example. Consider the model M of the top part of Figure 5, where we have abstracted away from the 2-successors of w_0 and w_1 and also have assumed that 1 is aware of all the four arguments at both w_0 and w_1. Suppose that 1 wants to convince 2 of the strong acceptance of a. Note that 1's goal is already achieved, but he believes that this is not the case. Moreover, 1 perceives d as persuasive ($\{d\}$ is an epistemic-based persuasive set, speaking in the paper's terminology). However, if 1 discloses d we obtain the model on the bottom part of Figure 5. Note that in the actual world after disclosing d (w'_0), the following hold:
  - 1's goal is not achieved (2 only accepts a weakly, since $\varnothing$ is a complete extension but $a \notin \varnothing$).
  - However, 1 wrongly things that his goal is achieved.
  - 1's goal is not achievable, since there is nothing new to say (2 is already aware of all arguments).

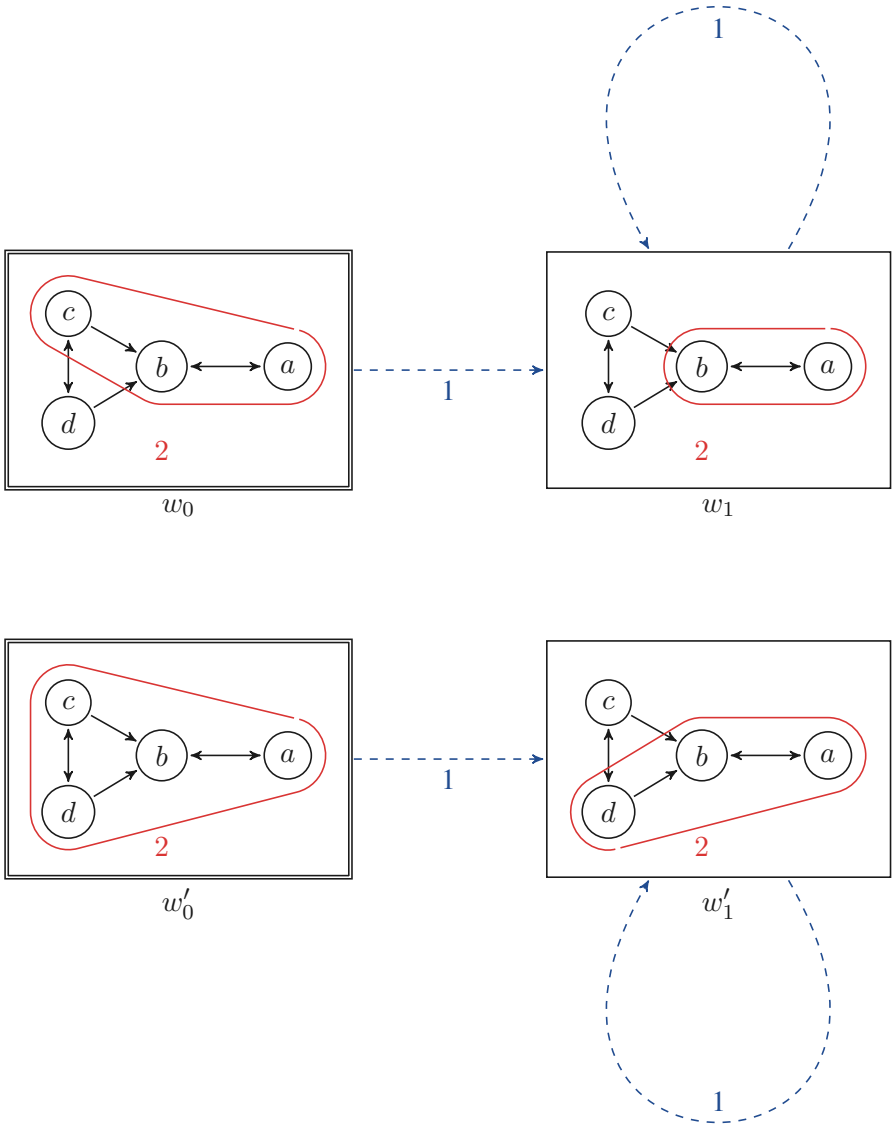


Figure 5: Example of irreparable argumentative mistake

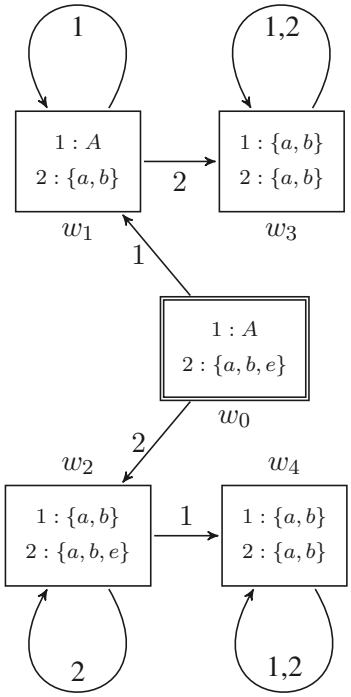


Figure 6: Correction to Figure 2 of Paper I

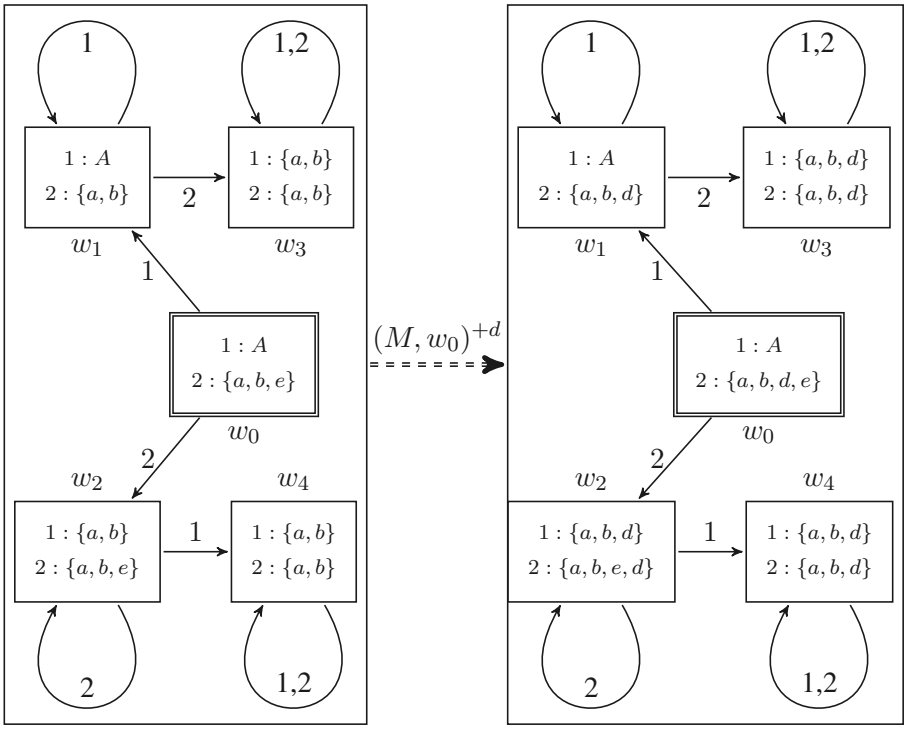


Figure 7: Correction to Figure 3 of Paper I

Paper II

Typos:

- (p.34, Definition 21, first hyphen), "$R \subseteq (A \times A^?) \cup (A \times A^?)$" should say "$R \subseteq (A \cup A^?) \times (A \cup A^?)$". Similarly, in the second hyphen "$R^?, \leftrightarrows \subseteq (A \times A^?) \cup (A \times A^?)$" should say "$R^?, \leftrightarrows \subseteq (A \cup A^?) \times (A \cup A^?)$".

Paper V

Typos:

- (p. 125, Definition 1, second BNF) "$\varphi \in \mathcal{L}_{\mathsf{BA}}$" should say "$\varphi \in \mathcal{F}$".
- (p. 126, Definition of $\mathcal{D}$) Although it is clear by context what we meant, "$\mathcal{D} \subseteq \mathcal{L}^n \times \mathcal{L}$ (with $n \in \mathbb{N}$)" should better say "$\mathcal{D} \subseteq \bigcup_{n \in \mathbb{N}} (\mathcal{F}^n \times \mathcal{F})$" or simply "$\mathcal{D} \subseteq \mathsf{SEQ}(\mathcal{F})$, where $\mathsf{SEQ}(\mathcal{F})$ denotes the set of all finite sequences of formulas" (just as it is done in Paper VI).
- (p. 128) "$\mathsf{Prem}^?(\alpha) := \{\varphi \in \mathsf{Prem}(\alpha) \mid \neg\Box\varphi \wedge \neg\Box\neg\varphi\}$" should say "$\mathsf{Prem}^?(\alpha) := \{\varphi \in \mathsf{Prem}(\alpha) \mid M, w \models \neg\Box\varphi \wedge \neg\Box\neg\varphi\}$".
- (p. 128, in the definition of $\mathcal{A}^?$) "$\mathsf{Prem}^? \neq \varnothing$" should say "$\mathsf{Prem}^?(\alpha) \neq \varnothing$".

Bibliography

L. Amgoud and S. Vesic. A new approach for preference-based argumentation frameworks. *Annals of Mathematics and Artificial Intelligence*, 63(2):149–183, 2011. DOI: 10.1007/s10472-011-9271-9.

L. Amgoud, N. Maudet, and S. Parsons. Modelling dialogues using argumentation. In *Proceedings Fourth International Conference on MultiAgent Systems*, pages 31–38, 2000. DOI: 10.1109/ICMAS.2000.858428.

O. Arieli and M. W. A. Caminada. A QBF-based formalization of abstract argumentation semantics. *J. Appl. Log.*, 11(2):229–252, 2013. DOI: 10.1016/j.jal.2013.03.009.

R. Arisaka, M. Hagiwara, and T. Ito. Formulating manipulable argumentation with intra-/inter-agent preferences. *CoRR*, abs/1909.03616, 2019a. URL http://arxiv.org/abs/1909.03616.

R. Arisaka, M. Hagiwara, and T. Ito. Deception/honesty detection and (mis)trust building in manipulable multi-agent argumentation: An insight. In M. Baldoni, M. Dastani, B. Liao, Y. Sakurai, and R. Zalila-Wenkstern, editors, *PRIMA 2019: Principles and Practice of Multi-Agent Systems*, volume 11873 of *LNCS*, pages 443–451. Springer, 2019b. DOI: 10.1007/978-3-030-33792-6_28.

S. Artemov. Justification awareness models. In S. Artemov and A. Nerode, editors, *Logical Foundations of Computer Science*, volume 10703 of *LNCS*, pages 22–36. Springer, 2018. DOI: 10.1007/978-3-319-72056-2_2.

S. Artemov. Justification awareness. *Journal of Logic and Computation*, 30(8):1431–1446, 2020. DOI: 10.1093/logcom/exaa043.

S. Artemov and M. Fitting. Justification logic. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, 2016.

S. Artemov and E. Nogina. Introducing justification into epistemic logic. *Journal of Logic and Computation*, 15(6):1059–1073, 2005. DOI: https://doi.org/10.1093/logcom/exi053.

G. Aucher. Consistency preservation and crazy formulas in BMS. In S. Hölldobler, C. Lutz, and H. Wansing, editors, *European Workshop on Logics in Artificial Intelligence*, volume 5293 of *LNCS*, pages 21–33. Springer, 2008. DOI: 10.1007/978-3-540-87803-2_4.

G. Aucher. Principles of knowledge, belief and conditional belief. In M. Rebuschi, M. Batt, G. Heinzmann, F. Lihoreau, M. Musiol, and A. Trognon, editors, *Interdisciplinary Works in Logic, Epistemology, Psychology and Linguistics: Dialogue, Rationality, and Formalism*, pages 97–134. Springer, 2014. DOI: 10.1007/978-3-319-03044-9_5.

G. Aucher and F. Schwarzentruber. On the complexity of dynamic epistemic logic. In B. Schipper, editor, *Proceedings of the 14th Theoretical Aspects of Rationality and Knowledge (TARK)*, pages 19–28. ACM, 2013.

P. Balbiani, A. Baltag, H. Van Ditmarsch, A. Herzig, T. Hoshi, and T. De Lima. 'knowable' as 'known after an announcement'. *The Review of Symbolic Logic*, 1(3):305–334, 2008. DOI: 10.1017/S1755020308080210.

P. Balbiani, H. van Ditmarsch, A. Herzig, and T. De Lima. Some truths are best left unsaid. In T. Bolander, T. Braüner, S. Ghilardi, , and L. Moss, editors, *Advances in modal logic*, volume 9, pages 36–54. College Publication, 2012.

A. Baltag and L. S. Moss. Logics for epistemic programs. *Synthese*, 139(2):165–224, 2004. DOI: B:SYNT.0000024912.56773.5e.

A. Baltag and B. Renne. Dynamic Epistemic Logic. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, 2016.

A. Baltag and S. Smets. A qualitative theory of dynamic interactive belief revision. In W. van der Hoek, G. Bonanno, and M. Wooldridge, editors, *Logic and the foundations of game and decision theory (LOFT 7)*, volume 3 of *Texts in Logic and Games*, pages 9–58. Amsterdam University Press, 2008. DOI: 10.1007/978-3-319-20451-2_39.

A. Baltag, L. S. Moss, and S. Solecki. The logic of common knowledge, public announcements, and private suspicions. In I. Gilboa, editor, *Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge (TARK)*, pages 43–56. Morgan Kaufmann Publishers, 1998.

A. Baltag, B. Renne, and S. Smets. The logic of justified belief change, soft evidence and defeasible knowledge. In L. Ong and R. de Queiroz, editors, *Logic, Language, Information and Computation. WoLLIC 2012. LNCS*, volume 7456, pages 168–190. Springer, 2012. DOI: 10.1007/978-3-642-32621-9_13.

A. Baltag, B. Renne, and S. Smets. The logic of justified belief, explicit knowledge, and conclusive evidence. *Annals of Pure and Applied Logic*, 165(1):49–81, 2014. DOI: 10.1016/j.apal.2013.07.005.

A. Baltag, N. Bezhanishvili, A. Özgün, and S. Smets. Justified belief and the topology of evidence. In J. Väänänen, Å. Hirvonen, and R. de Queiroz, editors, *Logic, Language, Information, and Computation*, pages 83–103. Springer, 2016. DOI: 10.1007/978-3-662-52921-8_6.

C. Barés Gómez and M. Fontaine. Argumentation and abduction in dialogical logic. In L. Magnani and T. Bertolotti, editors, *Springer Handbook of Model-Based Science*, pages 295–314. Springer, 2017. DOI: 10.1007/978-3-319-30526-4_14.

P. Baroni and M. Giacomin. On principle-based evaluation of extension-based argumentation semantics. *Artificial Intelligence*, 171(10-15):675–700, 2007. DOI: 10.1016/j.artint.2007.04.004.

P. Baroni, F. Cerutti, M. Giacomin, and G. Guida. Encompassing attacks to attacks in abstract argumentation frameworks. In C. Sossai and G. Chemello, editors, *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 83–94. Springer, 2009. DOI: 10.1007/978-3-642-02906-6_9.

P. Baroni, M. Caminada, and M. Giacomin. Abstract argumentation frameworks and their semantics. In P. Baroni, D. M. Gabbay, M. Giacomin, and L. van der Torre, editors, *Handbook of formal argumentation*, pages 159–236. College Publications, 2018a.

P. Baroni, D. M. Gabbay, M. Giacomin, and L. van der Torre. *Handbook of formal argumentation*. College Publications, 2018b.

R. Baumann. What does it take to enforce an argument? Minimal change in abstract argumentation. In L. D. Raedt, C. Bessiere, D. Dubois, P. Doherty, P. Frasconi, F. Heintz, and P. Lucas, editors, *Proceedings of ECAI*, volume 242 of *Frontiers in Artificial Intelligence and Applications*, pages 127–132. IOS Press, 2012. DOI: 10.3233/978-1-61499-098-7-127.

R. Baumann and G. Brewka. Expanding argumentation frameworks: Enforcing and monotonicity results. In P. Baroni, F. Cerutti, M. Giacomin, and G. R. Simari, editors, *Proceedings of the COMMA 2010*, volume 10 of *Frontiers in Artificial Intelligence and Applications*, pages 75–86. IOS Press, 2010. DOI: 10.3233/978-1-60750-619-5-75.

R. Baumann, G. Brewka, and M. Ulbricht. Revisiting the foundations of abstract argumentation - semantics based on weak admissibility and weak defense. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020*, pages 2742–2749. AAAI Press, 2020.

R. Baumann, S. Doutre, J. Mailly, and J. P. Wallner. Enforcement in formal argumentation. *IFCoLog Journal of Logics and Their Applications*, 8(6):1623–1678, 2021.

D. Baumeister, D. Neugebauer, and J. Rothe. Credulous and skeptical acceptance in incomplete argumentation frameworks. In S. Modgil, K. Budzynska, and J. Lawrence, editors, *Proceedings of the COMMA 2018*, volume 305 of *Frontiers in Artificial*

Intelligence and Applications, pages 181–192. IOS Press, 2018a. DOI: 10.3233/978-1-61499-906-5-181.

D. Baumeister, D. Neugebauer, J. Rothe, and H. Schadrack. Complexity of verification in incomplete argumentation frameworks. In S. A. McIlraith and K. Q. Weinberger, editors, *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18)*, pages 1753–1760. AAAI Press, 2018b.

D. Baumeister, D. Neugebauer, J. Rothe, and H. Schadrack. Verification in incomplete argumentation frameworks. *Artificial Intelligence*, 264:1–26, 2018c. DOI: 10.1016/j.artint.2018.08.001.

D. Baumeister, M. Järvisalo, D. Neugebauer, A. Niskanen, and J. Rothe. Acceptance in incomplete argumentation frameworks. *Artificial Intelligence*, 295:103470, 2021. DOI: 10.1016/j.artint.2021.103470.

M. Beirlaen, J. Heyninck, P. Pardo, and C. Straßer. Argument strength in formal argumentation. *IfCoLog Journal of Logics and their Applications*, 5(3):629–675, 2018.

P. Besnard and S. Doutre. Checking the acceptability of a set of arguments. In J. P. Delgrande and T. Schaub, editors, *Proceedings of the NMR*, pages 59–64. AAAI Press, 2004.

P. Besnard and A. Hunter. A review of argumentation based on deductive arguments. In *Handbook of formal argumentation*, pages 437–484. College Publications, 2018.

P. Besnard, S. Doutre, and A. Herzig. Encoding argument graphs in logic. In A. Laurent, O. Strauss, B. Bouchon-Meunier, and R. Yager, editors, *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, volume 443 of *Communications in Computer and Information Science*, pages 345–354. Springer, 2014a. DOI: 10.1007/978-3-319-08855-6_35.

P. Besnard, A. Garcia, A. Hunter, S. Modgil, H. Prakken, G. Simari, and F. Toni. Introduction to structured argumentation. *Argument & Computation*, 5(1):1–4, 2014b. DOI: 10.1080/19462166.2013.869764.

P. Besnard, C. Cayrol, and M.-C. Lagasquie-Schiex. Logical theories and abstract argumentation: A survey of existing works. *Argument & Computation*, 11(1-2):41–102, 2020. DOI: 10.3233/AAC-190476.

S. Bistarelli and F. Santini. Conarg: A constraint-based computational framework for argumentation systems. In *IEEE 23rd International Conference on Tools with Artificial Intelligence*, pages 605–612. IEEE Computer Society, 2011. DOI: 10.1109/ICTAI.2011.96.

E. Black, A. J. Coles, and C. Hampson. Planning for persuasion. In K. Larson, M. Winikoff, S. Das, and E. H. Durfee, editors, *Proceedings of the 16th Conference*

on Autonomous Agents and MultiAgent Systems, AAMAS 2017, São Paulo, Brazil, May 8-12, 2017, pages 933–942. ACM, 2017.

P. Blackburn, M. De Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, 2002. DOI: 10.1017/CBO9781107050884.

R. Booth, S. Kaci, T. Rienstra, and L. van der Torre. A logical theory about dynamics in abstract argumentation. In W. Liu, V. S. Subrahmanian, and J. Wijsen, editors, *Scalable Uncertainty Management*, volume 8070 of *LNCS*, pages 148–161. Springer, 2013. DOI: 10.1007/978-3-642-40381-1_12.

A. Burrieza and A. Yuste-Ginel. A justification logic for argument evaluation. In M. Blicha and I. Sedlár, editors, *The Logica Yearbook 2018*. College Publications, 2019. ISBN 978-1-84890-307-4.

A. Burrieza and A. Yuste-Ginel. Basic beliefs and argument-based beliefs in awareness epistemic logic with structured arguments. In H. Prakken, S. Bistarelli, F. Santini, and C. Taticchi, editors, *Proceedings of the COMMA 2020*, pages 123–134. IOS Press, 2020. DOI: 10.3233/FAIA200498.

A. Burrieza and A. Yuste-Ginel. Argument evaluation in multi-agent justification logics. *Logic Journal of the IGPL*, 29(4):672–696, 2021. DOI: 10.1093/jigpal/jzz046.

A. Burrieza and A. Yuste-Ginel. An awareness epistemic framework for belief, argumentation and their dynamics. In J. Y. Halpern and A. Perea, editors, *Proceedings Eighteenth Conference on Theoretical Aspects of Rationality and Knowledge (TARK)*, volume 335 of *EPTCS*, pages 69–83, 2021. DOI: 10.4204/EPTCS.335.6.

M. Caminada. On the issue of reinstatement in argumentation. In M. Fisher, W. van der Hoek, B. Konev, and A. Lisitsa, editors, *Logics in Artificial Intelligence. JELIA 2006*, volume 4160 of *LNCS*, pages 111–123. Springer, 2006. DOI: 10.1007/11853886_11.

M. Caminada. Rationality postulates: applying argumentation theory for non-monotonic reasoning. *Journal of Applied Logics*, 4(8):2707–2734, 2017.

M. Caminada and L. Amgoud. On the evaluation of argumentation formalisms. *Artificial Intelligence*, 171(5-6):286–310, 2007. DOI: 10.1016/j.artint.2007.02.003.

M. Caminada and C. Sakama. On the issue of argumentation and informedness. In M. Otake, S. Kurahashi, Y. Ota, K. Satoh, and D. Bekki, editors, *New Frontiers in Artificial Intelligence. JSAI-isAI 2015. LNCS*, volume 10091, pages 317–330. Springer, 2017. DOI: 10.1007/978-3-319-50953-2_22.

M. W. Caminada and D. M. Gabbay. A logical account of formal argumentation. *Studia Logica*, 93(2-3):109–145, 2009. DOI: 10.1007/s11225-009-9218-x.

M. W. A. Caminada, S. Modgil, and N. Oren. Preferences and unrestricted rebut. In S. Parsons, N. Oren, C. Reed, and F. Cerutti, editors, *Proceedings of the COMMA 2014*, Frontiers in Artificial Intelligence and Applications, pages 209–220. IOS Press, 2014. DOI: 10.3233/978-1-61499-436-7-209.

T. Cao Son, E. Pontelli, C. Baral, and G. Gelfond. Exploring the KD45 property of a Kripke model after the execution of an action sequence. In B. Bonet and S. Koenig, editors, *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

Á. Carrera and C. A. Iglesias. A systematic review of argumentation techniques for multi-agent systems research. *Artificial Intelligence Review*, 44(4):509–535, 2015. DOI: 10.1007/s10462-015-9435-9.

C. Cayrol and M.-C. Lagasquie-Schiex. On the acceptability of arguments in bipolar argumentation frameworks. In L. Godo, editor, *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, volume 3571 of *LNCS*, pages 378–389. Springer, 2005. DOI: 10.1007/11518655_33.

C. Cayrol, C. Devred, and M. C. Lagasquie-Schiex. Handling ignorance in argumentation: Semantics of partial argumentation frameworks. In K. Mellouli, editor, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pages 259–270. Springer, 2007. DOI: 10.1007/978-3-540-75256-1_25.

C. Cayrol, F. D. de Saint-Cyr, and M. Lagasquie-Schiex. Change in abstract argumentation frameworks: Adding an argument. *Journal of Artificial Intelligence Research*, 38: 49–84, 2010. DOI: 10.1613/jair.2965.

F. Cerutti, M. Cramer, M. Guillaume, E. Hadoux, A. Hunter, and S. Polberg. Empirical cognitive studies about formal argumentation. In *Handbook of formal argumentation (volume 2)*. College Publications, 2021.

M. Cramer and L. van der Torre. SCF2 - an argumentation semantics for rational human judgments on argument acceptability. In C. Beierle, M. Ragni, F. Stolzenburg, and M. Thimm, editors, *Proceedings of the 8th Workshop on Dynamics of Knowledge and Belief (DKB-2019)*, volume 2445 of *CEUR Workshop Proceedings*, pages 24–35. CEUR-WS.org, 2019.

K. Cyras, C. Schulz, F. Toni, and X. Fan. Assumption-based argumentation: Disputes, explanations, preferences. *IFCoLog Journal of Logics and Their Applications*, 4(8), 2018.

F. D. de Saint-Cyr, P. Bisquert, C. Cayrol, and M.-C. Lagasquie-Schiex. Argumentation update in YALLA (yet another logic language for argumentation). *International Journal of Approximate Reasoning*, 75:57–92, 2016. DOI: 10.1016/j.ijar.2016.04.003.

Y. Dimopoulos, J.-G. Mailly, and P. Moraitis. Control argumentation frameworks. In S. A. McIlraith and K. Q. Weinberger, editors, *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*. AAAI Press, 2018.

S. Doutre and J.-G. Mailly. Constraints and changes: A survey of abstract argumentation dynamics. *Argument & Computation*, 9(3):223–248, 2018. DOI: 10.3233/AAC-180425.

S. Doutre, A. Herzig, and L. Perrussel. A dynamic logic framework for abstract argumentation. In C. Baral, G. De Giacomo, and T. Eiter, editors, *Fourteenth International Conference on the Principles of Knowledge Representation and Reasoning*. AAAI Press, 2014.

S. Doutre, F. Maffre, and P. McBurney. A dynamic logic framework for abstract argumentation: adding and removing arguments. In S. Benferhat, K. Tabia, and M. Ali, editors, *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems.*, volume 10351 of *LNCS*, pages 295–305. Springer, 2017. DOI: 10.1007/978-3-319-60045-1_32.

S. Doutre, A. Herzig, and L. Perrussel. Abstract argumentation in dynamic logic: Representation, reasoning and change. In B. Liao, T. Ågotnes, and Y. N. Wang, editors, *Dynamics, Uncertainty and Reasoning*, pages 153–185. Springer, 2019. DOI: 10.1007/978-981-13-7791-4_8.

P. M. 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. DOI: 10.1016/0004-3702(94)00041-X.

W. Dvořák. On the complexity of computing the justification status of an argument. In S. Modgil, N. Oren, and F. Toni, editors, *Theorie and Applications of Formal Argumentation*, pages 32–49. Springer, 2012. DOI: 10.1007/978-3-642-29184-5_3.

W. Dvorák and P. E. Dunne. Computational problems in formal argumentation and their complexity. In *Handbook of formal argumentation*. College Publications, 2018.

W. Dvořák, S. Szeider, and S. Woltran. Abstract argumentation via monadic second order logic. In E. Hüllermeier, S. Link, T. Fober, and B. Seeger, editors, *Scalable Uncertainty Management*, volume 7520 of *LNCS*, pages 85–98. Springer, 2012. DOI: 10.1007/978-3-642-33362-0_7.

S. K. Dyrkolbotn and T. Pedersen. Arguably argumentative: A formal approach to the argumentative theory of reason. In V. C. Müller, editor, *Fundamental Issues of Artificial Intelligence*, pages 317–339. Springer, 2016. DOI: 10.1007/978-3-319-26485-1_19.

R. Fagin and J. Y. Halpern. Belief, awareness, and limited reasoning. *Artificial intelligence*, 34(1):39–76, 1987. DOI: 10.1016/0004-3702(87)90003-8.

R. Fagin, J. Y. Halpern, Y. Moses, and M. Vardi. *Reasoning about knowledge*. MIT press, 2004. DOI: 10.7551/mitpress/5803.001.0001.

B. Fazzinga, S. Flesca, and F. Furfaro. Revisiting the notion of extension over incomplete abstract argumentation frameworks. In *Proceedings of IJCAI-20*, pages 1712–1718. AAAI Press, 7 2020. DOI: 10.24963/ijcai.2020/237.

D. Fuenmayor Pelaez and A. Steen. A flexible approach to argumentation framework analysis using theorem proving. In *First International Workshop on Logics for New-Generation Artificial Intelligence*, pages 18–32. College Publications, 2021.

D. Gabbay, M. Giacomin, G. R. Simari, and M. Thimm, editors. *Handbook of formal argumentation (volume 2)*. College Publications, 2021.

D. M. Gabbay. Dung's argumentation is essentially equivalent to classical propositional logic with the peirce–quine dagger. *Logica universalis*, 5(2):255, 2011. DOI: 10.1007/s11787-011-0036-3.

D. M. Gabbay and V. B. Shehtman. Products of modal logics, part 1. *Logic journal of IGPL*, 6(1):73–146, 1998. DOI: 10.1093/jigpal/6.1.73.

A. J. García and G. R. Simari. Argumentation based on logic programming. In *Handbook of formal argumentation*. College Publications, 2018.

K. Genin and F. Huber. Formal Representations of Belief. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, 2021.

J. Gerbrandy and W. Groeneveld. Reasoning about information change. *Journal of logic, language and information*, 6(2):147–169, 1997.

E. Gettier. Is justified true belief knowledge? *Analysis*, 23(6):121–123, 1963.

L. Groarke. Informal logic. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, 2017.

D. Grossi. Argumentation in the view of modal logic. In P. McBurney, I. Rahwan, and S. Parsons, editors, *International Workshop on Argumentation in Multi-Agent Systems*, volume 6614 of *LNCS*, pages 190–208. Springer, 2010a. DOI: 10.1007/978-3-642-21940-5_12.

D. Grossi. On the logic of argumentation theory. In W. van der Hoek, G. Kaminka, Y. Lesperance, M. Luck, and S. Sen, editors, *9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, pages 409–416. IFAAMAS, 2010b.

D. Grossi and W. van der Hoek. Justified beliefs by justified arguments. In C. Baral, G. D. Giacomo, and T. Eiter, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourteenth International Conference*. AAAI Press, 2014.

D. Grossi and F. R. Velázquez-Quesada. *Twelve Angry Men*: A study on the fine-grain of announcements. In X. He, J. F. Horty, and E. Pacuit, editors, *Logic, Rationality, and Interaction, Second International Workshop, LORI 2009*, volume 5834 of *LNCS*, pages 147–160. Springer, 2009. DOI: 10.1007/978-3-642-04893-7_12.

D. Grossi and F. R. Velázquez-Quesada. Syntactic awareness in logical dynamics. *Synthese*, 192(12):4071–4105, 2015. DOI: 10.1007/s11229-015-0733-1.

J. Y. Halpern. Alternative semantics for unawareness. *Games and Economic Behavior*, 37(2):321–339, 2001. DOI: 10.1006/game.2000.0832.

J. Y. Halpern and R. Pucella. Dealing with logical omniscience: Expressiveness and pragmatics. *Artificial intelligence*, 175(1):220–235, 2011. DOI: 10.1016/j.artint.2010.04.009.

J. Y. Halpern and L. C. Rêgo. Interactive unawareness revisited. *Games and Economic Behavior*, 62(1):232–262, 2008. DOI: 10.1016/j.geb.2007.01.012.

S. O. Hansson and V. F. Hendricks. *Introduction to formal philosophy*. Springer, 2019.

A. Hasan and R. Fumerton. Foundationalist theories of epistemic justification. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, 2018.

A. Herzig. Dynamic epistemic logics: promises, problems, shortcomings, and perspectives. *Journal of Applied Non-Classical Logics*, 27(3-4):328–341, 2017. DOI: 10.1080/11663081.2017.1416036.

A. Herzig and A. Yuste-Ginel. Abstract argumentation with qualitative uncertainty: An analysis in dynamic logic. In P. Baroni, C. Benzmüller, and Y. N. Wáng, editors, *Logic and Argumentation*, volume 13040 of *LNCS*, pages 190–208. Springer, 2021a. ISBN 978-3-030-89391-0. DOI: 10.1007/978-3-030-89391-0_11.

A. Herzig and A. Yuste-Ginel. Multi-agent abstract argumentation frameworks with incomplete knowledge of attacks. In Z.-H. Zhou, editor, *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pages 1922–1928. IJCAI Organization, 2021b. DOI: 10.24963/ijcai.2021/265.

A. Herzig and A. Yuste-Ginel. On the Epistemic Logic of Incomplete Argumentation Frameworks. In M. Bienvenu, G. Lakemeyer, and E. Erdem, editors, *Proceedings of the 18th International Conference on Principles of Knowledge Representation and Reasoning*, pages 681–685, 11 2021c. DOI: 10.24963/kr.2021/69.

A. Herzig, E. Lorini, and F. Maffre. Possible worlds semantics based on observation and communication. In H. van Ditmarsch and G. Sandu, editors, *Jaakko Hintikka on Knowledge and Game-Theoretical Semantics*, pages 339–362. Springer, 2018. DOI: 10.1007/978-3-319-62864-6_14.

J. Heyninck and C. Straßer. Revisiting unrestricted rebut and preferences in structured argumentation. In C. Sierra, editor, *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pages 1088–1092, 2017. DOI: 10.24963/ijcai.2017/151.

J. Hintikka. *Knowledge and belief: an introduction to the logic of the two notions*. Cornell University Press, 1962.

J. J. Ichikawa and M. Steup. The analysis of knowledge. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, 2018.

T. Kelly. Evidence. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2016 edition, 2016.

K. Konolige. What awareness isn't: A sentential view of implicit and explicit belief. In *Proceedings of the 1st Conference on Theoretical Aspects of Reasoning about Knowledge (TARK)*, pages 241–250, 1986.

B. Kooi. Expressivity and completeness for public update logics via reduction axioms. *Journal of Applied Non-Classical Logics*, 17(2):231–253, 2007. DOI: 10.3166/jancl.17.231-253.

S. A. Kripke. A completeness theorem in modal logic. *The journal of symbolic logic*, 24 (1):1–14, 1959. DOI: 10.2307/2964568.

A. Kurucz, F. Wolter, M. Zakharyaschev, and D. M. Gabbay. *Many-dimensional modal logics: theory and applications*. Gulf Professional Publishing, 2003.

H. J. Levesque. A logic of implicit and explicit belief. In R. J. Brachman, editor, *Proceedings of the National Conference on Artificial Intelligence.*, pages 198–202. AAAI Press, 1984.

X. Li and Y. N. Wáng. A logic of knowledge and belief based on abstract arguments. In M. Dastani, H. Dong, and L. van der Torre, editors, *Logic and Argumentation*, volume 12061 of *LNCS*, pages 116–130. Springer, 2020. DOI: 10.1007/978-3-030-44638-3_8.

E. Lorini and P. Song. Grounding awareness on belief bases. In M. A. Martins and I. Sedlár, editors, *Dynamic Logic. New Trends and Applications*, volume 12569 of *LNCS*, pages 170–186. Springer, 2020. DOI: 10.1007/978-3-030-65840-3_11.

J.-G. Mailly. Yes, no, maybe, i don't know: Complexity and application of abstract argumentation with incomplete knowledge. *Argument & Computation*, (Preprint), 2021a. DOI: 10.3233/AAC-210010.

J.-G. Mailly. Constrained incomplete argumentation frameworks. In J. Vejnarová and N. Wilson, editors, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, volume 12897 of *LNCS*, pages 103–116. Springer, 2021b. DOI: 10.1007/978-3-030-86772-0_8.

H. Mercier and D. Sperber. Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(2):57, 2011. DOI: 10.1017/s0140525x10000968.

H. Mercier and D. Sperber. *The Enigma of Reason*. Harvard University Press, 2017.

J.-J. C. Meyer and W. van der Hoek. *Epistemic logic for AI and computer science*, volume 41 of *Cambridge tracts in theoretical computer science*. Cambridge University Press, 1995.

S. Modgil and H. Prakken. A general account of argumentation with preferences. *Artificial Intelligence*, 195:361–397, 2013.

S. Modgil and H. Prakken. The ASPIC+ framework for structured argumentation: a tutorial. *Argument & Computation*, 5(1):31–62, 2014. DOI: 10.1080/19462166.2013.869766.

S. Modgil and H. Prakken. Abstract rule-based argumentation. In *Handbook of formal argumentation*. College Publications, 2018.

A. Niskanen, D. Neugebauer, and M. Järvisalo. Controllability of control argumentation frameworks. In C. Bessiere, editor, *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, pages 1855–1861. IJCAI Organization, 2020. DOI: 10.24963/ijcai.2020/257.

N. Oren and T. J. Norman. Arguing using opponent models. In P. McBurney, I. Rahwan, S. Parsons, and N. Maudet, editors, *Proceedings of the 6th International Workshop Argumentation in Multi-Agent Systems (ArgMAS)*, volume 6057 of *LNCS*, pages 160–174. Springer, 2009. DOI: 10.1007/978-3-642-12805-9_10.

A. Özgün. *Evidence in epistemic logic: a topological perspective*. PhD thesis, Université de Lorraine, 2017.

E. Pacuit. *Neighborhood semantics for modal logic*. Springer, 2017.

S. Pandžić. A logic of defeasible argumentation: Constructing arguments in justification logic. *Argument & Computation*, (Preprint):1–45, 2019. DOI: 10.3233/AAC-200536.

S. Pandžić. Structured argumentation dynamics. *Annals of Mathematics and Artificial Intelligence*, pages 1–41, 2021. DOI: 10.1007/s10472-021-09765-z.

G. Pappas. Internalist vs. Externalist Conceptions of Epistemic Justification. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, 2017.

J. Plaza. Logics of public announcements. In M. Emrich, M. Pfeifer, M. Hadzikadic, and Z. Ras, editors, *Proceedings 4th International Symposium on Methodologies for Intelligent Systems*, pages 201–216. Oak Ridge National Laboratory, 1989.

J. L. Pollock. Defeasible reasoning. *Cognitive science*, 11(4):481–518, 1987. DOI: 10.1207/s15516709cog1104_4.

H. Prakken. Combining sceptical epistemic reasoning with credulous practical reasoning. In P. E. Dunne and T. J. M. Bench-Capon, editors, *Proceedings of COMMA 2006*, volume 144 of *Frontiers in Artificial Intelligence and Applications*, pages 311–322. IOS Press, 2006.

H. Prakken. An overview of formal models of argumentation and their application in philosophy. *Studies in logic*, 4(1):65–86, 2011.

H. Prakken. Historical overview of formal argumentation. *IfCoLog Journal of Logics and their Applications*, 4(8):2183–2262, 2017.

H. Prakken and M. D. Winter. Abstraction in argumentation: Necessary but dangerous. In S. Modgil, K. Budzynska, and J. Lawrence, editors, *Proceedings of COMMA 2018*, volume 305 of *Frontiers in Artificial Intelligence and Applications*, pages 85–96. IOS Press, 2018. DOI: 10.3233/978-1-61499-906-5-85.

C. Proietti. The dynamics of group polarization. In A. Baltag, J. Seligman, and T. Yamada, editors, *International Workshop on Logic, Rationality and Interaction*, volume 10455 of *LNCS*, pages 195–208. Springer, 2017. DOI: 10.1007/978-3-662-55665-8_14.

C. Proietti and A. Yuste-Ginel. Persuasive argumentation and epistemic attitudes. In L. Soares Barbosa and A. Baltag, editors, *Dynamic Logic. New Trends and Applications*, volume 12005 of *LNCS*, pages 104–123. Springer, 2020. DOI: 10.1007/978-3-030-38808-9_7.

C. Proietti and A. Yuste-Ginel. Dynamic epistemic logics for abstract argumentation. *Synthese*, 199(3):8641–8700, 2021. DOI: 10.1007/s11229-021-03178-5.

I. Rahwan and K. Larson. Argumentation and game theory. In G. Simari and I. Rahwan, editors, *Argumentation in Artificial Intelligence*, pages 321–339. Springer, 2009.

C. Reed and D. Walton. Argumentation schemes in argument-as-process and argument-asproduct. In *Proceedings of the Ontario Society for the Study of Argumentation Conference*, volume 5, 2003.

R. Rendsvig and J. Symons. Epistemic Logic. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2021 edition, 2021.

T. Rienstra, M. Thimm, and N. Oren. Opponent models with uncertainty for strategic argumentation. In F. Rossi, editor, *Twenty-Third International Joint Conference on Artificial Intelligence IJCAI 2013*. AAAI Press, 2013.

C. Sakama. Dishonest arguments in debate games. In B. Verheij, S. Szeider, and S. Woltran, editors, *Proceedings of the COMMA 2012*, Frontiers in Artificial Intelligence and Applications, pages 177–184. IOS Press, 2012.

C. Sakama and T. Cao Son. Epistemic argumentation framework. In A. C. Nayak and A. Sharma, editors, *PRICAI 2019: Trends in Artificial Intelligence*, pages 718–732. Springer, 2019. DOI: 10.1007/978-3-030-29908-8_56.

C. Sakama and T. Cao Son. Epistemic argumentation framework: Theory and computation. *Journal of Artificial Intelligence Research*, 69:1103–1126, 2020. DOI: 10.1613/jair.1.12121.

Y. D. Santos. A dynamic informational-epistemic logic. In A. Madeira and M. R. F. Benevides, editors, *Dynamic Logic. New Trends and Applications*, volume 10669 of *Lncs*, pages 64–81. Springer, 2017. DOI: 10.1007/978-3-319-73579-5_5.

B. C. Schipper. Awareness. In H. van Ditmarsch, J. Y. Halpern, W. van der Hoek, and B. Kooi, editors, *Handbook of Epistemic Logic*. London: College Publications, 2015.

F. Schwarzentruber, S. Vesic, and T. Rienstra. Building an epistemic logic for argumentation. In L. Fariñas del Cerro, A. Herzig, and J. Mengin, editors, *Logics in Artificial Intelligence*, volume 7519 of *LNCS*, pages 359–371. Springer, 2012. DOI: 10.1007/978-3-642-33353-8_28.

C. Shi. *Reason to believe*. PhD thesis, University of Amsterdam, 2018.

C. Shi. No false grounds and topology of argumentation. *Journal of Logic and Computation*, 31(4):1079–1101, 2021. DOI: 10.1093/logcom/exaa057.

C. Shi, S. Smets, and F. Velázquez-Quesada. Argument-based belief in topological structures. In J. Lang, editor, *Proceedings of the Sixteenth Conference on Theoretical Aspects of Rationality and Knowledge (TARK)*, EPTCS. Open Publishing Association, 2017. DOI: 10.4204/EPTCS.251.36.

C. Shi, S. Smets, and F. R. Velázquez-Quesada. Beliefs based on evidence and argumentation. In *Proceedings of WoLLIC 2018*, volume 10944 of *LNCS*, pages 289–306. Springer, 2018. DOI: 0.1007/978-3-662-57669-4_17.

C. Shi, S. Smets, and F. R. Velázquez-Quesada. Logic of justified beliefs based on argumentation. *Erkenntnis*, pages 1–37, 2021. DOI: 10.1007/s10670-021-00399-5.

A. Solaki. Where is epistemic logic in the rationality debate? 2021.

D. Sperber. Intuitive and reflective beliefs. *Mind & Language*, 12(1):67–83, 1997. DOI: 10.1111/j.1468-0017.1997.tb00062.x.

R. Stalnaker. The problem of logical omniscience, I. *Synthese*, pages 425–440, 1991.

R. C. Stalnaker. Possible worlds. *Noûs*, pages 65–75, 1976. DOI: 10.2307/2214477.

M. Thimm. Strategic argumentation in multi-agent systems. *KI-Künstliche Intelligenz*, 28(3):159–168, 2014. DOI: 10.1007/s13218-014-0307-2.

J. van Benthem. *Logical dynamics of information and interaction*. Cambridge University Press, 2011.

J. van Benthem and F. R. Velázquez-Quesada. The dynamics of awareness. *Synthese*, 177 (1):5–27, 2010. DOI: 10.1007/s11229-010-9764-9.

J. van Benthem, J. van Eijck, and B. Kooi. Logics of communication and change. *Information and computation*, 204(11):1620–1662, 2006. DOI: 10.1016/j.ic.2006.04.006.

J. van Benthem, D. Fernández-Duque, E. Pacuit, et al. Evidence logic: A new look at neighborhood structures. volume 9 of *Advances in modal logic*, pages 97–118. College Publications, 2012.

J. van Benthem, D. Fernández-Duque, and E. Pacuit. Evidence and plausibility in neighborhood structures. *Annals of Pure and Applied Logic*, 165(1):106–133, 2014. DOI: 10.1016/j.apal.2013.07.007.

W. van der Hoek, N. Troquard, and M. J. Wooldridge. Knowledge and control. In L. Sonenberg, P. Stone, K. Tumer, and P. Yolum, editors, *10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 719–726. IFAAMAS, 2011.

H. van Ditmarsch and B. Kooi. Semantic results for ontic and epistemic change. In W. van der Hoek, G. Bonanno, and M. Wooldridge, editors, *Logic and the foundations of game and decision theory (LOFT 7)*, volume 3 of *Texts in Logic and Games*, pages 9–58. Amsterdam University Press, 2008.

H. van Ditmarsch, W. van der Hoek, and B. Kooi. *Dynamic epistemic logic*. Springer, 2007.

H. van Ditmarsch, T. French, and F. R. Velázquez-Quesada. Action models for knowledge and awareness. In W. van der Hoek, L. Padgham, V. Conitzer, and M. Winikoff, editors, *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1091–1098. IFAAMAS, 2012.

H. P. van Ditmarsch, W. van der Hoek, and B. P. Kooi. Dynamic epistemic logic with assignment. In F. Dignum, V. Dignum, S. Koenig, S. Kraus, M. P. Singh, and M. J. Wooldridge, editors, *4th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2005), July 25-29, 2005, Utrecht, The Netherlands*, pages 141–148. ACM, 2005. DOI: 10.1145/1082473.1082495.

F. H. van Eemeren, B. Garssen, E. C. W. Krabbe, A. F. Snoeck Henkemans, B. Verheij, and J. H. M. Wagemans. *Handbook of Argumentation Theory*. Springer, 2014. DOI: 10.1007/978-90-481-9473-5.

G. H. Von Wright. An essay in modal logic. 1953.

D. Walton and E. C. Krabbe. *Commitment in dialogue: Basic concepts of interpersonal reasoning*. State University of New York Press, Albany, NY, 1995.

Y. Wang and Q. Cao. On axiomatizations of public announcement logic. *Synthese*, 190 (1):103–134, 2013. DOI: 10.1007/s11229-012-0233-5.

Y. N. Wáng and X. Li. A logic of knowledge based on abstract arguments. *Journal of Logic and Computation*, 2021. DOI: 10.1093/logcom/exab002.

Y. Wu and M. Caminada. A labelling-based justification status of arguments. *Studies in Logic*, 3(4):12–29, 2010.

Z. Yu, K. Xu, and B. Liao. Structured argumentation: Restricted rebut vs. unrestricted rebut. *Studies in Logic*, 11(3):3–17, 2018.

A. Yuste-Ginel. On some formal relations between arguing and believing. In F. Castagna, F. Mosca, J. Mumford, S. Sarkadi, and A. Xydis, editors, *Online Handbook of Argumentation for AI*, volume 2, pages 67–71, 2021.

Resumen

Tema y objetivo. *Creer* y *argumentar* son dos habilidades que juegan típicamente un papel crucial en el análisis de la estructura cognitiva de los humanos. Ambas nociones han recibido notable atención desde muy diversas disciplinas, incluyendo –de forma no exhaustiva– la lingüística, la filosofía, la psicología y las ciencias de la computación. El objetivo general de esta tesis consiste en el estudio, desde una perspectiva lógica (es decir, centrada en el razonamiento), de algunas de las relaciones existentes entre creencias y argumentación. Antes de avanzar, veamos dos ejemplos que ilustran de forma clara el tipo de relaciones en las que estamos pensando.

Ejemplo 1 (El robo de la chocolatina)**.** *Ana está intentando convencer a Roberto de que ella no se comió sin permiso su chocolatina. Para ello, piensa en dos posibles coartadas. De acuerdo con la primera, puede decir que es alérgica a los cacahuetes (ya que la chocolatina robada contenía este fruto seco). De acuerdo con la segunda coartada, puede declarar que se encontraba en su despacho (lejos de la chocolatina) en el momento en que el robo tuvo lugar. Ana cree que Roberto tiene acceso a las cámaras de seguridad de su oficina, pero no puede acceder a su historial médico. En consecuencia, Ana decide utilizar la supuesta alergia como coartada.*

Ejemplo 2 (Ana y el tiempo)**.** *Ana está en su despacho en la Universidad de Málaga. Es una oficina sin ventanas, al ser ella una pobre doctoranda. Ana se está preguntando si está lloviendo. Para averiguarlo, primero consulta a un colega, que le responde "bueno, el cielo parecía despejado cuando yo llegué, hace unas dos horas". Después de esto, abre su navegador y busca la predicción del tiempo en Málaga, que pronostica un 80 % de probabilidad de lluvia. Finalmente, Ana forma la creencia de que está lloviendo ahí fuera.*

Método. La metodología propuesta para el conseguimiento del objetivo general se basa en la combinación de dos familias de formalismos relativamente desconectadas: la *lógica epistémica* (Fagin et al., 2004; Meyer and van der Hoek, 1995) junto con sus extensiones dinámicas (van Ditmarsch et al., 2007; van Benthem, 2011), por un lado, y la *argumentación formal* (Baroni et al., 2018b; Gabbay et al., 2021), por el otro. Creemos que esta elección es natural, ya que la lógica epistémica es un enfoque conocido y exitoso para el modelado cualitativo de actitudes epistémicas (principalmente, conocimiento y creencia); y la argumentación formal es la extensa área de estudio donde se desarrollan representaciones matemáticas de fenómenos argumentativos. Además, nos servimos de las ideas y

herramientas de la *lógica de la conciencia* de Fagin and Halpern (1987), un tipo especial de lógica epistémica, para construir gran parte de esta combinación a lo largo de la tesis.

Estructura. La tesis está presentada como un compendio de publicaciones, es decir, que el corazón de la misma consiste en la reimpresión de la siguiente serie de trabajos ya publicados, ubicada en el Capítulo 4:

1. (Proietti and Yuste-Ginel, 2020) (abreviado I a lo largo de este resumen). Entrada bibliográfica completa: Proietti, C. y Yuste-Ginel, A. (2020). Persuasive argumentation and epistemic attitudes. En Soares Barbosa, L. y Baltag, A., editores, *Dynamic Logic. New Trends and Applications*, volumen 12005 de LNCS, pp. 104–123. Springer. DOI: 10.1007/978-3-030-38808-9_7. Una versión previa del trabajo está disponible en https://eprints.illc.uva.nl/id/eprint/1736/1/Proietti_Yuste_PAEP_preprint.pdf.

2. (Proietti and Yuste-Ginel, 2021) (abreviado II). Entrada bibliográfica completa: Proietti, C. y Yuste-Ginel, A. (2021). Dynamic epistemic logics for abstract argumentation. *Synthese* 199(3): 8641–8700, 2021. DOI: 10.1007/s11229-021-03178-5. Disponible en: https://link.springer.com/content/pdf/10.1007/s11229-021-03178-5.pdf.

3. (Herzig and Yuste-Ginel, 2021c) (abreviado III). Entrada bibliográfica completa: Herzig, A. y Yuste-Ginel, A. (2021). On the Epistemic Logic of Incomplete Argumentation Frameworks. En M. Bienvenu, G. Lakemeyer, and E. Erdem, editores, Proceedings of the 18th International Conference on Principles of Knowledge Representation and Reasoning, pp. 681–685. DOI: 10.24963/kr.2021/69. Disponible en: https://proceedings.kr.org/2021/69/.

4. (Herzig and Yuste-Ginel, 2021b) (abreviado IV). Entrada bibliográfica completa: Herzig, A. y Yuste-Ginel, A. (2021). Multi-agent abstract argumentation frameworks with incomplete knowledge of attacks. En Zhou, Z.-H., editor, *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pp. 1922– 1928. IJCAI Organization. DOI: 10.24963/ijcai.2021/265. Disponible en: https://doi.org/10.24963/ijcai.2021/265.

5. (Burrieza and Yuste-Ginel, 2020) (abreviado V). Entrada bibliográfica completa: Burrieza, A. y Yuste-Ginel, A. (2020). Basic beliefs and argument-based beliefs in awareness epistemic logic with structured arguments. En Prakken et al., editores, *Proceedings of the COMMA 2020*, pp. 123–134. IOS Press. DOI: 10.3233/FAIA200498. Disponible en: https://ebooks.iospress.nl/volumearticle/55364.

6. (Burrieza and Yuste-Ginel, 2021) (abreviado VI). Entrada bibliográfica completa: Burrieza, A. y Yuste-Ginel, A. (2021). An awareness epistemic framework for belief, argumentation and their dynamics. En Halpern F. y Perea A., editores, *Proceedings TARK 2021*, EPTCS 335, pp. 69–83. Open Publishing Association. DOI: 10.4204/EPTCS.335.6. Disponible en: https://doi.org/10.4204/EPTCS.335.6.

El propósito del resto de capítulos es el de tejer una unidad temática entre estas contribuciones. De forma más concreta, dicha unidad se persigue mediante la presentación introductoria del tema y problema de investigación, así como de las intuiciones centrales que subyacen a nuestra forma de modelar creencias y nociones argumentativas (Capítulo 1); una presentación sistemática y crítica de la metodología, en código de preliminares formales (Capítulo 2); la explicación, de forma panorámica, de cómo las publicaciones suponen un tratamiento del problema general de investigación (Capítulo 3); la discusión general de los resultados de todas las contribuciones, trazando puntos de conexión detallados entre ellas así como con otros trabajos de la bibliografía (Capítulo 5); y el establecimiento de unas conclusiones generales y de líneas abiertas para trabajo futuro (Capítulo 6).

Resumen detallado por capítulos. En el resto de este resumen, procedemos a exponer de forma más detallada el contenido de cada uno de los capítulos, deteniéndonos especialmente en los que creemos que pueden ofrecer una imagen más clara del carácter y contenido de la tesis.

En el Capítulo 1, y tras presentar los dos ejemplos motivadores que vimos anteriormente, realizamos, en la Sección 1.1, un análisis general del problema de investigación: el estudio de las relaciones existentes entre *actitudes epistémicas* (con un foco especial en la *creencia*) y *argumentación* (con un foco especial en los procesos de *evaluación de argumentos*). Parece evidente que la relación entre ambas dimensiones cognitivas es bidireccional. Por un lado,

C1 *La evaluación que un agente lleva a cabo de los argumentos que tiene disponibles está influenciada por sus actitudes epistémicas previas (en particular, por sus creencias).*

A modo de ejemplo, pensemos en cómo las creencias de Ana acerca de la información que posee Roberto en el Ejemplo 1 condicionan la elección acerca de qué coartada es mejor utilizar.[13] De hecho, este ejemplo expresa una lectura concreta de C1, véase:

$\mathsf{C1}_{\mathsf{rhetoric}}$ Las *actitudes epistémicas de orden superior*[14] condicionan la *evaluación retórica de argumentos* de dicho agente.[15]

Sin embargo, esta no es la única instancia plausible de C1. Volvamos nuestra atención al Ejemplo 2. Parece claro que Ana concibe el argumento de la predicción del tiempo como estrictamente más fuerte que el testimonio de su colega. Podemos imaginar varias explicaciones para este fenómeno. Una de ellas consiste en afirmar que, mientras Ana cree

[13] Estamos suponiendo que ambas coartadas se pueden entender en términos de sendos contraargumentos a un argumento que concluya la hipotética culpabilidad de Ana.

[14] Es decir, lo que un agente cree (sabe) que otro cree (sabe), o lo que uno cree (sabe) que otro cree (sabe) que otro cree (sabe), etc.

[15] Esto es, la fuerza persuasiva que el agente atribuye a cada argumento.

que la información mostrada en su monitor es correcta (se corresponde con la predicción real), sospecha que su compañero pueda estar mintiendo por alguna razón, es decir, no cree en la sinceridad de su testimonio. En tal caso, el principio evaluativo aplicado por Ana puede sintetizarse como:

$\mathsf{C1}_{\mathsf{epistemic}}$ *los argumentos con premisas creídas (o sabidas verdaderas) deben preferirse a aquellos argumentos cuyas premisas no son creídas (sabidas verdaderas).*

Además, en el mismo ejemplo, se puede apreciar cómo Ana aplica otro principio que pone en relación sus creencias con el proceso de evaluación de sus argumentos disponibles. Véase, cuando Ana forma la creencia de que llueve ahí fuera, lo hace porque el argumento más fuerte de aquellos que ha considerado habla en favor de la verdad de esta afirmación. Este es, probablemente, el principio más popular e intuitivo relacionando creencias y argumentación, y puede plasmarse de forma general como:

C2 un agente epistémico razonable debe tener en cuenta los argumentos que tiene disponibles a favor y en contra de una afirmación, así como la fuerza atribuida a estos, de cara a formar una u otra actitud epistémica con respecto a dicha afirmación (en particular, de cara a formar una u otra creencia).

De forma más breve, y centrándonos en las creencias, el principio sostiene que *la formación de creencias está condicionada por la evaluación de argumentos*.

Después de este sucinto análisis, podemos reformular de manera más precisa el objetivo general de esta tesis: modelar, desde un punto de vista lógico, los principios capturados en ambas instancias de C1 y en C2, mediante el uso combinado de la lógica epistémica y la argumentación formal.

En la Sección 1.2, presentamos las intuiciones que subyacen a los formalismos elegidos para modelar creencias y argumentación. La noción de creencia capturada por la lógica epistémica es *cualitativa* (en el sentido de no numérica), *total* (es decir, no admite grados), y prominentemente *multi-agente*, esto es, permite expresar creencias de orden superior. Por su parte, el objeto primordial de estas creencias (lo que se cree) puede atribuirse a dos entidades distintas: *proposiciones* u *oraciones* [*sentences*]. Nuestra visión, más orientada hacia representaciones sintácticas, se inclina por la segunda opción. Nótese, sin embargo, que en lógica epistémica estándar ambas opciones son equivalentes, ya que los agentes modelados son razonadores perfectos, y esto implica que un agente cree que una oración es verdadera si y sólo si cree que también lo son todas las que le son lógicamente equivalente (es decir, si cree en la proposición que expresan estas oraciones).

En lo que respecta a la visión de fenómenos argumentativos que subyace a las herramientas importadas en esta tesis desde el campo de la argumentación formal, estructuramos nuestra exposición en torno a dos preguntas, que en realidad están presentes en el núcleo de la teoría de la argumentación general (van Eemeren et al., 2014) y, en concreto, de la lógica informal (Groarke, 2017): *¿qué es un argumento?* y *¿cuán fuerte es un argumento?*

La respuesta a la primera pregunta depende fuertemente del formalismo elegido para modelar argumentación. En esta tesis, los argumentos se entienden bien como *nodos* de un

grafo dirigido o bien como *cadenas sintácticas* (que admiten una representación en forma de árbol). En el primer caso, que se corresponde con el uso de los *marcos abstractos de argumentación* [*abstract argumentation frameworks*], introducidos por Dung (1995), nos abstraemos de la estructura interna, origen y naturaleza de los argumentos, para centrarnos en sus relaciones dialécticas (e.g., ataque, derrota o defensa) con los demás argumentos. En el segundo caso, subyacente al uso de ASPIC$^+$ (Modgil and Prakken, 2013) y formalismos afines, entendemos que un argumento es una entidad sintáctica formada usando ciertas oraciones de partida (denominadas *premisas*) mediante la aplicación iterada de reglas de inferencias (ya sean estas deductivas o derrotables [*defeasible*]). Un ejemplo del segundo caso sería el argumento

$$\langle\langle\langle \mathsf{Ave}\rangle, \langle \mathsf{Ave} \to \mathsf{Alas}\rangle \twoheadrightarrow \mathsf{Alas}\rangle \Rightarrow \mathsf{Vuela}\rangle$$

en el cual, partiendo de la premisas 'Este objeto es un ave' y 'Si este objeto es un ave, entonces tiene alas' se infiere deductivamente ($\twoheadrightarrow$) que 'Este objeto tiene alas'. Además, desde esta última oración se infiere, derrotablemente ($\Rightarrow$), que dicho objeto vuela.

Con respecto a la noción de *fuerza argumentativa*, nos apoyamos en el artículo de Beirlaen et al. (2018) que, recogiendo y sistematizando el trabajo de décadas, descompone dicha noción en tres dimensiones distintas. La *dimensión del apoyo* [*support dimension*] se corresponde con la fuerza con la que las premisas y reglas de inferencia de un argumento apoyan las conclusión del mismo. La *dimensión dialéctica* se refiere a las relaciones dialécticas establecidas entre un grupo de argumentos (e.g., ataque, derrota o defensa). Por último, la *dimensión evaluativa* se centra en la pregunta de cómo seleccionar un conjunto de argumentos aceptables partiendo de sus relaciones dialécticas. A modo de ilustración, un principio evaluativo clásico es que dos argumentos no pueden aceptarse simultáneamente si son lógicamente incompatibles, por ejemplo, cuando la conclusión de uno coincide con la negación de la conclusión del otro.

Después de una breve guía de lectura (Sección 1.3), comenzamos el Capítulo 2, en el que se introducen las herramientas técnicas usadas en la tesis. Como señalamos antes, la metodología de la tesis consiste en la combinación de dos familias de formalismos, la lógica epistémica y sus extensiones dinámicas, a la que dedicamos la Sección 2.1, y la argumentación formal, presentada en Sección 2.2.

La *lógica epistémica estándar*, que introducimos a lo largo de la Sección 2.1.1, y cuyo nacimiento se puede remontar a las obras de Von Wright (1953) y Hintikka (1962), se encarga del estudio lógico y cualitativo de las *actitudes epistémicas proposicionales*: saber-que, creer-que, tener-una-opinión-sobre, etc. Desde un punto de vista semántico, la lógica epistémica suele trabajar con *modelos de Kripke multi-agente* (o, simplemente, *modelos epistémicos*), estos son tuplas de la forma $M = (\mathsf{W}, \mathcal{R}, \mathsf{V})$ donde W es un conjunto no vacío de *mundos* o *situaciones posibles*, $\mathcal{R} : \mathsf{Ag} \to \wp(\mathsf{W})$ es una función que asigna una *relación de accesibilidad* $\mathcal{R}_i$ a cada agente i de un conjunto finito y no vacío de agentes Ag; y $\mathsf{V} : \mathsf{At} \to \wp(\mathsf{W})$ es una *función de evaluación* que asigna a cada *variable proposicional* p (esta son, elementos de un conjunto numerable At) un conjunto de mundos $\mathsf{V}(p)$ (intuitivamente, el conjunto de mundos donde p es verdadera). La relación

de accesibilidad del agente i se interpreta como *indiscernibilidad epistémica* (o doxástica). Esto significa que si $(w, v) \in \mathcal{R}_i$, entonces, el agente i, si w fuese el mundo real, consideraría v como un candidato a serlo. Así, dado un *modelo puntuado* (un par (M, w) donde M es un modelo y w un mundo del mismo), y una proposición $P \subseteq \mathsf{W}$, decimos que i cree que P en (M, w) si y sólo si $(w, v) \in \mathcal{R}_i$ implica $v \in P$ (en palabras, i cree que P en (M, w) si y sólo si todos los mundos que considera como candidatos a ser el real son también mundos donde se da P). Desde un punto de vista más lingüístico, estas estructuras sirven para interpretar el *lenguaje epistémico multi-agente* $\mathcal{L}_{\Box}(\mathsf{Ag}, \mathsf{At})$, véase, el generado por la siguiente BNF:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Box_i\varphi \quad p \in \mathsf{At}, i \in \mathsf{Ag}$$

donde las fórmulas de lógica proposicional se leen como usualmente, y el nuevo tipo de fórmulas $\Box_i\varphi$ se lee como "el agente i cree/sabe que φ". Las fórmulas de este lenguaje se interpretan en modelos epistémicos puntuados, siendo la cláusula de verdad para el nuevo operador:

$$M, w \models \Box_i\varphi \text{ si y sólo si } (w, v) \in \mathcal{R}_i \text{ implica } M, v \models \varphi.$$

Así, un agente cree que una fórmula u oración φ es verdadera en (M, w) si y sólo si todos los mundos i-accesibles son mundos donde φ es verdadera.

Como es bien sabido desde Hintikka (1962), en este tipo de lógica epistémica, creer/saber proposiciones y creer/saber oraciones son dos conceptos equivalentes. Es más, los agentes aquí capturados son razonadores perfectos, en el sentido que si creen/saben algo, entonces creen/saben todas las consecuencias de este algo. Para evitar este problema, conocido como omnisciencia lógica, la *lógica de la conciencia* de Fagin and Halpern (1987), que presentamos en la Sección 2.1.2, propone introducir un componente sintáctico en los modelos epistémicos a través de la inclusión de una *función de conciencia* en los mismos. Una *función de conciencia* es una función $\mathsf{Aw} : (\mathsf{Ag} \times \mathsf{W}) \to \wp(\mathcal{L}_{\Box}(\mathsf{Ag}, \mathsf{At}))$ que asigna a cada par (agente, mundo) un conjunto de fórmulas (el conjunto de fórmulas de las que el agente es consciente en ese mundo). Estas funciones de conciencia permiten investigar de una forma flexible la definición de operadores de creencia y conocimiento que evitan el problema de la omnisciencia lógica. En esta tesis, proponemos una aplicación que, parcialmente, se desvía de dicho uso. A lo largo de nuestras contribuciones, con la notable excepción de IV, "elevamos" el rango de la función Aw, del conjunto potencia de todas las fórmulas, al conjunto potencia de todos los *argumentos*, siendo la naturaleza de este conjunto variable de un trabajo a otro, así como variables son las distintas aplicaciones propuestas.

Finalmente, en la Sección 2.1.3, introducimos las herramientas usadas en esta tesis importadas del campo de la *lógica epistémica dinámica* (van Ditmarsch et al., 2007; van Benthem, 2011). Esta puede entenderse como una extensión dinámica –es decir, que estudia cambios– de las estructuras y lenguajes que acabamos de presentar. En concreto, en nuestra exposición nos centramos en los conocidos como modelos de eventos [*event models*] (Baltag and Moss, 2004) enriquecidos con operadores de cambio proposicional (van Benthem et al., 2006; van Ditmarsch and Kooi, 2008). Este tipo de instrumentos

permite modelar cómo las actitudes epistémicas de un grupo de agentes evolucionan a través de diversos tipos de eventos: anuncios públicos, privados, observaciones correctas e incorrectas, sospechas, etc.

Nuestra breve presentación de los conceptos de argumentación formal usados a lo largo de la tesis sigue la distinción entre *modelos de argumentación abstractos* (Sección 2.2.1) y estructurados (Sección 2.2.2). Los modelos abstractos se caracterizan por no tener en cuenta la naturaleza, origen ni estructura de los argumentos modelados. En especial, los *marcos abstractos de argumentación* [*abstract argumentation frameworks*] (Dung, 1995), el más popular de estos modelos, se limita a tomar como elementos primitivos un conjunto de argumentos atómicos A y una relación binaria representando algún tipo de conflicto (típicamente llamado *ataque* o *derrota*) entre estos, formalmente $\mathsf{R} \subseteq \mathsf{A} \times \mathsf{A}$. Así, los marcos abstractos de argumentación sientan las bases formales para plantear la siguiente pregunta: dado un conjunto A de argumentos en conflicto, ¿qué subconjuntos de A deben ser aceptados por un agente racional? Esta pregunta, dicho sea de paso, es central en la dimensión evaluativa de la noción de fuerza argumentativa que mencionamos más arriba. Las distintas respuestas a la misma se formulan en términos de *semánticas argumentativas*. En la Sección 2.2.1 presentamos aquellas semánticas usadas a lo largo de la tesis (las mismas que introdujo Dung (1995)). Además, y partiendo de estas semánticas, ofrecemos distintas alternativas para definir la noción de *estado de justificación* o *aceptación* [*justification status*] de un argumento con respecto a un marco abstracto de argumentación. En cuanto a modelos estructurados de argumentación, nuestra exposición se centra en ASPIC$^+$ (Modgil and Prakken, 2013, 2014), ya que es este el formalismo que usamos como base en algunas de nuestras contribuciones. ASPIC$^+$ es un marco flexible y general –en el sentido de que otros son reducibles a él– que permite construir argumentos partiendo de información (premisas) con distintos niveles de fiabilidad, así como de reglas de inferencia deductivas y derrotables (como en el ejemplo del ave que vimos anteriormente). Además, incorpora en cierto nivel las semánticas argumentativas de Dung (1995), permitiendo capturar de forma simultánea las tres dimensiones de la noción de fuerza argumentativa que guían nuestro análisis desde el inicio.

En el Capítulo 3 elaboramos una explicación, en forma de panorámica, de cómo los trabajos reimpresos en el Capítulo 4 abordan el problema de investigación expuesto en la introducción. Así, aquí solo cabe resumir el primero (que es, en cierto sentido, ya un resumen del segundo). Los trabajos reimpresos están articulados en dos grandes bloques temáticos [*research tracks*] que atienden a la ya expuesta distinción entre modelos abstractos y estructurados de argumentación. El primer bloque, que se centra en modelos abstractos, están conformado por los trabajos I, II, III e IV. En todos ellos, nos dedicamos al análisis de $\mathsf{C1}_{\mathsf{rhetoric}}$. Además, y al tratar este principio desde la perspectiva de modelos abstractos de argumentación, todos estos trabajos tienen un carácter dialéctico, ya que la fuente principal para determinar la fuerza argumentativa de cada elemento consiste en observar sus relaciones de ataque y defensa con los demás.

En I, comenzamos a desarrollar $\mathsf{C1}_{\mathsf{rhetoric}}$ desde la perspectiva de la lógica epistémica. Para ello, primero ofrecemos una definición de persuasión en términos de la alineación,

después de que la comunicación se haya producido, entre el objetivo del hablante y el estado de justificación que el oyente concede a un cierto argumento (representando este último el *tema* de la discusión). Bajo este prisma, el objetivo de Ana en el Ejemplo 1 es que Roberto rechace totalmente el argumento que concluye su culpabilidad. Ana persuadirá a Roberto si, después de presentar su coartada, su objetivo se ve cumplido. Después de esto, mostramos, siguiendo el trabajo de Schwarzentruber et al. (2012), cómo puede usarse la semántica estándar de Kripke para expresar incertidumbre acerca de las bases de conocimiento de otros agentes (entendidas éstas en términos del conjunto de argumentos del que cada agente es consciente). Desde un punto de vista técnico, extendemos uno de sus lenguajes y semánticas de cara a capturar una forma simple de comunicación argumentativa, y procuramos un resultado de completud mediante *axiomas de reducción* (véase (Kooi, 2007)). Nuestra maquinaria formal permite distinguir con precisión los argumentos percibidos como persuasivos por un hablante (los que él cree que le harán lograr su objetivo) de los que realmente lo son. Terminamos el artículo aislando algunas condiciones suficientes para que ambas nociones colapsen, es decir, para que las creencias del agente sean lo suficientemente buenas como para garantizar su éxito en el debate.

En II, extendemos las herramientas lógicas usadas en I en tres direcciones distintas. Primero, aumentamos los tipos de variables proposicionales usados previamente, de cara a capturar en el lenguaje objeto la noción de *estado de justificación* (que antes solo describíamos en el meta-lenguaje). Esto permite, combinado con otros elementos, caracterizar también en el lenguaje objeto los argumentos que son percibidos como persuasivos por parte del hablante. Segundo, estudiamos de forma detallada y sistemática las posibles restricciones a la hora de definir la noción de *marco abstracto de argumentación multi-agente*, y cómo dichas restricciones pueden combinarse de forma adecuada con la semántica epistémica de Kripke. Ofrecemos, además, axiomatizaciones correctas y completas para algunas de estas combinaciones. Tercero, en el aspecto dinámico, saltamos de la acción simple de *comunicar un argumento*, estudiada en I, al marco expresivo de los *modelos de eventos* (Baltag and Moss, 2004) enriquecidos con operadores de cambio proposicional (van Ditmarsch et al., 2005; van Benthem et al., 2006), con la idea de capturar formas de cambio mucho más refinadas, que combinen dinámicas argumentativas y epistémicas. La lógica resultante es lo suficientemente expresiva como para subsumir y generalizar dos formalismos existentes, originalmente diseñados para razonar sobre incertidumbre cualitativa y dinámica argumentativa: los *marcos incompletos de argumentación* [*incomplete argumentation frameworks*] (Baumeister et al., 2018c) y los *marcos argumentativos con control* [*control argumentation frameworks*] (Dimopoulos et al., 2018), así como para razonar de forma sistemática sobre distintas maneras de instanciar la idea del *modelado de oponentes* [*opponent modelling*], central en el subcampo de la argumentación estratégica (Thimm, 2014).

En III, resolvemos un problema que había quedado abierto en nuestra contribución previa: ¿cuál es *la* lógica epistémica que subyace a los marcos incompletos de argumentación? Ofrecemos una respuesta a través del establecimiento de un fuerte nexo formal entre los marcos incompletos de argumentación y la *lógica epistémica de visibilidad* [*epistemic logic of visibility*] de Herzig et al. (2018). Tras esto, extendemos dicho nexo en dos direc-

ciones diferentes. Primero, mostramos cómo los problemas de aceptación de argumentos en marcos incompletos de argumentación pueden ser traducidos de forma natural a problemas de chequeo de modelos en la lógica epistémica de visibilidad. Segundo, presentamos una lógica epistémica mínima para los marcos incompletos de argumentación, para así desmadejar los supuestos epistémicos escondidos tras estas estructuras, que curiosamente resultan ser la consistencia de creencias y la distribución del operador epistémico sobre disyunciones de literales consistentes.

En IV trabajamos siguiendo el mismo espíritu que en las tres contribuciones anteriores, pero moviéndonos de la noción de *conciencia de argumentos* a la de *conocimiento de ataques*, de cara a representar la parte del marco abstracto de argumentación de la que cada agente está informado. Este cambio es interesante, desde un punto de vista teórico, por al menos tres razones. Primero, es más parsimonioso, ya que no requiere la introducción de variables proposicionales indexadas por agentes (como en el caso de la conciencia de argumentos). Esto se debe a que no podemos aplicar de forma natural operadores de conocimiento o creencia proposicional (saber-que-p/creer-que-p) a argumentos (uno puede conocer un argumento o ser consciente del mismo, pero no puede *saber que un argumento*); pero sí podemos hacerlo con ataques (uno puede afirmar naturalmente que sabe que un argumento ataca a otro). Segundo, la noción de conocimiento parcial de ataques encaja bien con algunos escenarios reales, como aquellos en los que los argumentos no son comunicados de forma completa (*entimemas*), o donde los agentes tienen capacidades de razonamiento limitadas, y por tanto fallan en detectar algunos de los ataques. Por último, parece que esta opción de modelado está mucho menos estudiada en la bibliografía (exceptuando el trabajo de Dyrkolbotn and Pedersen (2016)), así que también lo hacemos por el simple placer de la exploración. Además, en este artículo no partimos del marco expresivo de la lógica epistémica estándar, como hicimos antes, sino que lo hacemos desde marcos abstractos de argumentación multi-agente más simples, cuya definición está sin embargo inspirada por la lógica epistémica. Esto hace que nuestro enfoque sea más compacto que los anteriores, y por ende más cercano a la implementación. También mostramos que este nuevo tipo de estructuras son suficientes para modelar algunos casos de comportamiento argumentativo estratégico. El principal resultado técnico consiste en probar que esta versión de marcos de argumentación multi-agente, así como sus semánticas y sus actualizaciones mediante ataques comunicados por los distintos agentes, pueden ser caracterizados usando la conocida *lógica de anuncios públicos* (véanse (Plaza, 1989; van Ditmarsch et al., 2007)).

El segundo bloque de contribuciones, formado por V y VI, se centra en la combinación de la lógica epistémica con modelos estructurados de argumentación. Desde un punto de vista conceptual, recordemos primero que C1 sostiene que la evaluación de argumentos está condicionada por la formación de ciertas actitudes epistémicas, principalmente conocimiento y creencia. En estos trabajos, concretamos la interpretación de dicho principio general, centrándonos en su versión epistémica ($\mathsf{C1}_{\mathsf{epistemic}}$). Nuestras contribuciones continúan un par de publicaciones previas (Burrieza and Yuste-Ginel, 2019, 2021) donde analizamos dicho principio desde el punto de vista de la *lógica de la justificación* (Artemov and Fitting, 2016).

En V, analizamos la relación problemática que se da entre $\mathsf{C1}_{\mathsf{epistemic}}$ y $\mathsf{C2}$.[16] Si son adoptados sin restricciones, estos principios conducen de forma conjunta a una regresión al infinito, al preguntar al agente formalizado: "*¿por qué crees que cierta oración φ es (o no) verdadera?*". Nuestra propuesta para solucionar el problema consiste en distinguir entre creencias *básicas* (no-inferidas) y creencias *basadas en argumentos*, usando para ello un lenguaje formal que importa argumentos tipo ASPIC$^+$ dentro de los modelos de conciencia de Fagin and Halpern (1987). Desde un punto de vista conceptual, nuestro análisis aísla versiones restringidas de $\mathsf{C1}_{\mathsf{epistemic}}$ y $\mathsf{C2}$ que pueden adoptarse de forma conjunta y consistente (sin caer en la mencionada regresión al infinito). Por último, examinamos el formalismo propuesto bajo el prisma de los conocidos *postulados de racionalidad* para sistemas argumentativos de Caminada and Amgoud (2007).

Finalmente, en VI, extendemos primero el trabajo conceptual de nuestro último artículo. En particular, mostramos como este puede ser entendido como una primera aproximación para modelar la distinción entre creencias *intuitivas* y creencias *inferidas* de Sperber (1997) (creencias básicas y basadas en argumentos, en nuestra terminología). Esta distinción se encuentra en los fundamentos de la exitosa teoría argumentativa de la razón defendida por Mercier and Sperber (2011). Además, ampliamos la maquinaria técnica de V en dos direcciones. Primero, damos una axiomatización completa para el fragmento básico del lenguaje estático (esto es, sin el operador de creencia basada en argumentos), y lo extendemos asimismo con varios operadores dinámicos que se pueden eliminar de forma recursiva. En concreto, estudiamos las acciones de *volverse consciente de* y *olvidar* un argumento, *aprender/aceptar una regla derrotable* y *anunciar públicamente una fórmula* (acciones importadas de la tradición de la dinámica de la conciencia sintáctica (Grossi and Velázquez-Quesada, 2009, 2015)). En segundo lugar, extendemos nuestro análisis de postulados de racionalidad, encontrando condiciones bajo las cuales nuestra propuesta se convierte en una instancia de ASPIC$^+$ bien definida, satisfaciendo así todos los postulados de Caminada and Amgoud (2007).

La idea de combinar lógica epistémica y argumentación formal no nació en esta tesis. En el Capítulo 5 comparamos nuestras contribuciones con la bibliografía más cercana, además de analizar los puntos de conexión y divergencia entre las propias contribuciones. Una vez más, procedemos por bloques.

Con respecto a las publicaciones del primer bloque, mostramos primero cómo las lógicas presentadas en I son fragmentos de aquellas estudiadas en II. Asimismo, explicamos cómo adaptar la maquinaria desarrollada en II al tipo de codificación de nociones argumentativas utilizado en III y en IV. Además, recordamos que los modelos usados en ambas contribuciones son casos particulares de los utilizados en II. Esto nos permite concluir que II es el enfoque más general de todos presentados en el primer bloque. Por ello, nos sirve como pivote para llevar a cabo una comparación con obras de la bibliografía que están estrechamente relacionadas, en particular, con los trabajos de Schwarzentruber et al. (2012), Sakama and Cao Son (2020) y Dyrkolbotn and Pedersen (2016).

Con respecto a las publicaciones del segundo bloque, esbozar su relación interna re-

[16] Denominados, respectivamente, $\mathsf{P2}$ y $\mathsf{P1}$ en ambas publicaciones de este bloque.

quiere mucho menos trabajo, ya que una de ellas, VI, es la continuación explícita y directa de la otra, V. Así, nos centramos en su comparación con otras propuestas, de orientación mucho más semántica, para captar la noción de *creencia basada en argumentos* dentro del campo de la lógica epistémica. En particular, analizamos los trabajos de Shi et al. (2017, 2018, 2021); Shi (2021), Li and Wáng (2020); Wáng and Li (2021) y Grossi and van der Hoek (2014). Por último, remarcamos algunas de las características de nuestra adaptación de ASPIC$^+$ (Modgil and Prakken, 2013).

Finalmente, en el Capítulo 6, y después de recapitular, hacemos un par observaciones generales, a modo de conclusión. La primera consiste en señalar el papel esencial que la noción de *conciencia de argumentos* tiene a lo largo de la tesis (con la excepción de IV). Hacemos especial hincapié en las aplicaciones que esta noción tiene, tanto en su en su versión abstracta –servir de base para una teoría general de incertidumbre cualitativa y las multi-agencia acerca de los marcos abstractos de argumentación– como en su versión estructurada –ofrecer una alternativa sintáctica a las propuestas de modelado de creencias basadas en argumentos dentro de la lógica epistémica. La segunda observación se refiere a la noción de fuerza argumentativa, y de cómo nuestro trabajo supone una cierta *epistemización* de la misma. Tras esto, pasamos a esbozar los desafíos pendientes para trabajo futuro (más allá de los señalados al final de cada contribución). El primero se refiere a la introspección de las creencias basadas en argumentos dentro de nuestra propuesta. A diferencia de lo que ocurre con enfoques más semánticos (por ejemplo, con el trabajo de Shi et al. (2021)); aún desconocemos si existe alguna forma adecuada y conceptualmente clara de hacer que nuestros operadores de creencias basadas en argumentos cumplan los principios de introspección positiva e introspección negativa (normalmente asociados a las creencias dentro de la lógica epistémica). La segunda cuestión pendiente es la integración de los dos bloques que conforman esta tesis mediante la definición y el estudio de *modelos epistémicos multi-agentes para argumentación estructurada*. En este capítulo, esbozamos una propuesta parcial y la analizamos de forma breve, señalando sus ventajas e inconvenientes. Por último, mencionamos retos conceptuales que quedan abiertos, por ejemplo, el de extender nuestro estudio a otras de las relaciones existentes entre creencias y argumentación, como la plasmada en la siguiente versión argumentativa del sesgo de confirmación: *argumentar a favor de cosas en las que creo es más fácil que hacerlo a favor de aquellas en las que no creo*.