Taylor & Francis
Taylor & Francis Group

Journal of Applied Non-Classical Logics

ISSN: 1166-3081 (Print) 1958-5780 (Online) Journal homepage: www.tandfonline.com/journals/tncl20

A logic for reasoning about (negative) trust under uncertainty

Francesca Doneda, Francesco A. Genco & Giuseppe Primiero

To cite this article: Francesca Doneda, Francesco A. Genco & Giuseppe Primiero (29 Sep 2025): A logic for reasoning about (negative) trust under uncertainty, Journal of Applied Non-Classical Logics, DOI: 10.1080/11663081.2025.2551479

To link to this article: https://doi.org/10.1080/11663081.2025.2551479

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

Published online: 29 Sep 2025.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tncl20

JOURNAL OF APPLIED NON-CLASSICAL LOGICS
https://doi.org/10.1080/11663081.2025.2551479

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

A logic for reasoning about (negative) trust under uncertainty

Francesca Doneda [a], Francesco A. Genco [a,b] and Giuseppe Primiero [a]

[a]LUCI Lab, Department of Philosophy, University of Milan, Milan, Italy; [b]LLC, Department of Philosophy and Education Sciences, University of Turin, Turin, Italy

ABSTRACT
The notion of trust is a major player, in many epistemic and computational contexts. Such notion appears especially relevant in all those situations where verification or evaluation of knowledge is missing, not reachable or non-existent, and agents must rely on information received by others. This includes cases where expert knowers may not yet be able to ground their claims, and the public has to build an opinion by considering the dynamic of the information exchange. Formal logic approaches to this aim have been increasingly important and diverse in the last decades. A natural and important extension of such multi-agent contexts is where epistemic acts are performed under uncertainty, at several levels: the claims of the agents may be graded; information may reach agents with a certain degree of probability; and the degree of acceptance or rejection of the information received may not be binary. We present a multi-agent logic of negative trust applied to uncertain judgements. We offer a proof theory and a relational semantics for which standard soundness and completeness results hold.

ARTICLE HISTORY
Received 25 June 2024
Accepted 8 July 2025

KEYWORDS
Trust; uncertainty; information exchange; multi-agent systems

1. Introduction

The notion of trust is a major player in many epistemic and computational contexts. Such notion appears especially relevant in all those situations where verification or evaluation of information is missing, not reachable or non-existent, and agents must rely on communication with others. This includes cases where expert knowers may not yet be able to ground their claims, and the public has to build an opinion by considering the dynamic of the information exchange; and also cases in which decisions are taken in complex environments, with data being received from several autonomous agents whose reliability is doubtful or uncertain.

Formal logic approaches to this aim have been increasingly important and diverse in the last decades. Recently, the notion of negative trust has been formalised in terms of a proof theory and semantics of the logic `(un)SecureND` in Primiero (2016, 2020): this approach distinguishes between distrust, as the epistemic act of rejecting incoming contradictory information considered unreliable; and mistrust, as the epistemic act of updating one's own information state by removing previously held data in order

CONTACT Francesca Doneda francesca.doneda@unimi.it

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

to accommodate newly received information. A natural and important extension of such multi-agent contexts is where epistemic acts are performed under uncertainty, at several levels: the claims of the agents may be graded; information may reach agents with a certain degree of probability; and the degree of acceptance or rejection of the information received may not be binary. In the present work we introduce a multi-agent system $\texttt{(un)SecureND}_{grad}$ for modelling complex exchanges of *opinions* and formalising trusting (and non trusting) attitudes with respect to them. Here we stress the fact that $\texttt{(un)SecureND}_{grad}$ extends the previous system by referring to agents' opinions rather than knowledge claims.

To this aim, we model agents as sets of *opinions* that may vary over time. By opinion we mean here someone's personal evaluation or assessment concerning a particular topic. Technically, an opinion is modelled by a proposition associated to a numerical value indicating how much the agent is convinced of its truth. This, in turn, expresses that agents judge topics with a certain degree of (un)certainty. The conviction value is meant to formalise the degree of conviction with which an agent asserts an opinion, and not the force with which such opinion is asserted. Someone might, indeed, wish to assert very strongly that she does not have a strong conviction on a topic. Notice moreover that the degree of conviction cannot be directly related to notions such as that of *force* from *speech act* theory, see Austin (1975). Indeed, the force of an utterance, in this context, usually comprises also qualitative aspects related to the meaning of what is uttered, its communicative significance and the intentions of the utterer, see for instance Holdcroft (1979).

As for the dynamics of opinion exchange, we formalise it by enabling the agents to perform the following actions:

Writing The action of divulging, making public or broadcasting one's opinion – as in the creation of a social network post.

Reading The action of taking into consideration an opinion accessible to the agent.

Notice that an agent's choice of reading or writing an opinion exclusively depends on her.

Trusting (and non trusting) attitudes, on the other hand, are modelled through the following mechanisms that regulate whether an opinion is added or not (and how) to the set of opinions of an agent when she considers (*reads*) it.

Trusting The action, or decision, of accepting a considered opinion, which leads the agent to add it to her opinion set. This action is triggered when the newly considered opinion is consistent with the convictions of the agent.

Distrusting The action, or decision, of rejecting a considered opinion, which leads the agent to reject it and thus not add it to her opinion set. This action is triggered when the newly considered opinion partly conflicts with the agent's convictions and the exposure to it is not strong enough to do otherwise.

Mistrusting The action, or decision, of rejecting some of the agent's present opinions in favour of a newly considered one. This action is triggered when the newly considered opinion partly conflicts with the agent's convictions but the exposure to it is strong enough to make the agent change her mind.

The choice of an agent between distrusting and mistrusting is an aspect of our system in reference to the handling of conflicting opinions inherited from `(un)SecureND` but applied here for the first time to uncertain judgments. Formally, this choice will depend on a numerical value indicating how much the agent is exposed to the newly considered opinion over the total amount of available information: if this value is great enough, the agent will accept the opinion, otherwise, we will have a case of distrust. The underlying intuition is that the more we are exposed to an opinion, the more we are inclined to think that it accurately reflects reality, which is a very well established notion in the psychological literature, see Pennycook and Rand (2020, Section 1.1) for a general discussion of the issue that dates back to some classical works such as Allport and Lepkin (1945), Begg et al. (1992) and Whittlesea (1993). Obviously, other approaches are possible to discriminate between distrust and mistrust, e.g. the influence that the information source has over the receiving agent.

The rest of this paper is structured as follows. In Section 2 we overview the literature related to the present investigation. In Section 3 we introduce the proof-theory of the logic `(un)SecureND`$_{grad}$ and in Section 4 we investigate its meta-theoretical properties. In Section 5 we provide a relational semantics and in Section 6 we illustrate its workings through some examples. We conclude in Section 7 highlighting the next steps of this research.

2. Related work

In the present section we discuss some of the related literature and formal systems with most similar motivations to `(un)SecureND`$_{grad}$, in order to contextualise our analysis, its scope and differences with existing proposals.

The literature concerning the notion of trust is particularly extensive: many definitions have been proposed and discussed, but no consensus has been reached, see Castelfranchi and Falcone (2010). Nevertheless, some important distinctions can be drawn within this broad discussion. We have collected the references to related work by grouping them according to such main differences. The first distinction to be drawn concerns the reference discipline. As suggested in Rousseau et al. (1998) and Sherchan et al. (2013), trust is a multidisciplinary notion that originated in psychology and sociology before entering computer science. While there would be plenty to say about psychological motivations that might trigger any of the operations formally defined in `(un)SecureND`$_{grad}$, our focus will be on the notion of trust intended as a computational feature. Within the latter field, trust can be classified into two general categories: *user trust* and *system trust*. Our work clearly refers to the former. This kind of trust is usually defined as a relational notion concerning subjective expectations (Mui, 2002; Sherchan et al., 2013), while the latter is about the expectations about a system eventually fulfilling its purpose. A last distinction, as remarked in Primiero and Kosolosky (2016), concerns the difference between the definition of trust as a first-order relation among agents or as a second-order property of relations (Demolombe, 2017). In the present work, following Primiero (2020), we consider trust as a second-order property of first-order interactions between agents: in our model, trust is of the form 'Agent *A* trusts the asserted opinion of agent *B*', implying a communication between the two agents.

As far as the technical literature is concerned, many formalisms have been proposed to analyse different facets of the notion of trust. These systems have different characteristics, some of which are related to features of the intended context of application. While not providing a survey, the remainder of this section is concerned with highlighting the most relevant similarities and differences with other formal systems of interest.

The first group of formal systems comprise analyses that arise from the consideration that trust relations are, in some contexts, affected by underlying preexisting hierarchical structures among the considered entities, as originated in the so-called BLP Model (Bell & LaPadula, 1973). The multi-agent systems introduced in Ferraiolo et al. (2001) and Primiero (2020), for instance, feature a hierarchical order between the agents which directly affects trust between them. Technically, the hierarchy influences the possibility that agents have to access pieces of information divulged by others. Another example of systems analysing trust by relying on hierarchical orders concerns the so-called *trust-aware role-based access control models*, developed and studied, for instance, in Bertino et al. (2001), Chandran and Joshi (2005), Chakraborty and Ray (2006) and Oleshchuk (2012). Another body of work concerned with the interplay between trust and hierarchical structures comprises the formalisms introduced in Ceolin and Primiero (2019), Ceolin et al. (2021) and Primiero et al. (2023). Here, a hierarchical structure – more precisely, a *trustworthiness ranking* – is derived from the trust relations, rather than defining them. Technically, Ceolin et al. (2021) and Primiero et al. (2023) builds on a combination of the mechanisms presented in Primiero (2020) and Ceolin and Primiero (2019). $\texttt{(un)SecureND}_{grad}$ departs from such approach in such that it does not define the mechanism of acceptance and rejection of information on the basis of a definable order on agents; rather, acceptance or rejection is based on a comparison between the strength of the message as expressed by the sender and different acceptance and inconsistency thresholds associated to the receiver. We believe this type of information selection mechanism is more appropriate for a variety of contexts of application, especially where information flows freely as, for instance, in news, social environments, and social networks.

A second way to categorise formal approaches to trust is by looking at systems that model its behaviour in terms of a modal logic. In this line, Demolombe (2017) offers another example of work concerned with trustworthiness. A feature of interest of this analysis – which is also shared by Ceolin et al. (2021) and Primiero et al. (2023) – is that it models communication contexts in which information sources may not be reliable, as is, for instance, the one of information access on the web. Just as Demolombe (2017) does, other works analyse trust by modal logic operators, see, for instance, Liau (2003), Demolombe (2004), Lorini and Demolombe (2009), Demolombe (2011), Aldini et al. (2021) and Genco (2024). Among these, trust as a dynamic attitude is defined in Rodenhäuser (2014) by a dynamic epistemic logic operator – see e.g. Van Ditmarsch et al. (2007) – which allows to come to accept information received from a source. As a first remark, it is appropriate to mention that while the operators $\texttt{W}, \texttt{R}, \texttt{Tr}$ and $\texttt{DisTr}$ of $\texttt{(un)SecureND}_{grad}$ can be seen as modal operators, the basic language on which they are defined cannot be simply qualified as propositional. In particular, our language is syntactically equipped with a number of features that enables us to explicitly talk about agents and their actions, frequency

of appearance of messages, and the strength – or conviction – with which an agent sends a particular message. These features technically make it a quantifier-free predicative language. In this sense `(un)SecureND`$_{grad}$ diverges from a standard modal logic approach to trust. Moreover, while the other works only comprise axiomatisations and semantical analyses, the trust operators in our language enjoy both a proof-theoretic and a semantic definition, which enables us to provide also an inferential analysis of their behaviour along with the semantical one.

Further expanding on this observation, another formal characterisation of trust concerns its modelling under uncertainty – namely, when information being divulged and received is not certain. Also with respect to this element, works abound. The work in Kohlas et al. (2008), for instance, is based on the Theory of Probabilistic Argumentation (TPA). The one in Sensoy et al. (2013) instead combines DL-Lite – a tractable subset of Description Logic – and Dempster-Shafer theory of evidence. Ries et al. (2011) offers a model for the evaluation of propositional logic expressions under uncertainty, where the latter is modelled probabilistically. From this point of view, our logic relies on standard probabilistic properties which are not specific to our approach only. Rather, the distinguishing feature of our logic is to be identified in the intended interpretation of values associated to propositions: as shortly mentioned above, we refer to values in the real interval [0, 1] to express the degree of conviction that an agent has in the content under consideration.

Another comparison element with the existing literature is represented by the interaction in the language between agents' opinions and beliefs. Part of the literature precisely analyses trust in the context of belief revision. Yasser and Ismail (2021), for example, argue that in order to formalise trust and belief revision, it is necessary to treat them jointly in an interdependent way. Also Demolombe and Liau (2000) propose a logic for reasoning in a combined way about graded trust and belief, and they focus their analysis on different kinds of trust in information sources. Liau (2003) presents a modal logic system to formalise the influence of trust on the assimilation of information into an agent's belief, which has been extended by Dastani et al. (2005) to also capture the interaction between trust and topics. A different modal logic has been introduced by Lorini et al. (2014) in order to reason about trust-based belief change. In this system, indeed, belief change depends on the degree of trust the receiver has in the source of information. By still using modal logic, Demolombe (2009) proposes a definition of trust based on the truster's belief about the regularity of some trustee's property. This leads to a definition of graded trust in terms of graded belief and graded regularities. A different approach is that of Falcone et al. (2003), who propose a socio-cognitive analysis of the sub-components on which the final decision to trust or not to trust is taken, that is, the basic beliefs of the agents. We note here that while these logics typically delegate the formal expression of information transmission operations to the semantic structure of the underlying logic, our attempt may be considered to deal with this aspect differently. Namely, we explicitly provide operations for agents to read and write information, and define trust and distrust on that basis by considering individual thresholds of acceptability in view of possibly contradictory information previously held. In this sense, trust and its negative counterpart become operations that are instrumental to modify the belief base of the agents, rather than being doxastic operators themselves. Moreover, the explicit definition of the acceptability threshold

for individual pieces of information seems to be a non standard mechanism that our logic introduces, whose closest counterpart could be found in the notion of epistemic entrenchment with respect to sets of formulas from standard belief revision theory. This mechanism to handle the integration of newly received conflicting opinions in the opinion set of an agent is similar in spirit to some other existing methods, see for instance Dubois and Prade (2013). While the latter rely on a possibilistic approach – which is a refinement of a purely logical belief revision method that employs variants of probabilistic notions – we handle the integration of newly received information by employing purely logical means, that is, by using a quantitative notion related to logical inconsistency, see, for instance, Definition 3.4.

Finally, also some purely technical aspects distinguish $\texttt{(un)SecureND}_{grad}$ from a majority of the existing approaches in belief revision. First of all, not all belief revision formalisms rely on real values. Secondly, while real values in the interval [0, 1] are mostly interpreted probabilistically in the context of belief revision, they are employed in $\texttt{(un)SecureND}_{grad}$ in order to quantify the conviction that an agent has in an opinion, without necessarily referring to the probability that the opinion is actually true. Thirdly – and this is a direct consequence of the different interpretations of real values in the two formalisms – while in belief revision adjustments of the values of formulae are usually instrumental to the change of the belief set of an agent, in $\texttt{(un)SecureND}_{grad}$, we can only have an increase in the value associated to a formula and, rather than being instrumental to the change of the set of opinions of an agent, it represents the agent's change of opinions resulting from her actions and those of other agents. Notice, indeed, that in $\texttt{(un)SecureND}_{grad}$ the revision of previously held opinions is established on the basis of their values, but it is implemented by retracting some opinions rather than by adjusting their values.

A slightly different area of analysis concerns the relationship between trust and concepts such as reliability, reputation and sincerity. Muller and Vercouter (2003), for instance, propose a trust-based model for the reliability of agent communications. In this system, agents can detect lies and update their trust attitudes towards other agents on the basis of inconsistencies in the communication. The work in Herzig et al. (2010), is concerned with the study of trust and reputation and their relevant properties. In particular, starting from a distinction of trust into two general types (occurrent trust and dispositional trust), they present a definition of reputation, similar to the definition of trust, based on a concept of belief relative to a collective dimension, that is a concept of group belief. Also the work in Leturc and Bonnet (2018) presents an attempt at extending trust to a collective dimension, but focussing on the notion of sincerity. They propose a modal logic to reason about an agent's trust in the sincerity of a statement formulated by another agent, and then they extend it to a collective notion of shared trust.

On a related note, the work in Hunter (2022) interprets the notion of sincerity from the point of view of the sender in terms of honesty. The receiver keeps track of all information received and stores an account of who is considered honest in order to establish which information should be ignored. In this system, indeed, the receiver is able to infer the honesty of other agents through observation, and this influences, in turn, how she forms beliefs. Finally, in Bentahar et al. (2022) the group dimension is treated by a formal analysis of quantitative group trust. In particular, they employ a

graded branching temporal logic, including quantitative operators for trust within a group, that allows them to express and reason about collective trust in multi-agent systems. Several of these aspects are not modelled in $\texttt{(un)SecureND}_{grad}$ where the only elements of evaluation, from the receiver's point of view, are the conviction that its proponents have in it, the frequency of messages concerning it, and its coherence with the opinions previously held by the receiver. This does not amount to either reliability or honesty, but it expresses a different mechanism to select plausible information. Moreover, $\texttt{(un)SecureND}_{grad}$ does not have yet group operators.

A very active field of investigation essentially concerned with the study and modelling of the structure of interactions between entities is that of *network science*. The level of analysis adopted in network science is usually much more abstract than the relevant one for formalising trust in systems like the one introduced in the present work. Indeed, formalisms concerned with trust mostly deal with exchanges of complex information, often endowed with a propositional structure. On the other hand, network science endeavours mostly focus on the exchange of atomic pieces of information. An excellent attempt at the integration of the fields of agent-based modelling and network science is Namatame and Chen (2016). It provides an interesting distinction that suits us well in order to show the main differences between network science and the present endeavour. It points out that network inspired formalisms like those presented in Kermack et al. (1927), Kephart and White (1991) and Van Mieghem et al. (2009) could be employed for studying the diffusion of information, ideas, and opinions if we are interested in the ways in which they spread. Nevertheless, these models formalise an epidemic-type diffusion, because they were originally meant for the study of epidemic diseases. Since they treat the relations between all pairs of agents as equal, they fail to account for the potential effects of individual heterogeneity on the decision to adopt an idea or opinion. On the other hand, according to Namatame and Chen (2016), works such as Schelling (1978), Granovetter (1978) and Watts (2002) consider a model of diffusion where an agent's behaviour is the result of her deliberate choice. These models take into consideration a threshold determining, for each agent, the proportion of neighbours with which the communication will happen. This is precisely the reason why these models, as compared to those for epidemic-type diffusion, are closer to our system, even though thresholds are used in a slightly different way.

At the intersection of network science studies and the formal and epistemological analysis of trust, the works in Primiero et al. (2016, 2017) and Prandi and Primiero (2020, 2022) present analyses of information diffusion in networks of agents characterised by different epistemic attitudes towards information contents (also depending on ranking criteria) and are defined by logical rules of mistrust and distrust from the logic $\texttt{(un)SecureND}$. Such models have also been applied in contexts of swarm robotics (Primiero et al., 2018) and vehicular ad-hoc networks (Primiero et al., 2018), although yet without uncertainty, for which the present work is the first extension.

3. The proof theory of $\texttt{(un)SecureND}_{grad}$

Before introducing the technical details of the calculus, let us intuitively exemplify a simple interaction between agents of the kind we will formalise by the system.

Example 3.1: Suppose two agents, Amira (A) and Balthazar (B), are discussing the SARS-CoV-2 pandemic. Agent A is not at all convinced by the following opinion

$$\phi = \text{'the situation concerning SARS-CoV-2 is critical'}$$

This formally corresponds to the expression $(\phi^A, 0.1)$ in which we indicate that Amira holds the opinion ϕ with a conviction degree of 0.1. Logically considering the issue, Amira reaches the conclusion that she is actually quite strongly convinced that

$$\neg\phi = \text{'the situation concerning SARS-CoV-2 is not critical'}$$

This formally corresponds to the expression $(\neg\phi^A, 0.9)$ which can be formally obtained from $(\phi^A, 0.1)$ by employing the logical operations that we are going to formalise later – see also Example 3.3.

Now that $(\neg\phi^A, 0.9)$ is explicitly part of Amira's opinions, she decides to make it public by what we call a *write* operation. Since now Balthazar is exposed to this opinion by Amira, he decides to consider it as well by what we call a *read* operation.

Balthazar is actually rather convinced of ϕ. Hence, after considering Amira's opinion $(\neg\phi^A, 0.9)$, he reaches the conclusion that it is too strongly in conflict with his own previous opinions to be simply accepted. Moreover, he considers that it is not worth changing his own opinions just because one agent alone disagrees with him. Therefore, Balthazar distrusts $(\neg\phi^A, 0.9)$ and does not change in any way his own opinions.

We now formally introduce $\texttt{(un)SecureND}_{grad}$ by defining and explaining its language.

Definition 3.1 (Syntax of $\texttt{(un)SecureND}_{grad}$):

$$\begin{aligned}
&\mathcal{A} := A \mid B \mid C \ldots \\
&\mathcal{P}^{\mathcal{A}} := a^{\mathcal{A}} \mid b^{\mathcal{A}} \mid c^{\mathcal{A}} \ldots \\
&\phi := \mathcal{P} \mid (\neg\phi^{\mathcal{A}})^{\mathcal{A}} \mid (\phi^{\mathcal{A}} \to \phi^{\mathcal{A}})^{\mathcal{A}} \mid (\phi^{\mathcal{A}} \wedge \phi^{\mathcal{A}})^{\mathcal{A}} \mid (\phi^{\mathcal{A}} \vee \phi^{\mathcal{A}})^{\mathcal{A}} \\
&\mathcal{E} := (\phi, p) \mid \texttt{W}(\phi, p) \mid \texttt{R}\left((\phi, p), \frac{m}{n}\right) \mid \texttt{Tr}(\phi, p) \mid \texttt{DisTr}(\phi, p) \\
&\mathbb{G}^{\mathcal{A}} := \emptyset \mid \mathbb{G}^{\mathcal{A}}, (\mathcal{P}^{\mathcal{A}}, p) \\
&\mathbb{D} := \emptyset \mid \mathbb{D}, (\phi, p) \\
&\mathbb{L} := \emptyset \mid \mathbb{L}, w(\mathcal{A}, (\phi, p)) \mid \mathbb{L}, r(\mathcal{A}, (\phi, p)) \\
&S := \mathbb{L}; \mathbb{G}; \mathbb{D} \vdash \mathcal{E}
\end{aligned}$$

where $p \in [0, 1]$ and $m, n \in \mathbb{N}$.

Notice that, while a formula ϕ can have different agent name labels occurring in it, all $\mathcal{P}^{\mathcal{A}}$s occurring in an element of $\mathbb{G}^{\mathcal{A}}$ are signed by the same agent name. This is indicated by the fact that ϕ does not explicitly depend on the variable $\mathcal{A}$ while $\mathbb{G}^{\mathcal{A}}$ does.

The set $\mathcal{A}$ contains agent names $A, B, C \ldots$ that we will also use, with a slight abuse of notation, as metavariables for agent names – that is, A and B will be such that possibly

$A = B$. A set $\mathcal{P}^A$, for some $A \in \mathcal{A}$, contains propositional variables labelled (or *signed*) by agent name A. Also in this case, we will use propositional variables as metavariables for propositional variables. The metavariable ϕ^A – we will also use $\psi^A, \chi^A \ldots$ in the following – represents a a formula labelled (or *signed*) by possibly different agent names. These represent the content of opinions, i.e. opinions without any conviction degree assigned, each of which is indexed by the agent asserting it. The set $\mathcal{E}$ contains expressions, with either of the following forms: (ϕ^A, p) representing an opinion along with its degree of conviction; or an operational formula indicating the action of writing, reading, trusting or distrusting an opinion denoted by an expression of the form (ϕ^A, p). Note that an expression of the form $\mathtt{R}((\phi^A, p), \frac{m}{n})$ does not simply denote the action of reading (ϕ^A, p) but, more precisely, the fact that the considered agent has been exposed to the opinion (ϕ^A, p) with a frequency of $\frac{m}{n}$ over all opinions to which she has been exposed. In other terms, we do not only indicate that the agent has been exposed to a certain opinion, but also indicate how often she has been exposed to it. An element of $\mathbb{G}^A$, for some $A \in \mathcal{A}$ – a generic element of which we will denote by Γ^A – is a possibly empty list of the form $(a_1^A, p_1), \ldots, (a_n^A, p_n)$. These lists are used to represent the innate opinions of agent A in the form of a list of atomic opinions. Syntactically complex opinions can then be derived from atomic ones, as explained later in this section. An element of $\mathbb{D}$ is a possibly empty list of the form $(\phi_1^{A_1}, p_1), \ldots, (\phi_1^{A_n}, p_n)$. These lists – denoted by the metavariables $\Delta, \Xi, \Theta \ldots$ – are used to represent the acquired opinions of the agent under consideration. Hence, if we consider a sequent of the form

$$\ldots; \Gamma^A; \Delta \vdash \ldots$$

we regard Γ as the set of innate opinions and Δ as the set of acquired opinions of agent A. The latter are opinions that an agent acquires over time by interacting with other agents. It should be noted that a complex formula occurring in the set Δ may contain subformulae signed by different agent names. The elements of $\mathbb{L}$ – for which we will use the metavariable $\mathfrak{L}$ – are lists of actions of the form $w(A, (\phi^B, p))$ or $r(A, (\phi^B, p))$ indicating, respectively, that agent A has communicated (*written*) the opinion (ϕ^B, p) and that agent A has considered and evaluated (*read*) the opinion (ϕ^B, p). As we will specify in Definition 3.2, these lists, called *literatures*, are going to be used when modelling an exchange of opinions, as records of the actions performed so far. We refer to a set of innate and acquired opinions of an agent along with a suitable literature with the term *agent's profile* and use formally $\mathfrak{L}; \Gamma^A; \Delta :: \textit{Profile}$ to declare one for agent A. The type declaration :: *Profile* is a simple adaptation of the standard type declaration : *Context* used to indicate that what stands to the left of : is a context. While 'context' simply indicates a part of a judgement, we use the term 'profile' to stress that the syntactical object is used here to describe certain features of an agent. In the following we use accordingly either term to refer to the syntactic object or to its intended meaning. The notation ;, on the other hand, is a standard device to separate ,-separated lists inside a context.

A sequent of the form

$$\mathfrak{L}; \Gamma^A; \Delta \vdash (\phi^B, p)$$

has thus the intuitive meaning that, on the basis of what was previously read or written, of the innate opinions contained in Γ^A and of the acquired opinions contained in Δ, agent A assigns the degree p of conviction to the opinion ϕ^B originally asserted by agent B. In other terms, agent A accepts Bs opinion ϕ with a conviction value of p. On the other hand, a sequent of the form

$$\mathfrak{L}; \Gamma^A; \Delta \vdash \mathcal{E}$$

where $\mathcal{E}$ is not of the form (ϕ^B, p) but of the form $\mathtt{W}(\phi^B, p)$, $\mathtt{R}((\phi^B, p), \frac{m}{n})$, $\mathtt{Tr}(\phi^B, p)$, or $\mathtt{DisTr}(\phi^B, p)$ expresses that agent A writes, reads, trusts or distrusts a certain opinion.

Notice that while $\mathtt{W}$ and $\mathtt{R}$ are used as connectives in formulae expressing that an agent *writes* or *reads*, the symbols w and r are exclusively used in expressions of the form $w(A, (\phi^B, p))$ or $r(A, (\phi^B, p))$, which are not, technically, formulae of our language but simply elements of $\mathfrak{L}$. Before introducing the calculus for $\mathtt{(un)SecureND}_{grad}$, let us formally define some useful notions.

Definition 3.2 (Literature and communication records): For any two agents C, C' and expression $(\phi^{C'}, p)$, the following are communication records: $w(C, (\phi^{C'}, p))$ (writing record) and $r(C, (\phi^{C'}, p))$ (reading record). A literature $\mathfrak{L}$ is a list of communication records. For any communication record c and literatures $\mathfrak{L}, \mathfrak{L}'$, we indicate by $\mathfrak{L}^{++c}$ the list obtained by appending c to the list $\mathfrak{L}$; by $\mathfrak{L}^{++\mathfrak{L}'}$ the list obtained by concatenating the list $\mathfrak{L}$ and the list $\mathfrak{L}'$; by $\#\mathfrak{L}(c)$ the number of occurrences of c in $\mathfrak{L}$. We indicate by $\#\mathfrak{L}(w)$, respectively $\#\mathfrak{L}(r)$, the number of writing records, respectively reading records, occurring in $\mathfrak{L}$.

Definition 3.3 (Derivations and subderivations): A *derivation* is inductively defined as follows:

- Any instance of the rules in Figure 1 is a derivation;
- Any application of one of the rules in Figures 2–6 to one or more derivations is a derivation.

The notion of *subderivation* of a derivation δ is inductively defined as follows:

- δ is a subderivation of δ;
- If δ is obtained by applying one of the rules in Figures 2–6 to the derivations $\delta_1, \ldots, \delta_n$, then all subderivations of $\delta_1, \ldots, \delta_n$ are also subderivations of δ.

Contexts $\mathfrak{L}; \Gamma^C; \Delta$ are constructed by the $::$ rule (as far as Γ^A is concerned) and by the $++$ rules (as far as the literature $\mathfrak{L}$ is concerned) introduced in Figure 1. The $::$ rule, indeed, generates a context with an empty literature which can be populated by $++$ rule applications. The *start innate* ($\mathtt{StI}$) and *start acquired* ($\mathtt{StA}$) rules in Figure 1 can either be applied with an empty literature – hence, immediately after a $::$ rule application – or with a non-empty one – when we employ them at the bottom of a derivation in which the operational rules $\mathtt{W}$ and $\mathtt{R}$ have been applied, as we will see below. The rules $\mathtt{StI}$ and $\mathtt{StA}$ enable us to derive any element of Γ^A and of Δ. The

$$\frac{\vdash (a_1^A, p_1) \quad \ldots \quad \vdash (a_n^A, p_n)}{;(a_1^A, p_1), \ldots, (a_n^A, p_n); :: \mathit{Profile}} \; ::$$

where $a_1^A, \ldots, a_n^A$ are all different

$$\frac{\mathfrak{L}; \Gamma^A; \Delta :: \mathit{Profile} \quad p > 0.5}{\mathfrak{L}^{++w(B,(\phi^C,p))}; \Gamma^A; \Delta :: \mathit{Profile}} \; ++ \qquad \frac{\mathfrak{L}; \Gamma^A; \Delta :: \mathit{Profile} \quad w(B, (\phi^C, p)) \in \mathfrak{L}}{\mathfrak{L}^{++r(B',(\phi^C,p))}; \Gamma^A; \Delta :: \mathit{Profile}} \; ++$$

where $A \neq B$

$$\frac{\mathfrak{L}; \Gamma^A, (\phi^A, p); \Delta :: \mathit{Profile}}{\mathfrak{L}; \Gamma^A, (\phi^A, p); \Delta \vdash (\phi^A, p)} \; \texttt{StI} \qquad \frac{\mathfrak{L}; \Gamma^A; \Delta, (\phi^C, p) :: \mathit{Profile}}{\mathfrak{L}; \Gamma^A; \Delta, (\phi^C, p) \vdash (\phi^C, p)} \; \texttt{StA}$$

Figure 1. The System $\texttt{(un)SecureND}_{grad}$: Profile Construction and Start Rules.

$$\frac{\mathfrak{L}; \Gamma^A; \Delta \vdash (\phi^C, p)}{\mathfrak{L}; \Gamma^A; \Delta \vdash ((\neg\phi^C)^A, 1-p)} \; \neg I \qquad \frac{\mathfrak{L}; \Gamma^A; \Delta \vdash ((\neg\phi^C)^{C'}, p)}{\mathfrak{L}; \Gamma^A; \Delta \vdash (\phi^C, 1-p)} \; \neg E$$

$$\frac{\mathfrak{L}; \Gamma^A; \Delta \vdash (\phi^C, p) \quad \mathfrak{L}; \Gamma^A; \Delta \vdash (\psi^{C'}, p')}{\mathfrak{L}; \Gamma^A; \Delta \vdash ((\phi^C \wedge \psi^{C'})^A, \frac{p \cdot p'}{p+p'})} \; \wedge I$$

only when $p + p' \neq 0$

$$\frac{\mathfrak{L}; \Gamma^A; \Delta \vdash ((\phi^C \wedge \psi^{C'})^{C''}, p) \quad \mathfrak{L}; \Gamma^A; \Delta \vdash (\psi^{C'}, p')}{\mathfrak{L}; \Gamma^A; \Delta \vdash (\phi^C, \frac{p' \cdot p}{p'-p})} \; \wedge E$$

$$\frac{\mathfrak{L}; \Gamma^A; \Delta \vdash ((\phi^{C'} \wedge \psi^C)^{C''}, p) \quad \mathfrak{L}; \Gamma^A; \Delta \vdash (\phi^{C'}, p')}{\mathfrak{L}; \Gamma^A; \Delta \vdash (\psi^C, \frac{p' \cdot p}{p'-p})} \; \wedge E$$

only when $p' - p > 0$ and $p' \cdot p \leq p' - p$

Figure 2. The System $\texttt{(un)SecureND}_{grad}$: Logical Rules, 1.

latter case is obviously only possible after Δ has been populated by the operational rules that we will introduce below.

Technically, the $::$ rule enables us to derive a statement of the form $\mathfrak{L}; \Gamma^A; \Delta ::$ *Profile* with $\mathfrak{L} = \Delta = \emptyset$ and Γ^A suitably constructed – that is, only including pairs of the form (a^A, p), each of which assigns a value p to a different atom a. This statement indicates that $\mathfrak{L}; \Gamma^A; \Delta$ can be suitably used to start a derivation syntactically formalising the actions of an agent A. The $++$ rules enable us to add elements to a

$$\frac{\mathfrak{L};\Gamma^A;\Delta \vdash (\phi^C, p) \quad \mathfrak{L};\Gamma^A;\Delta \vdash (\psi^{C'}, p')}{\mathfrak{L};\Gamma^A;\Delta \vdash ((\phi^C \vee \psi^{C'})^A, 1 - \frac{(1-p)\cdot(1-p')}{1-p+1-p'})} \vee I$$

only when $1 - p + 1 - p' \neq 0$

$$\frac{\mathfrak{L};\Gamma^A;\Delta \vdash ((\phi^C \vee \psi^{C'})^{C''}, p) \quad \mathfrak{L};\Gamma^A;\Delta \vdash (\psi^{C'}, p')}{\mathfrak{L};\Gamma^A;\Delta \vdash (\phi^C, 1 - \frac{(1-p')\cdot(1-p)}{(1-p')-(1-p)})} \vee E$$

$$\frac{\mathfrak{L};\Gamma^A;\Delta \vdash ((\phi^{C'} \vee \psi^C)^{C''}, p) \quad \mathfrak{L};\Gamma^A;\Delta \vdash (\phi^{C'}, p')}{\mathfrak{L};\Gamma^A;\Delta \vdash (\psi^C, 1 - \frac{(1-p')\cdot(1-p)}{(1-p')-(1-p)})} \vee E$$

only when $(1 - p') - (1 - p) > 0$ and $(1 - p') \cdot (1 - p) \leq (1 - p') - (1 - p)$

$$\frac{\mathfrak{L};\Gamma^A;\Delta \vdash (\phi^C, p) \quad \mathfrak{L};\Gamma^A;\Delta \vdash (\psi^{C'}, p')}{\mathfrak{L};\Gamma^A;\Delta \vdash ((\phi^C \to \psi^{C'})^A, 1 - \frac{p\cdot(1-p')}{p+1-p'})} \to I$$

only when $p + 1 - p' \neq 0$

$$\frac{\mathfrak{L};\Gamma^A;\Delta \vdash ((\phi^{C'} \to \psi^C)^{C''}, p) \quad \mathfrak{L};\Gamma^A;\Delta \vdash (\phi^{C'}, p')}{\mathfrak{L};\Gamma^A;\Delta \vdash (\psi^C, 1 - \frac{p'\cdot(1-p)}{p'-(1-p)})} \to E$$

only when $p' - (1 - p) > 0$ and $p' \cdot (1 - p) \leq p' - (1 - p)$

The occurrences of C, C', C'' in the elimination rules must be such that either $C = C' = C'' = A$ or $C \neq A$.

Figure 3. The System `(un)SecureND`$_{grad}$: Logical Rules, 2.

$$\frac{\mathfrak{L};\Gamma^A;\Delta \vdash (\phi^C, p) \quad p > 0.5}{\mathfrak{L}^{++w(A,(\phi^C,p))};\Gamma^A;\Delta \vdash \mathtt{W}(\phi^C, p)} \mathtt{W} \qquad \frac{\mathfrak{L}_0;\Gamma^C,\Xi \vdash \mathtt{W}(\phi^{C'}, p) \quad \mathfrak{L}_1;\Gamma^A;\Delta :: \mathit{Profile}}{\mathfrak{L}_1^{++r(A,(\phi^{C'},p))};\Gamma^A;\Delta \vdash \mathtt{R}((\phi^{C'}, p), \frac{m}{n})} \mathtt{R}$$

where $m = \sum_{D\in\mathcal{A}} \#\mathfrak{L}_1(w(D, (\phi^{C'}, p)))$ for the set $\mathcal{A}$ of agent names, $n = \#\mathfrak{L}_1(w)$, and $\mathfrak{L}_1 = \mathfrak{L}_0^{++\mathfrak{L}}$, for some possibly empty literature $\mathfrak{L}$

Figure 4. The System `(un)SecureND`$_{grad}$: Operational Rules, 1.

literature $\mathfrak{L}$ in order to account for the actions executed by an agent B. The condition on the rule requires that B is different from A, the actions of which the derivation is supposed to formalise. This restriction is due to the fact that, while actions of an agent $B \neq A$ only affect the literature $\mathfrak{L}$, the actions executed by A are also supposed to transform the sets of opinions Γ^A and Δ and hence must be formalised by more complex rules – which we will present later in Figures 4–6. Finally, the `StI` and `StA` rules enable us to infer from a statement of the form $\mathfrak{L};\Gamma^A;\Delta :: \mathit{Profile}$ a sequent of the form $\mathfrak{L};\Gamma^A;\Delta \vdash (\phi^D, p)$ – with (ϕ^D, p) either in Γ^A or in Δ – that can be used to start a derivation formalising the actions of an agent A.

The derivability relation $\vdash$ for complex formulae in Γ^A and Δ^A is given by the rules in Figures 2 and 3.

Notice that the introduction rules enable us to compute a value for a complex formula by applying different arithmetical functions to the values related to the immediate subformulae of the complex formula at hand. For instance, $\wedge I$ attributes a value $\frac{p\cdot p'}{p+p'}$ to the conjunction $(\phi^C \wedge \psi^{C'})^{C''}$ in case the premisses attribute values p and p',

respectively, to the formulae ϕ^C and $\psi^{C'}$. The arithmetical functions associated to the elimination rules are simply the duals of those associated to the relative introduction rule. This duality is explicitly established in Theorem 4.1. For some of the connectives, the choice of these arithmetical functions is rather natural; for others, it is the product of adjustments that enabled us to obtain dual functions with respect to which the unit interval $[0, 1]$ is closed. For negation we simply use the complement function $1-p$ on the unit interval $[0, 1]$. For the conjunction, we adopt a modification of the product function – a rather obvious choice, also related to the probability of the conjunction of independent events. The impossibility of using the product of the values of the conjuncts is due to the fact that the unit interval is not closed with respect to the dual of product – that is, division. For the disjunction, we use the complement of the value of the conjunction of the complements of the values of the disjuncts. In other words, we treat disjunction as if defined according to the classical translation $\phi \vee \psi \equiv \neg(\neg\phi \wedge \neg\psi)$. For implication, we use the complement of the value of the conjunction of the value of the antecedent and the complement of the value of the succedent. In other words, we treat implicatoin as if defined according to the classical translation $\phi \rightarrow \psi \equiv \neg(\phi \wedge \neg\psi)$. We remark that the conditions on elimination rules – that is, $p' - p > 0$, $(1 - p') - (1 - p) > 0$ and $p' - (1 - p) > 0$ – correspond to rather basic intuitions on the relationship between the value of a complex formula and the value of its constituents, such as the idea that a conjunction should have a value which is smaller than that of each conjunct. Since the calculus is rather flexible with respect to mixing more or less coherent opinions, these conditions do not imply that the values of all complex formulae must behave like this, but simply impose that, if we wish to use a complex formula and one of its constituents together to derive the value of the other constituent – which reasonably presupposes the decision of considering them, to some extent, coherent – then we must make sure that the relationship between the value of the complex formula and that of the constituent complies with these conditions. Similarly goes for the conditions on the relationship between numerator and denominator of the fractions used here. These are aimed at guaranteeing that the considered values fall inside the unit interval. In other terms, an agent can use a complex formula and one of its constituents in order to derive the value of the other constituent only if the value of the complex formula can be seen as a suitable combination of two values from the $[0, 1]$ interval: the existing value of the constituent and the newly obtained one. If this is not the case, the agent has no way to consider the two existing values as coherent, so to say, and thus cannot use them in combination to obtain a value for the other constituent. Let us exemplify a case of this kind.

Example 3.2: Suppose that $(\phi^A, 0.6)$ is among Amira's (A) innate opinions. Suppose, moreover, that she reads and trusts Balthazar's (B) opinion $((\phi^A \wedge \psi^B)^B, 0.5)$. Notice that Balthazar, in composing the opinion $((\phi^A \wedge \psi^B)^B, 0.5)$, might have used an old opinion by Amira concerning ϕ^A with a conviction value other than 0.6.

Now Amira would like to infer the conviction value of ψ^B by combining the value of the complex opinion $((\phi^A \wedge \psi^B)^B, 0.5)$ and the value of her opinion $(\phi^A, 0.6)$. She cannot do this, though, since $0.6 \cdot 0.5 \not\leq 0.6 - 0.5$ and this indicates that it is impossible to obtain the value 0.5 as the result, for a value $p \in [0, 1]$, of $\frac{0.6+p}{0.6 \cdot p}$ – that is, the value of a conjunction of opinions with values 0.6 and p. Amira, in more intuitive terms, realises

that the values 0.5 for $(\phi^A \wedge \psi^B)^B$ and 0.6 for ϕ^A cannot be considered as coherent since no conjunctive combination of 0.6 and a conviction value can yield 0.5; she thus refrains from using these two values to obtain a value for ψ^B.

Another, more technical, consequence of these conditions is the closure of the unit interval $[0, 1]$ with respect to the arithmetical operations that we use to interpret the connectives. We prove it formally.

Proposition 3.1: *If $p, p' \in [0, 1]$, then all numerical values computed in applying the rules in Figures 2 and 3 are in* $[0, 1]$.

Proof: We reason, case by case, on the single arithmetical expressions employed in Figures 2 and 3.

As for $1-p$, it is clearly in $[0, 1]$ provided that $p \in [0, 1]$.

Let us then consider $\frac{p \cdot p'}{p+p'}$ under the assumption that $p + p' \neq 0$. Now, when $p, p' \in [0, 1]$, $p \cdot p' \leq p + p'$ and both are nonnegative. But this means that $\frac{p \cdot p'}{p+p'} \in [0, 1]$.

Let us consider $\frac{p' \cdot p}{p'-p}$ under the assumption that $p' - p > 0$ and that $p' \cdot p \leq p' - p$. Now, if $p' - p > 0$ and $p, p' \in [0, 1]$, both $p' \cdot p$ and $p' - p$ are nonnegative. Therefore, we just need to prove that $\frac{p' \cdot p}{p'-p} \not> 1$. Suppose, by reasoning indirectly, that $\frac{p' \cdot p}{p'-p} > 1$. Then we need to have also that $p' \cdot p > p' - p$, which contradicts the hypothesis that $p' \cdot p \leq p' - p$. Hence, we can conclude that $\frac{p' \cdot p}{p'-p} \in [0, 1]$.

Let us consider $1 - \frac{(1-p) \cdot (1-p')}{1-p+1-p'}$ under the assumption that $1 - p + 1 - p' \neq 0$. Now, since $1 - q \in [0, 1]$ whenever $q \in [0, 1]$, we can reduce this case to the one of $\frac{p \cdot p'}{p+p'}$.

Let us consider $1 - \frac{(1-p') \cdot (1-p)}{(1-p')-(1-p)}$ under the assumption that $(1 - p') - (1 - p) > 0$ and that $(1 - p') \cdot (1 - p) \leq (1 - p') - (1 - p)$. Now, since $1 - q \in [0, 1]$ whenever $q \in [0, 1]$, we can reduce this case to the one of $\frac{p' \cdot p}{p'-p}$.

Let us consider $1 - \frac{p \cdot (1-p')}{p+1-p'}$ under the assumption that $p + 1 - p' \neq 0$. Now, since $1 - p' \in [0, 1]$ whenever $p' \in [0, 1]$, we can reduce this case to the one of $\frac{p \cdot p'}{p+p'}$.

Let us consider $1 - \frac{p' \cdot (1-p)}{p'-(1-p)}$ under the assumption that $p' - (1 - p) > 0$ and that $p' \cdot (1 - p) \leq p' - (1 - p)$. Now, since $1 - p \in [0, 1]$ whenever $p \in [0, 1]$, we can reduce this case to the one of $\frac{p' \cdot p}{p'-p}$. ■

Finally, we remark that the conditions on C, C' and C'' in the elimination rules guarantee that an agent does not reach conclusions about her own opinions on the basis of opinions that she has received from other agents. Technically, she either exclusively reasons about her own opinions or does not conclude an opinion signed by her.

In order to introduce the rules governing the agents' operations of reading, (not) trusting and writing information, we introduce some simple formal machinery and some useful notions.

$$\frac{\mathfrak{L};\Gamma^A;\Delta \vdash \mathrm{R}((\phi^C,p),\frac{m}{n}) \quad val^{\perp\!\!\!\perp}(\Gamma^A;\Delta,(\phi^C,p)) \leq \varepsilon^A_\perp}{\mathfrak{L};\Gamma^A;\Delta \vdash \mathtt{Tr}(\phi^C,p)}\ \mathtt{Tr}$$

$$\frac{\mathfrak{L};\Gamma^A;\Delta \vdash \mathtt{Tr}(\phi^C,p)}{\mathfrak{L};\Gamma^A;\Delta,(\phi^C,p) :: Profile_{temp}}\ \mathtt{ATr}$$

$$\frac{\mathfrak{L};\Gamma^A;\Delta \vdash \mathrm{R}((\phi^C,p),\frac{m}{n}) \quad val^{\perp\!\!\!\perp}(\Gamma^A;\Delta,(\phi^C,p)) > \varepsilon^A_\perp \quad \frac{m}{n} \leq \varepsilon^A_{freq}}{\mathfrak{L};\Gamma^A;\Delta \vdash \mathtt{DisTr}(\phi^C,p)}\ \mathtt{DisTr}$$

$$\frac{\mathfrak{L};\Gamma^A;\Delta \vdash \mathrm{R}((\phi^C,p),\frac{m}{n}) \quad val^{\perp\!\!\!\perp}(\Gamma^A;\Delta,(\phi^C,p)) > \varepsilon^A_\perp \quad \frac{m}{n} > \varepsilon^A_{freq}}{\mathfrak{L};\Gamma'^A;\Delta',(\phi^C,p) :: Profile_{temp}}\ \mathtt{AMis}$$

where $\Gamma'^A = \Gamma^A \setminus (\Gamma^A \perp\!\!\!\perp_{\varepsilon^A_\perp} (\Delta,(\phi^C,p)))$ and $\Delta' = \Delta \setminus (\Delta \perp\!\!\!\perp_{\varepsilon^A_\perp} (\Gamma^A,(\phi^C,p)))$

Figure 5. The System $\mathtt{(un)SecureND}_{grad}$: Operational Rules, 2.

First of all, to each agent A in our language, we associate three thresholds – that is, real numbers in the unit interval [0, 1]:

$$\varepsilon^A_\perp, \varepsilon^A_{freq}, \varepsilon^A_{int}$$

Intuitively, $\varepsilon^A_\perp$ will be used to determine whether an agent simply accepts a newly considered opinion, namely when the inconsistency value of the new opinion, as defined in Definition 3.4, is not greater than $\varepsilon^A_\perp$; or to determine when she cannot simply accept it, namely when the inconsistency value of the new opinion is greater than $\varepsilon^A_\perp$; in the latter case, then the agent has to decide whether to reject it or to revise some of the previous opinions in order to accept it. This latter choice, in turn, is based on ε^A_{freq}: if the frequency of the newly considered opinion over all opinions to which the agent has access – that is, the frequency $\frac{m}{n}$ employed in rule `R` from Figure 4 and in rules `Tr`,`DisTr`,`AMis` from Figure 5 – is greater than ε^A_{freq}, then the agent will accept it and revise some of the previous opinions; if the frequency of the newly considered opinion over all opinions to which the agent has access is not greater than ε^A_{freq}, then the agent will reject it. Finally, ε^A_{int} will be used to determine whether an agent simply accepts the opinion of another agent when the conviction value of the opinion is not greater than ε^A_{int}; or she also makes the newly acquired opinion a personal opinion by signing it, when the conviction value of the opinion is greater than ε^A_{int}.

The following definitions will introduce the notions of inconsistency value and of inconsistent parts of a profile. They will be used in combination with $\varepsilon^A_\perp$ in order to trigger mistrust operations and to handle them by a revision strategy on the opinions of the agent. This strategy consists in reducing the opinions of the agent in such a way that the inconsistency value of its profile is kept below the threshold $\varepsilon^A_\perp$ of acceptable

$$\frac{\mathfrak{L}; \Gamma^A; \Delta \vdash \mathtt{W}(\phi^C, p)}{\mathfrak{L}; \Gamma^A; \Delta :: \mathit{Profile}} :: \mathtt{W} \qquad \frac{\mathfrak{L}; \Gamma^A; \Delta \vdash \mathtt{DisTr}(\phi^C, p)}{\mathfrak{L}; \Gamma^A; \Delta :: \mathit{Profile}} :: \mathtt{DisTr}$$

$$\frac{\mathfrak{L}; \Gamma^A; \Delta, (\phi^C, p) :: \mathit{Profile}_{temp} \quad p \leq \varepsilon^A_{int}}{\mathfrak{L}; \Gamma^A, \Delta, (\phi^C, p) :: \mathit{Profile}} \ \mathtt{DisI}$$

$$\frac{\mathfrak{L}; \Gamma^A; \Delta, (\phi^C, p) :: \mathit{Profile}_{temp} \quad (\phi^A, q) \in \Delta \quad p \leq q \quad p > \varepsilon^A_{int}}{\mathfrak{L}; \Gamma^A, \Delta, (\phi^C, p) :: \mathit{Profile}} \ \mathtt{DisI}$$

$$\frac{\mathfrak{L}; \Gamma^A; \Delta, (\phi^C, p) :: \mathit{Profile}_{temp} \quad (\phi^A, q) \notin \Delta \quad p > \varepsilon^A_{int}}{\mathfrak{L}; \Gamma^A, \Delta, (\phi^C, p), (\phi^A, p) :: \mathit{Profile}} \ \mathtt{I}$$

where ϕ^A contains at least one agent name different from A

$$\frac{\mathfrak{L}; \Gamma^A; \Delta, (\phi^C, p) :: \mathit{Profile}_{temp} \quad (\phi^A, q) \in \Delta \quad q < p > \varepsilon^A_{int}}{\mathfrak{L}; \Gamma'^A, \Delta', (\phi^C, p), (\phi^A, p) :: \mathit{Profile}} \ \mathtt{I}$$

where $\Delta' = \Delta \setminus \{(\phi^A, q)\}$ and ϕ^A contains at least one agent name different from A

Figure 6. The System $(\mathtt{un})\,\mathtt{SecureND}_{grad}$: Operational Rules, 3.

inconsistency. The role of all these notions will be formally determined by their use in the rules of the calculus.

Definition 3.4 (Inconsistency value of a profile): For any Γ^A and Δ such that $\Gamma^A; \Delta :: \mathit{Profile}$, agent C and formula ϕ, let Γ^{A*}, Δ^* be the sets of expressions obtained from Γ^A, Δ by replacing each element (χ, p) where $p < \frac{1}{2}$ with $(\neg\chi, 1-p)$. Then $\mathit{Set}^{\perp\!\!\!\perp}(\Gamma^A, \Delta, (\phi^C, p))$ is defined as the set of all pairs $(\{\gamma_1, \ldots, \gamma_n, \delta_1, \ldots, \delta_m, \phi\}, i)$ such that

- $n \geq 0 \leq m$
- $\{(\gamma_1^A, q_1), \ldots, (\gamma_n^A, q_n)\} \subseteq \Gamma^{A*}$
- $\{(\delta_1^{C_1}, r_1), \ldots, \delta_m^{C_m}, r_m)\} \subseteq \Delta^*$
- $\{\gamma_1, \ldots, \gamma_n, \delta_1, \ldots, \delta_m, \phi\}$ is a classically inconsistent set
- all proper subsets of $\{\gamma_1, \ldots, \gamma_n, \delta_1, \ldots, \delta_m, \phi\}$ are consistent
- $i = q_1 \cdot \ldots \cdot q_n \cdot r_1 \cdot \ldots \cdot r_m \cdot p$

Moreover, $\mathit{val}^{\perp\!\!\!\perp}(\Gamma^A, \Delta, (\phi^C, p)) = \max(\{j \mid (Z, j) \in \mathit{Set}^{\perp\!\!\!\perp}(\Gamma^A, \Delta, (\phi^C, p))\})$.

Definition 3.5 (Inconsistent parts of a profile): For any real number ε, $\Gamma^A \perp\!\!\!\perp_\varepsilon (\Delta, (\phi^C, p))$ is defined as the subset of Γ^A that contains all pairs in Γ^A whose first element occurs in any set Z such that $(Z, j) \in \mathit{Set}^{\perp\!\!\!\perp}(\Gamma^A, \Delta, (\phi^C, p))$ and $j > \varepsilon$. For any real number ε, $\Delta \perp\!\!\!\perp_\varepsilon (\Gamma^A, (\phi^C, p))$ is defined as the subset of Δ that contains all pairs in Δ the first element of which occurs in any set Z such that $(Z, j) \in \mathit{Set}^{\perp\!\!\!\perp}(\Gamma^A, \Delta, (\phi^C, p))$ and $j > \varepsilon$.

In Figure 4, we present the operational rules that enable us to derive statements concerning the actions agents execute. The rule $\mathtt{W}$ enables us to derive a statement describing a writing action executed by the agent A that we are considering at the

moment. The literature $\mathfrak{L}$ is updated accordingly. The rule `R` similarly enables us to derive a statement concerning a reading action executed by *A*. We recall that 'reading' here is meant to refer to the action of taking into consideration an opinion and not simply to the action of physically reading about an opinion – which might have already happened without the agent properly conducting an evaluation of the opinion itself. Therefore, when an agent *reads* an opinion, we attach to it its frequency $\frac{m}{n}$ over all opinions to which the agent has access. This frequency indicates indeed the degree of exposure that the agent had to that opinion, and this will influence the attitude of the agent towards that opinion. Notice that, in the rule `W`, if *A* is writing an opinion by an agent *C* different from *A*, then we will have that the literature contains a record of *A* reading that opinion and, even before, one of *C* writing it. There is no need to specify this by the premisses of the rule since it is a consequence of how the rules of the calculus are defined.

In Figure 5, we present the operational rules for the the agent to react to a new opinion that she has considered (*read)*. Notice that $Profile_{temp}$ is simply a notational device employed inside the calculus to indicate that the opinion set of an agent is being modified, but does not yet faithfully represent the consequences of the last performed action. Without it, the corresponding states in the semantics would have no counterpart, hindering our completeness proof.

The rule `Tr` – for *trust* – enables us to derive a statement establishing that agent *A* trusts the opinion that has been read. The rule `ATr` – for *acquire* because of a *trusting attitude* – enables us to integrate in the agent's profile an opinion which has been trusted by the agent. The rule `DisTr` – for *distrust* – enables us to derive a statement establishing that agent *A* does not trust an opinion which has been considered by the agent. Finally, the rule `AMis` – for *acquire* because of a *mistrusting attitude* – enables us to formalise a mistrust attitude towards a read opinion, that is, we can update the profile of the agent in case an opinion has been considered and accepted – formally indicated by $\frac{m}{n} > \varepsilon^{A}_{freq}$ – even though it conflicts with previous opinions of the agent. Notice that the usage of $\perp\!\!\!\perp_{\varepsilon^{A}_{\perp}}$ that we do here guarantees that the inconsistency values of the profile at the conclusion of the rule are below the inconsistency threshold $\varepsilon^{A}_{\perp}$. Notice that the definition of $\perp\!\!\!\perp_{\varepsilon^{A}_{\perp}}$ could be modified in order to remove elements from the old profile in a more minimal way – as it were – while still keeping inconsistencies involving the newly acquired opinion at bay. On the one hand, the strategy adopted here is more costly in terms of loss of information since we remove all elements of the inconsistent sets which exceed the amount of inconsistency accepted by the agent. On the other hand, this strategy has the advantage that it does not require us to take into account the content of opinions in order to establish a criterion for deciding which opinions should be removed and which should be kept.

In Figure 6, we present the operational rules that enable us to complete the update of an agent's profile after an opinion has been read and its content considered. The rules `::W` and `::DisTr` simply leave the profile as it was since the agent has simply written an opinion and has not read anything or, respectively, not trusted a considered opinion. These rules are required since *writing* and *distrusting* operations do not change the profile of the agent – indeed, after these operations are executed, no internalisation of opinions is in order – but something in the

derivation still needs to signal that the agent's profile is reinitialised for the following operation.

The rules `I` – for *internalisation* – add to the profile a statement (ϕ^A, p) indicating that agent A personally embraces ϕ and will not simply report C's assertion of ϕ and agree with it but will also be disposed to directly assert ϕ as a personal opinion. The conditions on ϕ^A in these rules are meant to exclude cases in which the agent A internalises some of her own opinions because someone else cited her. Obviously, the case in which A internalises one of her own present opinions is not problematic – in any case, A's present opinions can be derived from Γ^A and Δ by `StA`, `StI` and logical rules – but it is rather problematic, from a conceptual perspective, if A changes her own present opinions on the basis of the fact that some other agent cited an *old* opinion of A. Hence, we introduce conditions that exclude this.

The rules `DisI` – for *no internalisation* – leave the profile unchanged since the accepted opinion (ϕ^C, p) is already in the updated profile and it is not to be internalised by the agent, that is, the agent will not personally embrace ϕ but only report C's assertion of ϕ and express agreement with it.

Let us now discuss a few simple examples that show rules of different kinds at work.

Example 3.3 (continued): The logical deduction from $(\phi^A, 0.1)$ to $(\neg\phi^A, 0.9)$ discussed in Example 3.1 can be formalised as follows:

$$\frac{;\Gamma^A; \vdash (\phi^A, 0.1)}{;\Gamma^A; \vdash (\neg\phi^A, 1-0.1)}$$

Example 3.4 (Logical rules): Consider an agent A and suppose that

$$\Gamma^A = (b^A, 0.3), (c^A, 0.7), (d^A, 0.5)$$

is the context containing her innate opinions. The literature $\mathfrak{L}_0$, for the moment, does not record any exchange of information, and it will be updated later by adding records of exchanges when the agents will start to read and write opinions.

Consider, for instance, the formula $((\neg b^A \wedge c^A)^A \vee d^A)^A$. It can be derived with conviction value that can be approximated to 0.717 from the set of innate opinions of agent A. A derivation that formally proves it is as follows:

$$\dfrac{\dfrac{\dfrac{\mathfrak{L}_0; \Gamma^A; \Delta \vdash (b^A, 0.3)}{\mathfrak{L}_0; \Gamma^A; \Delta \vdash ((\neg b^A)^A, 1-0.3)} \quad \mathfrak{L}_0; \Gamma^A; \Delta \vdash (c^A, 0.7)}{\mathfrak{L}_0; \Gamma^A; \Delta \vdash (((\neg b^A)^A \wedge c^A)^A, \frac{0.7\cdot 0.7}{0.7+0.7})} \quad \mathfrak{L}_0; \Gamma^A; \Delta \vdash (d^A, 0.5)}{\mathfrak{L}_0; \Gamma^A; \Delta \vdash (((\neg b^A \wedge c^A)^A \vee d^A)^A, 1 - \frac{(1-\frac{0.7\cdot 0.7}{0.7+0.7})\cdot(1-0.5)}{1-\frac{0.7\cdot 0.7}{0.7+0.7}+1-0.5})}$$

where $\mathfrak{L}_0 = \Delta = \emptyset$.

Example 3.5 (Operational rule `W`): Suppose that we wish to model the event consisting in A writing an opinion, for $\Gamma^A = (b^A, 0.3), (c^A, 0.7), (d^A, 0.5)$. In order to derive a sequent indicating this, we simply need to be able to derive from A's context $\mathfrak{L}_1; \Gamma^A; \Delta$

the expression that A writes as shown in Example 3.4. Then, we can apply the $\mathtt{W}$ rule as follows:

$$\frac{\begin{array}{c}\vdots\\ \mathfrak{L}_0;\Gamma^A;\Delta\vdash(((\neg b^A\wedge c^A)^A\vee d^A)^A,0.717)\quad 0.717>0.5\end{array}}{\mathfrak{L}_1;\Gamma^A;\Delta\vdash\mathtt{W}(((\neg b^A\wedge c^A)^A\vee d^A)^A,0.717)}\ \mathtt{W}$$

where $\Delta=\emptyset$ and $\mathfrak{L}_1=\mathfrak{L}_0^{++w(A,(((\neg b^A\wedge c^A)^A\vee d^A)^A,0.717))}=w(A,(((\neg b^A\wedge c^A)^A\vee d^A)^A,$ $0.717))$ since $\mathfrak{L}_0=\emptyset$.

Notice that, for the sake of simplicity, an agent can only write expressions with value greater than 0.5. Writing a formula ϕ with value $p\leq 0.5$ is the same as writing $\neg\phi$ with value $1-p$.

Now we know that A actually wrote the opinion $((\neg b^A\wedge c^A)^A\vee d^A)^A$ with a conviction value of 0.717. Hence, we also update her literature, that is, the record of exchanges of information by adding to the old literature $\mathfrak{L}_0$ the item $w(A,(((\neg b^A\wedge c^A)^A\vee d^A)^A,0.717))$: $\mathfrak{L}_0^{++w(A,(((\neg b^A\wedge c^A)^A\vee d^A)^A,0.717))}$.

Example 3.6 (Operational rule $\mathtt{R}$): Suppose now that agent B wishes to read the opinion written by A in Example 3.5. We can derive a sequent indicating it as follows:

$$\frac{\mathfrak{L}_1;\Gamma^A,\Delta\vdash\mathtt{W}(((\neg b^A\wedge c^A)^A\vee d^A)^A,0.717)\quad \mathfrak{L}_1;\Gamma^B;\Theta::\mathit{Profile}}{\mathfrak{L}_1^{++r(B,((((\neg b^A\wedge c^A)^A\vee d^A)^A,0.717)))};\mathfrak{L}_1;\Gamma^B;\Theta\vdash\mathtt{R}((((\neg b^A\wedge c^A)^A\vee d^A)^A,0.717),\frac{1}{1})}\ \mathtt{R}$$

where $\mathfrak{L}_1=w(B,(((\neg b^A\wedge c^A)^A\vee d^A)^A,0.717))$ and $\mathfrak{L}_1^{++r(A,((((\neg b^A\wedge c^A)^A\vee d^A)^A,0.717)))}=$ $w(A,(((\neg b^A\wedge c^A)^A\vee d^A)^A,0.717)),r(A,((((\neg b^A\wedge c^A)^A\vee d^A)^A,0.717)))$.

Here, the opinion being read has been written once, and it is the only opinion written by any agent. Hence, $\frac{1}{1}$ indicates how frequently B has been exposed to the opinion $((((\neg b^A\wedge c^A)^A\vee d^A)^A,0.717))$. That is, in this particular case, this opinion is the only one B has ever been exposed to. This will help determine how inclined will B be to accept the content of the opinion in case it is strongly in conflict with B's opinions.

4. Proof-theoretical properties of the system

We present now a study of the proof-theoretical properties of the calculus for the logic $\mathtt{(un)SecureND}_{grad}$.

Intuitively, Theorem 4.1 below guarantees that, if $\mathfrak{L};\Gamma^A;\Delta$ is a coherent context, then the derivability of $\mathfrak{L};\Gamma^A;\Delta\vdash(\phi^C,p)$ and $\mathfrak{L};\Gamma^A;\Delta\vdash(\phi^C,q)$ implies that $p=q$. In other words, the use of a coherent context implies the uniqueness of the value assigned to a signed formula ϕ^A. Notice that this result can be seen as guaranteeing a certain form of consistency of the logical fragment of $\mathtt{(un)SecureND}_{grad}$. Indeed, this can also be proved by a normalisation argument, i.e. a proof showing that there exists a procedure to remove all *detours* – unnecessary steps consisting of an introduction of a connective immediately followed by its elimination – occurring in any purely logical derivation, see also Prawitz (2006, Chapter II). We do not use a normalisation result in order to prove this – but prove it instead by reasoning on the arithmetical functions related to connectives – since the proof is simpler and more direct. We remark, nevertheless,

that the logical rules of the calculus admit a terminating detour elimination procedure. Theorem 4.2 below certifies, instead, that the logical fragment of the calculus can be seen as a conservative extension of classical logic. Theorem 4.3 generalises this result to the whole calculus. Obviously, these two theorems require some restrictions to the considered derivations ensuring that we are using the calculus as a reasoning system for a single agent in a classical setting.

Let us first introduce some useful notation.

Definition 4.1 (The arithmetic of connectives): The arithmetical operations $\circ_\neg, \circ_\wedge, \circ_\vee$ and $\circ_\rightarrow$ are defined as follows:

$$\circ_\neg x = 1 - x \qquad x \circ_\wedge y = \frac{x \cdot y}{x + y} \qquad x \circ_\vee y = 1 - \frac{(1-x)\cdot(1-y)}{1 - x + 1 - y}$$

$$x \circ_\rightarrow y = 1 - \frac{x \cdot (1-y)}{x + (1-y)}$$

The following theorem shows that, if the two parts Γ^A and Δ of a given context coherently assign values to signed formulae, then a unique value can be assigned to each formula for which we can assign a value. Technically, the coherence of Γ^A and Δ is guaranteed by the existence of the function f – which appears as a condition in the statement. This theorem also implies that, if all its conditions are met, then a unique value can be derived for each signed formula for which any value can be derived.

Theorem 4.1 (Uniqueness of value for logically derivable formulae): *Given a profile* $\mathfrak{L}; \Gamma^A; \Delta$, *if there exists a total function f from the set of signed formulae to* [0, 1] *such that*

- *for any* $(\psi^{C'}, p) \in \Gamma^A \cup \Delta, f(\psi^{C'}) = p$,
- *for any* $\psi^{C'}$ *and* C'', $f((\neg\psi^{C'})^{C''}) = \circ_\neg f(\psi^{C'})$, *and*
- *for any* $\psi_1^{C'}, \psi_2^{C''}, C'''$ *and* $\star \in \{\wedge, \vee \rightarrow\}, f((\psi_1^{C'} \star \psi_2^{C''})^{C'''}) = f(\psi_1^{C'}) \circ_\star f(\psi_2^{C''})$;

then, if $\mathfrak{L}; \Gamma^A; \Delta \vdash (\phi^C, q)$ *is derivable by only using logical rules,* $f(\phi^C) = q$.

Proof: The proof is by induction on the number of rule applications in the derivation of $\mathfrak{L}; \Gamma^A; \Delta \vdash (\phi^C, q)$.

In the base case, (ϕ^C, q) is either an element of Γ^A or an element of Δ. The statement holds by the first condition on f.

Suppose now that the statement holds for all derivations containing less than n rule applications. We prove that it holds also for any derivation containing n rule applications. We consider a generic derivation containing n rule applications and we reason by cases on the last rule applied in it.

Suppose the rule application is an introduction rule for $\neg$. By inductive hypothesis, the function f and the pair occurring as premiss of the rule application agree on the value of the formula ψ^C occurring in the pair. Since the value of f on the complex formula $(\neg(\psi^C))^{C_1}$ at the conclusion of the rule is $1 - f(\psi^C)$, since $f(\psi^C)$ is the value p that the premiss of the rule assigns to ψ^C and since the $\neg$ introduction rule assigns the value $1-p$ to the formula at its conclusion, we have that the statement holds.

Suppose the rule is an introduction rule for a connective $\star \in \{\wedge, \vee, \rightarrow\}$. By inductive hypothesis, the function f and the pairs occurring as premisses of the rule application agree on the values of the formulae $\psi_1^{C_1}$ and $\psi_2^{C_2}$ in these pairs. Since the value of f on the complex formula $(\psi_1^{C_1} \star \psi_2^{C_2})^{C_3}$ at the conclusion of the rule is computed by $\circ_\star$ on the formulae $\psi_1^{C_1}$ and $\psi_2^{C_2}$ at the premisses of the rule, and since the introduction rule employs the same arithmetical operation, we have that the statement holds.

Suppose the rule application is an elimination of $\neg$ with a premiss of the form $\ldots \vdash ((\neg(\psi^C))^{C_1}, p)$ and with conclusion of the form $\ldots \vdash (\psi^C, 1-p)$. By inductive hypothesis, $f((\neg(\psi^C))^{C_1}) = p$. By definition of f, we have that $f((\neg(\psi^C))^{C_1}) = 1 - f(\psi^C)$. But since $f((\neg(\psi^C))^{C_1}) = p$, we have that $p = 1 - f(\psi^C)$. From $p = 1 - f(\psi^C)$ it follows that $f(\psi^C) = 1 - p$. Since, moreover, the conclusion of the rule is of the form $\ldots \vdash (\psi^C, 1-p)$, we have that the statement holds for the considered derivation too.

Suppose the rule application is an elimination of $\wedge$ with premisses of the form $\ldots \vdash ((\psi_1^{C_1} \wedge \psi_2^{C_2})^C, p)$ and $\ldots \vdash (\psi_2^{C_2}, p_2)$, and with conclusion of the form $\ldots \vdash (\psi_1^{C_1}, \frac{p_2 \cdot p}{p_2 - p})$ where $p_2 - p \neq 0$. Moreover, by the condition on the rule, $p_2 - p \neq 0$. By inductive hypothesis, $f((\psi_1^{C_1} \wedge \psi_2^{C_2})^C) = p$ and $f(\psi_2^{C_2}) = p_2$. By definition of f, we have that $f((\psi_1^{C_1} \wedge \psi_2^{C_2})^C) = \frac{f(\psi_1^{C_1}) \cdot f(\psi_2^{C_2})}{f(\psi_1^{C_1}) + f(\psi_2^{C_2})}$. But since $f(\psi_2^{C_2}) = p_2$ and $f((\psi_1^{C_1} \wedge \psi_2^{C_2})^C) = p$, we have that

$$\begin{aligned}
\frac{f(\psi_1^{C_1}) \cdot p_2}{f(\psi_1^{C_1}) + p_2} &= p \\
f(\psi_1^{C_1}) \cdot p_2 &= p \cdot (f(\psi_1^{C_1}) + p_2) \\
f(\psi_1^{C_1}) \cdot p_2 &= (p \cdot f(\psi_1^{C_1})) + (p \cdot p_2) \\
(f(\psi_1^{C_1}) \cdot p_2) - (p \cdot f(\psi_1^{C_1})) &= p \cdot p_2 \\
f(\psi_1^{C_1}) \cdot (p_2 - p) &= p \cdot p_2 \\
f(\psi_1^{C_1}) &= \frac{p \cdot p_2}{p_2 - p}
\end{aligned}$$

Since, moreover, the conclusion of the rule is of the form $\ldots \vdash (\psi_1^{C_1}, \frac{p \cdot p_2}{p_2 - p})$, we have that the statement holds for the considered derivation too.

Suppose the rule application is an elimination of $\vee$ with premisses of the form $\ldots \vdash ((\psi_1^{C_1} \vee \psi_2^{C_2})^C, p)$ and $\ldots \vdash (\psi_2^{C_2}, p_2)$, and with conclusion of the form $\ldots \vdash (\psi_1^{C_1}, 1 - \frac{(1-p_2)\cdot(1-p)}{(1-p_2)-(1-p)})$ where $(1-p_2) - (1-p) \neq 0$. By inductive hypothesis, $f((\psi_1^{C_1} \vee \psi_2^{C_2})^C) = p$ and $f(\psi_2^{C_2}) = p_2$. By definition of f, we have that $f((\psi_1^{C_1} \vee \psi_2^{C_2})^C) = 1 - \frac{(1-f(\psi_1^{C_1}))\cdot(1-f(\psi_2^{C_2}))}{1-f(\psi_1^{C_1})+1-f(\psi_2^{C_2})}$. But since $f(\psi_2^{C_2}) = p_2$ and $f((\psi_1^{C_1} \vee \psi_2^{C_2})^C) = p$, we have that

$$1 - \frac{(1 - f(\psi_1^{C_1})) \cdot (1 - p_2)}{1 - f(\psi_1^{C_1}) + 1 - p_2} = p$$

$$
\begin{aligned}
1-p &= \frac{(1-f(\psi_1^{C_1}))\cdot(1-p_2)}{1-f(\psi_1^{C_1})+1-p_2} \\
(1-p)\cdot(1-f(\psi_1^{C_1})+1-p_2) &= (1-f(\psi_1^{C_1}))\cdot(1-p_2) \\
(1-p)\cdot(1-f(\psi_1^{C_1}))+(1-p)\cdot(1-p_2) &= (1-f(\psi_1^{C_1}))\cdot(1-p_2) \\
(1-p)\cdot(1-p_2) &= (1-f(\psi_1^{C_1}))\cdot(1-p_2) \\
&\quad -(1-p)\cdot(1-f(\psi_1^{C_1})) \\
(1-p)\cdot(1-p_2) &= (1-f(\psi_1^{C_1}))\cdot((1-p_2)-(1-p)) \\
\frac{(1-p)\cdot(1-p_2)}{(1-p_2)-(1-p)} &= 1-f(\psi_1^{C_1}) \\
f(\psi_1^{C_1}) &= 1-\frac{(1-p)\cdot(1-p_2)}{(1-p_2)-(1-p)}
\end{aligned}
$$

Since, moreover, the conclusion of the rule is of the form $\ldots \vdash (\psi_1^{C_1}, 1-\frac{(1-p_2)\cdot(1-p)}{(1-p_2)-(1-p)})$, the statement holds for the considered derivation too.

Suppose the rule application is an elimination of $\rightarrow$ with premisses of the form $\ldots \vdash ((\psi_1^{C_1} \rightarrow \psi_2^{C_2})^C, p)$ and $\ldots \vdash (\psi_1^{C_1}, p_1)$, and with conclusion of the form $\ldots \vdash (\psi_2^{C_2}, 1-\frac{(1-p)\cdot p_1}{p_1-(1-p)})$ where $p_1-(1-p) \neq 0$. By inductive hypothesis, $f((\psi_1^{C_1} \rightarrow \psi_2^{C_2})^C) = p$ and $f(\psi_1^{C_1}) = p_1$. By definition of f, we have that $f((\psi_1^{C_1} \rightarrow \psi_2^{C_2})^C) = 1-\frac{f(\psi_1^{C_1})\cdot(1-f(\psi_2^{C_2}))}{f(\psi_1^{C_1})+1-f(\psi_2^{C_2})}$. But since $f(\psi_1^{C_1}) = p_1$ and $f((\psi_1^{C_1} \rightarrow \psi_2^{C_2})^C) = p$, we have that

$$
\begin{aligned}
1-\frac{p_1\cdot(1-f(\psi_2^{C_2}))}{p_1+1-f(\psi_2^{C_2})} &= p \\
1-p &= \frac{p_1\cdot(1-f(\psi_2^{C_2}))}{p_1+1-f(\psi_2^{C_2})} \\
(1-p)\cdot(p_1+1-f(\psi_2^{C_2})) &= p_1\cdot(1-f(\psi_2^{C_2})) \\
(1-p)\cdot p_1+(1-p)\cdot(1-f(\psi_2^{C_2})) &= p_1\cdot(1-f(\psi_2^{C_2})) \\
(1-p)\cdot p_1 &= p_1\cdot(1-f(\psi_2^{C_2}))-(1-p)\cdot(1-f(\psi_2^{C_2})) \\
(1-p)\cdot p_1 &= (1-f(\psi_2^{C_2}))\cdot(p_1-(1-p)) \\
\frac{(1-p)}{p_1-(1-p)}\cdot p_1 &= 1-f(\psi_2^{C_2}) \\
f(\psi_2^{C_2}) &= 1-\frac{(1-p)}{p_1-(1-p)}\cdot p_1 \\
f(\psi_2^{C_2}) &= 1-\frac{(1-p)\cdot p_1}{p_1-(1-p)}
\end{aligned}
$$

Since, moreover, the conclusion of the rule is of the form $\ldots \vdash (\psi_2^{C_2}, 1-\frac{(1-p)\cdot p_1}{p_1-(1-p)})$, we have that the statement holds for the considered derivation too. ■

We remark that the first condition on f also implies that Γ^A and Δ agree on all values that they directly assign to formulae. Moreover, the second and third conditions on f states that all agents ('for any C''...') agree on the value of complex formulae constructed by combining the same formulae. Notice, finally, that a derivation δ that does not use part of a context can be reformulated as a derivation δ' which is structurally equivalent (same rules applied in the same order to the same pairs) but with a context that does not contain some of the parts, or all of the parts, which were not used in δ. Hence, the statement also applies to derivations that only use coherent parts of possibly incoherent profiles.

Next theorem guarantees that, if we only consider one agent and we restrict the conviction values to 0 and 1, the logical rules of our calculus behave in a sound way with respect to the connectives of classical logic. In other words, it tells us that our system can be seen, in a sense, as a conservative extension of classical logic. To prove this, we first provide a translation of formulae of our logic $\mathcal{E} = (\phi^C, q)$ where $q \in \{0, 1\}$ to formulae of the language of classical logic.

Definition 4.2 (Classical translation $\tau_{\mathsf{CL}}(\,)$ of an expression): For any signed formula ϕ^C, we denote by $(\phi^C)^\emptyset$ the formula in the language of classical logic obtained from ϕ^C by erasing all agent labels.

For any expression $\mathcal{E} = (\phi^C, q)$ where $q \in \{0, 1\}$, its translation $\tau_{\mathsf{CL}}((\phi^C, q))$ in the language of classical logic is defined as follows:

- if $q = 1$, then $\tau_{\mathsf{CL}}((\phi^C, q)) = (\phi^C)^\emptyset$,
- if $q = 0$, then $\tau_{\mathsf{CL}}((\phi^C, q)) = \neg(\phi^C)^\emptyset$.

Theorem 4.2 (Classical soundness): *If all of the following points hold:*

- *a sequent of the form $\mathfrak{L}; \Gamma^A; \Delta \vdash (\phi^C, p)$ is derivable from a set of start rule conclusions of the form*

$$\mathfrak{L}; \Gamma^A; \Delta \vdash (\psi_1^C, p_1) \quad \cdots \quad \mathfrak{L}; \Gamma^A; \Delta \vdash (\psi_n^C, p_n)$$

  by only using logical rules,
- *all numerical values occurring in the derivation are either 0 or 1, and*
- *all agent symbols occurring in $\phi^C, \psi_1^C, \ldots, \psi_n^C$ are equal to S*

then $\tau_{\mathsf{CL}}((\phi^C, p))$ classically follows from $\tau_{\mathsf{CL}}((\psi_1^C, p_1)), \ldots, \tau_{\mathsf{CL}}((\psi_n^C, p_n))$.

Proof: We show how to derive $\tau_{\mathsf{CL}}((\phi^C, p))$ from the hypotheses

$$\tau_{\mathsf{CL}}((\psi_1^C, p_1)), \ldots, \tau_{\mathsf{CL}}((\psi_n^C, p_n))$$

in a standard natural deduction calculus for classical logic, see, for instance, Prawitz (1971). We proceed by induction on the number of rules applied below the start rule conclusions in the derivation δ of $\mathfrak{L}; \Gamma^A; \Delta \vdash (\phi^C, p)$.

If no rules are applied below the start rule conclusions in δ, then

$$\mathfrak{L}; \Gamma^A; \Delta \vdash (\phi^C, p) \in \{\mathfrak{L}; \Gamma^A; \Delta \vdash (\psi_1^C, p_1) \quad \cdots \quad \mathfrak{L}; \Gamma^A; \Delta \vdash (\psi_n^C, p_n)\}$$

and thus also

$$\tau_{\mathsf{CL}}((\phi^C, p)) \in \{\tau_{\mathsf{CL}}((\psi_1^C, p_1)), \ldots, \tau_{\mathsf{CL}}((\psi_n^C, p_n))\}$$

and the classical derivation translating δ is the trivial one with hypothesis $\tau_{\mathsf{CL}}((\phi^C, p))$ and conclusion $\tau_{\mathsf{CL}}((\phi^C, p))$.

Suppose now that the statement holds for all derivations δ' that contain less than $n \neq 0$ rule applications below the start rule conclusions, we show that the statement holds also for any derivation δ that contains n rule applications below the start rule conclusions.

Consider the last rule applied in δ and call it R. According to the statement, R must be a logical rule.

- Suppose $R = \neg I$. By inductive hypothesis, the derivation of the premiss $\ldots \vdash (\xi, r)$ of $\neg I$ can be translated into a classical natural deduction derivation that complies with the requirements of the statement. According to the statement, r is either 0 or 1. If $r = 0$, then our premiss has the form $\ldots \vdash (\xi, 0)$ and the classical translation of the derivation of $\ldots \vdash (\xi, 0)$ has conclusion $\tau_{\mathsf{CL}}((\xi, 0)) = \neg\xi'$ for some ξ'. Since an application of $\neg I$ to a sequent of the form $\ldots \vdash (\xi, 0)$ has the form $\ldots \vdash (\neg\xi, 1)$, we already have our classical derivation of $\neg\xi' = \tau_{\mathsf{CL}}((\neg\xi, 1))$ that translates δ.

  If, on the other hand, $r = 1$, then our premiss is of the form $\ldots \vdash (\xi, 1)$ and the classical translation of the derivation of $\ldots \vdash (\xi, 1)$ has conclusion $\tau_{\mathsf{CL}}((\xi, 1)) = \xi'$ for some ξ'. Since an application of $\neg I$ to a sequent of the form $\ldots \vdash (\xi, 1)$ has the form $\ldots \vdash (\neg\xi, 0)$, in order to translate δ, we need to classically derive $\tau_{\mathsf{CL}}((\neg\xi, 0)) = \neg\neg\xi'$ from ξ' without using any additional hypothesis. But this can be done by the following derivation:

$$\dfrac{\dfrac{[\neg\xi']^1 \quad \begin{array}{c}\vdots\\ \xi'\end{array}}{\bot}}{\neg\neg\xi'}\,1$$

- Suppose $R = \neg E$. By inductive hypothesis, the derivation of the premiss $\ldots \vdash (\neg\xi, r)$ of $\neg E$ can be translated into a classical natural deduction derivation that complies with the requirements of the statement. According to the statement, r is either 0 or 1. If $r = 0$, then our premiss has the form $\ldots \vdash (\neg\xi, 0)$ and the classical translation of the derivation of $\ldots \vdash (\neg\xi, 0)$ has conclusion $\tau_{\mathsf{CL}}((\neg\xi, 0)) = \neg\neg\xi'$ for some ξ'. Since an application of $\neg E$ to a sequent of the form $\ldots \vdash (\neg\xi, 0)$ has the form $\ldots \vdash (\xi, 1)$, in order to translate δ, we need to classically derive $\tau_{\mathsf{CL}}((\xi, 1)) = \xi'$ from $\neg\neg\xi'$ without using any additional hypothesis. But this can

be done by the following derivation:

$$\dfrac{\dfrac{[\neg\xi']^1 \quad \begin{array}{c}\vdots\\ \neg\neg\xi'\end{array}}{\bot}}{\xi'}\,1$$

where the last inference is an application of the classical absurdity rule that discharges the hypothesis $\neg\xi'$.

If, on the other hand, $r = 1$, then our premiss is of the form $\ldots \vdash (\neg\xi, 1)$ and the classical translation of the derivation of $\ldots \vdash (\neg\xi, 1)$ has conclusion $\tau_{\mathsf{CL}}((\neg\xi, 1)) = \neg\xi'$ for some ξ'. Since an application of $\neg E$ to a sequent of the form $\ldots \vdash (\neg\xi, 1)$ has conclusion $\ldots \vdash (\xi, 0)$, we already have our classical derivation of $\neg\xi' = \tau_{\mathsf{CL}}((\xi, 0))$ that translates δ.

- Suppose $R = \wedge I$. By inductive hypothesis, the derivations of each premiss $\ldots \vdash (\xi_1, r_1)$ and $\ldots \vdash (\xi_2, r_2)$ of $\wedge I$ can be translated into a classical natural deduction derivation that complies with the requirements of the statement. According to the statement, each value r_1, r_2 can either be 0 or 1. If $r_1 = r_2 = 1$, then the classical translation of the derivations of the premisses $\ldots \vdash (\xi_1, r_1)$ and $\ldots \vdash (\xi_2, r_2)$ have conclusions $\tau_{\mathsf{CL}}((\xi_1, 1)) = \xi_1'$ and $\tau_{\mathsf{CL}}((\xi_2, 1)) = \xi_2'$ respectively. The translation $\tau_{\mathsf{CL}}((\xi_1 \wedge \xi_2, 1)) = \xi_1' \wedge \xi_2'$ of the conclusion of δ can then be derived as follows:

$$\dfrac{\begin{array}{c}\vdots\\ \xi_1'\end{array} \quad \begin{array}{c}\vdots\\ \xi_2'\end{array}}{\xi_1' \wedge \xi_2'}$$

Suppose, on the other hand, that either $r_1 = 0$ or $r_2 = 0$. Without loss of generality, we say that $r_1 = 0$, the other cases being analogous. Then, by inductive hypothesis, there exists a classical derivation of $\tau_{\mathsf{CL}}((\xi_1, 0)) = \neg\xi_1'$. If $r_1 = 0$, moreover, the conclusion of δ is of the form $\ldots \vdash (\xi_1 \wedge \xi_2, 0)$. Therefore, in order to translate δ, we need to derive $\tau_{\mathsf{CL}}((\xi_1 \wedge \xi_2, 0)) = \neg(\xi_1' \wedge \xi_2')$ from $\neg\xi_1'$ with no additional hypotheses. But this can be done as follows:

$$\dfrac{\dfrac{\begin{array}{c}\vdots\\ \neg\xi_1'\end{array} \quad \dfrac{[\xi_1' \wedge \xi_2']^1}{\xi_1'}}{\bot}}{\neg(\xi_1' \wedge \xi_2')}\,1$$

- Suppose $R = \wedge E$. Without loss of generality, we consider the case in which the premisses of $\wedge E$ are $\ldots \vdash (\xi_1 \wedge \xi_2, r_1)$ and $\ldots \vdash (\xi_2, r_2)$, the other case being analogous. By inductive hypothesis, the derivations of the premisses of $\wedge E$ can be translated into classical natural deduction derivations that comply with the requirements of the statement. According to the statement, each value r_1, r_2 can either be 0 or 1, but, according to the restrictions on the rule $\wedge E$, only r_1 can be 0. Suppose then that $r_1 = r_2 = 1$. By inductive hypothesis, then, there exists a classical derivation of $\tau_{\mathsf{CL}}((\xi_1 \wedge \xi_2, 1)) = \xi_1' \wedge \xi_2'$. Moreover, the conclusion of δ

is of the form $\ldots \vdash (\xi_1, 1)$. Hence, in order to translate δ, it is enough to derive $\tau_{\mathsf{CL}}((\xi_1, 1)) = \xi_1'$ from the hypothesis $\xi_1' \wedge \xi_2'$ alone as follows:

$$\dfrac{\begin{array}{c}\vdots\\ \xi_1' \wedge \xi_2'\end{array}}{\xi_1'}$$

Suppose, on the other hand, that $r_1 = 0$. Recall that r_2 must be, in any case, 1 due to the restrictions on $\wedge E$. By inductive hypothesis, then, there exists a classical derivation of $\tau_{\mathsf{CL}}((\xi_1 \wedge \xi_2, 0)) = \neg(\xi_1' \wedge \xi_2')$ and a classical derivation of $\tau_{\mathsf{CL}}((\xi_2, 1)) = \xi_2'$. The conclusion of δ, moreover, is of the form $\ldots \vdash (\xi_1, 0)$. Hence, in order to translate δ, we need to derive $\tau_{\mathsf{CL}}((\xi_1, 0)) = \neg\xi_1'$ from the hypotheses $\neg(\xi_1' \wedge \xi_2')$ and ξ_2'. This can be done as follows:

$$\dfrac{\dfrac{\begin{array}{c}\vdots\\ \neg(\xi_1' \wedge \xi_2')\end{array} \quad \dfrac{[\xi_1']^1 \quad \begin{array}{c}\vdots\\ \xi_2'\end{array}}{\xi_1' \wedge \xi_2'}}{\bot}}{\neg\xi_1'}\,1$$

- Suppose $R = \vee I$. By inductive hypothesis, the derivations of each premiss $\ldots \vdash (\xi_1, r_1)$ and $\ldots \vdash (\xi_2, r_2)$ of $\vee I$ can be translated into a classical natural deduction derivation that complies with the requirements of the statement. According to the statement, each value r_1, r_2 can either be 0 or 1. Suppose that at least one of the values r_1, r_2 is 1. Without loss of generality we say that $r_1 = 1$, the other cases being analogous. Then the classical translation of the derivation of the premiss $\ldots \vdash (\xi_1, r_1)$ has conclusion $\tau_{\mathsf{CL}}((\xi_1, 1)) = \xi_1'$. The translation $\tau_{\mathsf{CL}}((\xi_1 \vee \xi_2, 1)) = \xi_1' \vee \xi_2'$ of the conclusion of δ can thus be derived by the following derivation:

$$\dfrac{\begin{array}{c}\vdots\\ \xi_1'\end{array}}{\xi_1' \vee \xi_2'}$$

Suppose, on the other hand, that $r_1 = r_2 = 0$. Then, by inductive hypothesis, there are classical derivations of $\tau_{\mathsf{CL}}((\xi_1, 0)) = \neg\xi_1'$ and $\tau_{\mathsf{CL}}((\xi_2, 0)) = \neg\xi_2'$. If $r_1 = r_2 = 0$, moreover, the conclusion of δ is of the form $\ldots \vdash (\xi_1 \vee \xi_2, 0)$. Therefore, in order to translate δ, we need to derive $\tau_{\mathsf{CL}}((\xi_1 \vee \xi_2, 0)) = \neg(\xi_1' \vee \xi_2')$ from $\neg\xi_1'$ and $\neg\xi_2'$ without any additional hypothesis. But this can be done as follows:

$$\dfrac{\dfrac{[\xi_1' \vee \xi_2']^1 \quad \dfrac{\begin{array}{c}\vdots\\ \neg\xi_1'\end{array} \quad [\xi_1']^2}{\bot} \quad \dfrac{\begin{array}{c}\vdots\\ \neg\xi_2'\end{array} \quad [\xi_2']^2}{\bot}}{\bot}\,2}{\neg(\xi_1' \vee \xi_2')}\,1$$

- Suppose $R = \vee E$. Without loss of generality, we consider the case in which the premisses of $\wedge E$ are $\ldots \vdash (\xi_1 \vee \xi_2, r_1)$ and $\ldots \vdash (\xi_2, r_2)$, the other case being analogous. According to the statement, each value r_1, r_2 can either be 0 or 1, but, according to the restrictions on the rule $\vee E$, only r_1 can be 1. Suppose then that $r_1 = 0$. By inductive hypothesis, then, there exists a classical derivation of $\tau_{\mathrm{CL}}((\xi_1 \vee \xi_2, 0)) = \neg(\xi_1' \vee \xi_2')$. Moreover, the conclusion of δ is of the form $\ldots \vdash (\xi_1, 0)$. Hence, in order to translate δ, it is enough to derive $\tau_{\mathrm{CL}}((\xi_1, 0)) = \neg\xi_1'$ from the hypothesis $\neg(\xi_1' \vee \xi_2')$ alone as follows:

$$\dfrac{\dfrac{\begin{array}{c}\vdots\\ \neg(\xi_1' \vee \xi_2')\end{array} \quad \dfrac{[\xi_1']^1}{\xi_1' \vee \xi_2'}}{\bot}}{\neg\xi_1'}\,1$$

Suppose, on the other hand, that $r_1 = 1$. Recall that r_2 must be, in any case, 0 due to the restrictions on $\vee E$. By inductive hypothesis, then, there exists a classical derivation of $\tau_{\mathrm{CL}}((\xi_1 \vee \xi_2, 1)) = \xi_1' \vee \xi_2'$ and a classical derivation of $\tau_{\mathrm{CL}}((\xi_2, 0)) = \neg\xi_2'$. The conclusion of δ, moreover, is of the form $\ldots \vdash (\xi_1, 1)$. Hence, in order to translate δ, we need to derive $\tau_{\mathrm{CL}}((\xi_1, 1)) = \xi_1'$ from the hypotheses $\xi_1' \vee \xi_2'$ and $\neg\xi_2'$. This can be done as follows:

$$\dfrac{\begin{array}{c}\vdots\\ \xi_1' \vee \xi_2'\end{array} \quad [\xi_1']^1 \quad \dfrac{\begin{array}{c}\vdots\\ \neg\xi_2'\end{array} \quad [\xi_2']^1}{\bot}}{\xi_1'}\,1$$

where the inference that enables us to derive ξ_1' from $\bot$ is an application of the *ex falso quodlibet* rule.

- Suppose $R = \rightarrow I$. By inductive hypothesis, the derivations of the premisses $\ldots \vdash (\xi_1, r_1)$ and $\ldots \vdash (\xi_2, r_2)$ of $\rightarrow I$ can be translated into classical natural deduction derivations that comply with the requirements of the statement. According to the statement, each value r_1, r_2 can either be 0 or 1, but, according to the restrictions on the rule $\rightarrow I$, only r_2 can be 0. Suppose then that $r_2 = 1$. By inductive hypothesis, then, there exists a classical derivation of $\tau_{\mathrm{CL}}((\xi_2, 1)) = \xi_2'$. Moreover, the conclusion of δ is of the form $\ldots \vdash (\xi_1 \rightarrow \xi_2, 1)$. Hence, in order to translate δ, it is enough to derive $\tau_{\mathrm{CL}}((\xi_1 \rightarrow \xi_2, 1)) = \xi_1' \rightarrow \xi_2'$ from the hypothesis ξ_1' alone as follows:

$$\dfrac{\begin{array}{c}\vdots\\ \xi_2'\end{array}}{\xi_1' \rightarrow \xi_2'}$$

Suppose, on the other hand, that $r_2 = 0$. Recall that r_1 must be, in any case, 1 due to the restrictions on $\rightarrow I$. By inductive hypothesis, then, there exists a classical derivation of $\tau_{\mathrm{CL}}((\xi_2, 0)) = \neg\xi_2'$ and a classical derivation of $\tau_{\mathrm{CL}}((\xi_1, 1)) = \xi_1'$. The conclusion of δ, moreover, is of the form $\ldots \vdash (\xi_1 \rightarrow \xi_2, 0)$. Hence, in

order to translate δ, we need to derive $\tau_{\mathsf{CL}}((\xi_1 \to \xi_2, 0)) = \neg(\xi_1' \to \xi_2')$ from the hypotheses $\neg(\xi_2')$ and ξ_1'. This can be done as follows:

$$\dfrac{\dfrac{\begin{array}{c}\vdots\\ \neg\xi_2'\end{array} \qquad \dfrac{[\xi_1' \to \xi_2']^1 \quad \begin{array}{c}\vdots\\ \xi_1'\end{array}}{\xi_2'}}{\bot}}{\neg(\xi_1' \to \xi_2')}\,1$$

- Suppose $R = \to E$. By inductive hypothesis, the derivations of the premisses $\ldots \vdash (\xi_1 \to \xi_2, r_1)$ and $\ldots \vdash (x_1, r_2)$ of $\to E$ can be translated into classical natural deduction derivations that comply with the requirements of the statement. According to the statement, each value r_1, r_2 can either be 0 or 1, but, according to the restrictions on $\to E$, none of the two values can be 0. We necessarily have, then, that $r_1 = r_2 = 1$. By inductive hypothesis, there are classical derivations of $\tau_{\mathsf{CL}}((\xi_1 \to \xi_2, 1)) = \xi_1' \to \xi_2'$ and $\tau_{\mathsf{CL}}((\xi_1, 1)) = \xi_1'$. Moreover, the conclusion of δ is of the form $\ldots \vdash (\xi_2, 1)$. Hence, in order to translate δ, it is enough to derive $\tau_{\mathsf{CL}}((\xi_2, 1)) = \xi_2'$ from the hypotheses $\xi_1' \to \xi_2'$ and ξ_1' as follows:

$$\dfrac{\begin{array}{c}\vdots\\ \xi_1' \to \xi_2'\end{array} \quad \begin{array}{c}\vdots\\ \xi_1'\end{array}}{\xi_2'}$$

■

The last theorem of this section generalises Theorem 4.1 to the case in which not only logical rules are used. Here it is essential that, when we restrict the conviction values to 0 and 1, inconsistency values collapse to 0 and 1 too, indicating respectively full consistency and inconsistency.

Theorem 4.3: *If $\varepsilon_\perp^A \neq 1$, given any context $\mathfrak{L}; \Gamma^A; \Delta$ such that $\tau_{\mathsf{CL}}(\Gamma^A \cup \Delta)$ is classically consistent, if two sequents $\mathfrak{L}; \Gamma^A; \Delta \vdash (\phi^C, p)$ and $\mathfrak{L}; \Gamma^A; \Delta \vdash (\phi^C, q)$ are derivable by derivations only containing the numerical values 0 and 1 and in which only one agent name C occurs to the right of $\vdash$, then $p = q$.*

Proof: If the context $\mathfrak{L}; \Gamma^A; \Delta$ only contains the numerical values 0 and 1 and $r \in \{0, 1\}$, then, for any expression (ψ^C, r), the inconsistency value $val^{\perp\!\!\!\perp}(\Gamma^A; \Delta, (\psi^C, r))$ can only be 0 or 1, see Definition 3.4. In particular, $val^{\perp\!\!\!\perp}(\Gamma^A; \Delta, (\psi^C, r)) = 0$ if, and only if, the set $\Gamma^A \cup \Delta \cup \{(\psi^C, r)\}$ is classically consistent. By applying the `ATr` rule, then, we can add a pair (ψ^C, r) where $r \in \{0, 1\}$ to the profile $\mathfrak{L}; \Gamma^A; \Delta$ only in case one of the following holds:

- $r = 1$ and ψ is classically consistent with the set of formulae $\{\xi \mid (\xi, 1) \in \Gamma^A \cup \Delta\} \cup \{\neg\xi \mid (\xi, 0) \in \Gamma^A \cup \Delta\}$
- $r = 0$ and $\neg\psi$ is classically consistent with the set of formulae $\{\xi \mid (\xi, 1) \in \Gamma^A \cup \Delta\} \cup \{\neg\xi \mid (\xi, 0) \in \Gamma^A \cup \Delta\}$

where by ψ and χ we indicate, with a slight abuse of notation, formulae that do not contain agent name labels.

By the AMis rule, on the other hand, if all values we are considering are either 0 or 1, we cannot add a pair (ψ^C, r) to a context without removing all inconsistencies.

Now, by Theorem 4.2, if all our values are either 0 or 1, the existence of a purely logical derivation of $\mathfrak{L}; \Gamma^A; \Delta \vdash (\psi^C, r)$ in which only the agent name C occurs to the right of $\vdash$ implies

- if $r = 1$, the classical derivability of ψ from $\{\xi \mid (\xi^C, 1) \in \Gamma^A \cup \Delta\} \cup \{\neg\xi^C \mid (\xi, 0) \in \Gamma^A \cup \Delta\}$ and
- if $r = 0$, the classical derivability of $\neg\psi$ from $\{\xi \mid (\xi^C, 1) \in \Gamma^A \cup \Delta\} \cup \{\neg\xi \mid (\xi^C, 0) \in \Gamma^A \cup \Delta\}$

Suppose now, for the sake of contradiction, that $\mathfrak{L}; \Gamma^A; \Delta \vdash (\phi^C, p)$ and $\mathfrak{L}; \Gamma^A; \Delta \vdash (\phi^C, q)$ are derivable by derivations only containing the numerical values 0 and 1 and in which only one agent name occurs to the right of $\vdash$, and that $p \neq q$. If this is so, we either have added inconsistent formulae to our context by acquire rules, which is impossible, or we have that two contradictory formulae are derivable in classical logic from a consistent context, which is impossible as well. Therefore, the statement holds. ■

5. The semantics of (un) SecureND$_{grad}$

We now introduce the relational semantics of (un) SecureND$_{grad}$, closely representing a situation in which agents – each associated to a set of opinions – interact by exchanging opinions and change their own sets of opinions depending on the occurring interactions.

Let us begin by defining what a model is.

Definition 5.1 (Model): A model $\mathcal{M}$ is a tuple $(\{K^i_j\}_{i \in T, j \in \mathbb{A}}, T, \mathbb{A}, \mathfrak{C}, \varepsilon)$ where

- T is a set of time instants (natural numbers);
- $\mathbb{A}$ is a set of agents;
- K^t_A, for any $t \in T$ and $A \in \mathbb{A}$, is the set of expressions representing, on the one hand, the opinion state of agent A at time t (K^0_A being, for instance, the set of innate opinions initially possessed by A) and, on the other hand, a record of all communicative actions that have been taken by the agent A at time t;
- $\mathfrak{C} : T \mapsto \{r(B, (\phi^C, p)), w(B, (\phi^C, p))\}$ is an action choice function that associates to each time point $t \in T$ the name of the action (read or write) that agent B chooses to execute at time t;
- ε is a triple $(\varepsilon_\perp, \varepsilon_{int}, \varepsilon_{freq})$ of threshold functions mapping agent names to real numbers between 0 and 1: for any $i \in \mathbb{A}$, $\varepsilon^i_\perp, \varepsilon^i_{int}, \varepsilon^i_{freq} \in [0, 1]$;

that complies with the following four conditions:

(1) if $\mathfrak{C}(t) = w(A, (\phi^C, p))$, then $(\phi^C, p) \in \mathrm{cl}_A(K^t_A)$ and $p > 0.5$, where $\mathrm{cl}_A(X)$ denotes the deductive closure of the set X according to the agent A as defined below in Definition 5.6;

(2) if $\mathfrak{C}(t) = r(A, (\phi^C, p))$, then there exists a $t' \in T$ and an $A' \in \mathbb{A}$ such that $t' < t$ and $\mathfrak{C}(t') = w(A', (\phi^C, p))$;
(3) K_A^0, for any $A \in \mathbb{A}$, is a set $\{(a_1^A, p_1), \ldots, (a_n^A, p_n)\}$ such that $a_1^A, \ldots, a_n^A$ are all different;
(4) $K_A^{t+1} = Up(K_A^t, \mathfrak{C}(t))$ as specified in Definition 5.5 below.

Let us now introduce a few formal definitions required to specify how states are updated during agents' interactions.

Definition 5.2 (Inconsistency value of a set of expressions): For any set K_A^t of expressions, let K_A^{t*} be the set of expressions obtained from K_A^t by replacing each element (χ, p) where $p < \frac{1}{2}$ with $(\neg\chi, 1-p)$. Then $Set^{\perp\!\!\!\perp}(K_A^t, (\phi^C, p))$ is defined as the set of all pairs $(\{\psi_1, \ldots, \psi_n, \phi\}, i)$ such that

- $n \geq 0 \leq m$;
- $\{(\psi_1^{C_1}, q_1), \ldots, (\psi_n^{C_n}, q_n)\} \subseteq K_A^{t*}$;
- $\{\psi_1, \ldots, \psi_n, \phi\}$ is a classically inconsistent set;
- all proper subsets of $\{\psi_1, \ldots, \psi_n, \phi\}$ are consistent;
- $i = q_1 \cdot \ldots \cdot q_n \cdot p$.

Let, moreover, $val^{\perp\!\!\!\perp}(K_A^t, (\phi^C, p)) = \max(\{j \mid (Z, j) \in Set^{\perp\!\!\!\perp}(K_A^t, (\phi^C, p))\})$.

Definition 5.3 (Inconsistent part of a set of expressions): For any real number ε, $K_A^t \perp\!\!\!\perp_\varepsilon ((\phi^C, p))$ is defined as the subset of K_A^t that contains all pairs in K_A^t the first element of which occurs in any set Z such that $(Z, j) \in Set^{\perp\!\!\!\perp}(K_A^t, (\phi^C, p))$ and $j > \varepsilon$.

Definition 5.4 (Opinion set of a state): For any state K_A^t, let $Op(K_A^t)$ be defined as $\{(\phi^B, p) \mid (\phi^B, p) \in K_A^t\}$.

In Definition 5.5, we present the function that determines the temporal evolution of the opinion states in the system. The function computes the opinion state K_A^{t+1} of an agent A at time $t+1$ on the basis of two arguments: the state K_A^t – which is the state related to the same agent A for the previous time t – and the value of the function $\mathfrak{C}$ applied to time t. Depending on $\mathfrak{C}(t)$, which indicates the action that happened between time t and time $t+1$, the function operates differently. As specified by clause (1) of Definition 5.5, if $\mathfrak{C}(t)$ is an action by an agent $B \neq A$, then the state K_A^{t+1} is the same as K_A^t minus the operational formulae relative to time t, which are removed inasmuch as they are outdated. As specified by clause (2), if $\mathfrak{C}(t)$ is a writing action executed by A, then K_A^{t+1} simply contains all elements of K_A^t minus the operational formulae relative to time t plus the expression $\mathbb{W}(\phi^C, p)$ indicating that A wrote the pair (ϕ^C, p). If $\mathfrak{C}(t)$ is a reading action, the result of $Up(K_A^t, \mathfrak{C}(t))$ depends on several factors determining whether:

- A trusts but does not internalise the newly received opinion (ϕ^C, p) since its conviction value p is not higher than A's threshold ε_{int} for internalisation – clause (3)(a);

- trusts but does not internalise the newly received opinion (ϕ^C, p) since A already has a stronger opinion (ϕ^A, q) on the topic – clause (3)(b);
- trusts and internalises the newly received opinion – clause (3)(c);
- distrusts the newly received opinion – clause (3)(d);
- accepts the newly received opinion but, as a consequence, has to mistrust some previously held opinions – clauses (3)(e)–(g).

The differences in (3)(e)–(g) roughly reflect those in (3)(a)–(c). Indeed, (3)(e) concerns the case in which there is no internalisation because of the threshold ε_{int}, just as (3)(a); (3)(f) concerns the case in which the internalisation happens, just as (3)(c); and, finally, (3)(g) concerns the case in which there is no internalisation because a stronger opinion on the same topic is held by A, just as (3)(b).

Definition 5.5 (Update function): The function Up is defined as follows:

(1) if $B \neq A$, $Up(K^t_A, w(B, (\phi^C, p))) = Up(K^t_A, r(B, (\phi^C, p))) = Op(K^t_A)$
(2) $Up(K^t_A, w(A, (\phi^C, p))) = Op(K^t_A) \cup \{\mathtt{W}(\phi^C, p)\}$
(3) $Up(K^t_A, r(A, (\phi^C, p)))$ is defined as follows:
  (a) $Op(K^t_A) \cup \{\mathtt{R}((\phi^C, p), \frac{m}{n}), \mathtt{Tr}(\phi^C, p), (\phi^C, p)\}$ if $val^{\perp\!\!\!\perp}(K^t_A, (\phi^C, p)) \leq \varepsilon^A_\perp$ and $p \leq \varepsilon^A_{int}$
  (b) $Op(K^t_A) \cup \{\mathtt{R}((\phi^C, p), \frac{m}{n}), \mathtt{Tr}(\phi^C, p), (\phi^C, p)\}$ if $val^{\perp\!\!\!\perp}(K^t_A, (\phi^C, p)) \leq \varepsilon^A_\perp$, $p > \varepsilon^A_{int}$ and $(\phi^A, q) \in K^t_A$ with $p \leq q$
  (c) $Op(K^t_A) \cup \{\mathtt{R}((\phi^C, p), \frac{m}{n}), \mathtt{Tr}(\phi^C, p), (\phi^C, p), (\phi^A, p)\}$ if $val^{\perp\!\!\!\perp}(K^t_A, (\phi^C, p)) \leq \varepsilon^A_\perp$, $p > \varepsilon^A_{int}$ and, for any q, $(\phi^A, q) \notin K^t_A$
  while
  $(Op(K^t_A) \setminus \{(\phi^A, q)\}) \cup \{\mathtt{R}((\phi^C, p), \frac{m}{n}), \mathtt{Tr}(\phi^C, p), (\phi^C, p), (\phi^A, p)\}$
  if $(\phi^A, q) \in K^t_A$, $val^{\perp\!\!\!\perp}(K^t_A, (\phi^C, p)) \leq \varepsilon^A_\perp$, $q < p > \varepsilon^A_{int}$ and ϕ^A contains at least one agent name different from A
  (d) $Op(K^t_A) \cup \{\mathtt{R}((\phi^C, p), \frac{m}{n}), \mathtt{DisTr}(\phi^C, p)\}$ if $val^{\perp\!\!\!\perp}(K^t_A, (\phi^C, p)) > \varepsilon^A_\perp$ and $\frac{m}{n} \leq \varepsilon^A_{freq}$
  (e) $(Op(K^t_A) \setminus (K^t_A \perp\!\!\!\perp_{\varepsilon^A_\perp} (\phi^C, p))) \cup \{\mathtt{R}((\phi^C, p), \frac{m}{n}), (\phi^C, p)\}$ if $val^{\perp\!\!\!\perp}(K^t_A, (\phi^C, p)) > \varepsilon^A_\perp$ and $\frac{m}{n} > \varepsilon^A_{freq}$ and $p \leq \varepsilon^A_{int}$
  (f) $(Op(K^t_A) \setminus (K^t_A \perp\!\!\!\perp_{\varepsilon^A_\perp} (\phi^C, p))) \cup \{\mathtt{R}((\phi^C, p), \frac{m}{n}), (\phi^C, p), (\phi^A, p)\}$ if $val^{\perp\!\!\!\perp}(K^t_A, (\phi^C, p)) > \varepsilon^A_\perp$ and $\frac{m}{n} > \varepsilon^A_{freq}$ and $p > \varepsilon^A_{int}$ and $(\phi^A, q) \notin K^t_A$ for any q
  $((Op(K^t_A) \setminus (K^t_A \perp\!\!\!\perp_{\varepsilon^A_\perp} (\phi^C, p))) \setminus \{(\phi^A, q)\}) \cup \{\mathtt{R}((\phi^C, p), \frac{m}{n}), (\phi^C, p), (\phi^A, p)\}$ if $val^{\perp\!\!\!\perp}(K^t_A, (\phi^C, p)) > \varepsilon^A_\perp$ and $\frac{m}{n} > \varepsilon^A_{freq}$, $(\phi^A, q) \in K^t_A$, $q < p > \varepsilon^A_{int}$ and ϕ^A contains at least one agent name different from A
  (g) $(Op(K^t_A) \setminus (K^t_A \perp\!\!\!\perp_{\varepsilon^A_\perp} (\phi^C, p))) \cup \{\mathtt{R}((\phi^C, p), \frac{m}{n}), (\phi^C, p)\}$ if $val^{\perp\!\!\!\perp}(K^t_A, (\phi^C, p)) > \varepsilon^A_\perp$ and $\frac{m}{n} > \varepsilon^A_{freq}$ and $p > \varepsilon^A_{int}$ and $(\phi^A, q) \in K^t_A$ with $p \leq q$

where

$$n = |\{t' \mid t' < t,\ \mathfrak{C}(t') = w(D, (\phi^C, p))\}|$$

$$m = |\{t' \mid t' < t,\ \mathfrak{C}(t') = w(E, (\psi^F, q))\}|$$

for any agents D, E, F, formula ψ^F and real number q.

Finally, the function defining the closure of opinion states with respect to logical derivability is provided below. It allows the agent to have in her opinion state all she could obtain through logical reasoning. Technically, we define an operation *ded* that adds opinions to a set if these can be derived from its elements and then define, as per usual practice, the closure of X as the smallest fixpoint Y of *ded* containing X. Since Y is the *smallest* fixpoint of *ded* containing X, Y can always be obtained from X by a possibly infinite number of *ded* applications.

Definition 5.6 (Closure): For any agent A and set of opinions X, $\mathrm{cl}_A(X)$ is the smallest superset Y of X such that $ded_A(Y) = Y$. The set $ded_A(Y)$ is defined by the following clauses:

(1) If $(\phi^C, p) \in Y$, $ded_A(Y) \supseteq Y \cup \{((\neg\phi^C)^A, 1-p)\}$
(2) If $((\neg\phi^C)^{C'}, p) \in Y$ and either $C = C' = A$ or $C \neq A$, $ded_A(Y) \supseteq Y \cup \{(\phi^C, 1-p)\}$
(3) If $(\phi^C, p), (\psi^{C'}, p') \in Y$, $ded_A(Y) \supseteq Y \cup \{((\phi^C \wedge \psi^{C'})^A, \frac{p \cdot p'}{p+p'})\}$
(4) If $((\phi^C \wedge \psi^{C'})^{C''}, p), (\psi^{C'}, p') \in Y$ and $\frac{p}{p'}$ is defined and either $C = C' = C'' = A$ or $C \neq A$, $ded_A(Y) \supseteq Y \cup \{(\phi^C, \frac{p' \cdot p}{p'-p})\}$
(5) If $((\phi^C \wedge \psi^{C'})^{C''}, p), (\phi^C, p') \in Y$ and $\frac{p}{p'}$ is defined and either $C = C' = C'' = A$ or $C' \neq A$, $ded_A(Y) \supseteq Y \cup \{(\psi^{C'}, \frac{p' \cdot p}{p'-p})\}$
(6) If $(\phi^C, p), (\psi^{C'}, p') \in Y$, $ded_A(Y) \supseteq Y \cup \{((\phi^C \vee \psi^{C'})^A, 1 - \frac{(1-p)\cdot(1-p')}{1-p+1-p'})\}$
(7) If $((\phi^C \vee \psi^{C'})^{C''}, p), (\psi^{C'}, p') \in Y$ and $\frac{(1-p')\cdot(1-p)}{(1-p')-(1-p)}$ is defined and either $C = C' = C'' = A$ or $C \neq A$, $ded_A(Y) \supseteq Y \cup \{(\phi^C, 1 - \frac{(1-p')\cdot(1-p)}{(1-p')-(1-p)})\}$
(8) If $((\phi^C \vee \psi^{C'})^{C''}, p), (\phi^C, p') \in Y$ and $\frac{(1-p')\cdot(1-p)}{(1-p')-(1-p)}$ is defined and either $C = C' = C'' = A$ or $C' \neq A$, $ded_A(Y) \supseteq Y \cup \{(\psi^{C'}, 1 - \frac{(1-p')\cdot(1-p)}{(1-p')-(1-p)})\}$
(9) If $(\phi^C, p), (\psi^{C'}, p') \in Y$ and $\frac{p'}{p}$ is defined, $ded_A(Y) \supseteq Y \cup \{((\phi^C \rightarrow \psi^{C'})^A, 1 - \frac{p\cdot(1-p')}{p+1-p'})\}$
(10) If $((\phi^C \rightarrow \psi^{C'})^{C''}, p), (\phi^C, p') \in Y$ and $0 \neq p'$ and either $C = C' = C'' = A$ or $C' \neq A$ then $ded_A(Y) \supseteq Y \cup \{(\psi^{C'}, \frac{p'\cdot(1-p)}{p})\}$

Intuitively, we do not formalise the truth of complex formulae in a traditional Tarskian way and we use instead the closure of the set K^i_j with respect to logical derivability since we do not handle all formulae compositionally. In other words, an agent might be able to derive the value of some formulae from the value of their components; for some other formulae, instead, she might not be able to do so and she might only be able to attribute a value to them in bulk. The former kind of formulae contains the innate opinions of the agent, which the agent can evaluate atomically; the latter kind contains the acquired opinions of the agent, which she might have received composed but not in atomic form. Hence, we would be able to use a set of Tarskian-style semantic clauses only for the truth of formulae which in the calculus are derived from the innate set of opinions. For the other complex formulae, we might not be able to do it. In order to have a homogeneous way of treating all complex formulae, we only use

a semantical construction based on the identification of set-containment and truth at a state.

Intuitively, the set $ded_A(Y)$ contains all elements of Y plus all pairs that can be obtained by applying a logical introduction or elimination rule to some elements of Y. Technically, then, the clauses defining $ded_A(Y)$ are structured according to the syntactical introductions and eliminations of connectives. For instance, clause (4) – which defines $ded_A(Y)$ with respect to disjunctions occurring in the set Y – states that, if $((\phi^C \wedge \psi^{C'})^{C''}, p)$ and $(\psi^{C'}, p')$ are elements of Y, then $(\phi^C, \frac{p' \cdot p}{p' - p})$ must be an element of the set $ded_A(Y)$. Rather obviously, the conviction value $\frac{p' \cdot p}{p' - p}$ occurring in the pair which must be added to Y in order to obtain $ded_A(Y)$ is computed according to the same arithmetical function used in the conjunction elimination rule. In other words, if the set Y contains one or two pairs, the form of which corresponds to those displayed in the premisses of a logical rule of the calculus, then $ded_A(Y)$ must contain at least all elements of Y and the pair occurring on the right hand side of the conclusion of that logical rule.

5.1. Soundness and completeness

We now prove the equivalence between calculus and semantics. Let us begin with a definition establishing a correspondence between sequent contexts and states in models, essential to state the soundness result.

Definition 5.7 (Profile–opinion state correspondence): For any profile $P = \mathfrak{L}; \Gamma^A; \Delta$ and state K_B^t in a model $\mathcal{M} = (\{K_j^i\}_{i \in T, j \in \mathbb{A}}, T, \mathbb{A}, \mathfrak{C}, \varepsilon)$, K_B^t is a P-state and P is a K_B^t-profile if, and only if, all following points hold:

- $A = B$;
- the thresholds for A are the same in the calculus and in $\mathcal{M}$;
- $\{(\phi^C, p) \mid (\phi^C, p) \in K_A^t\} = \Gamma^A \cup \Delta$, and
- $\mathfrak{C}(0), \ldots, \mathfrak{C}(t-1) = \mathfrak{L}$.

Notice that the profile $P = \mathfrak{L}; \Gamma^A; \Delta$ used as context of a sequent might represent an impossible situation. Indeed, the literature $\mathfrak{L}$ might describe an impossible exchange of opinions: one in which an agent writes an opinion that cannot be derived from its opinion state or one in which an agent reads an opinion which has not been written by any agent. If this is the case, the following statement trivially holds. This might seem odd, but it is absolutely comparable with the case in which we consider an instance $A, \neg A \vDash B$ of the consequence relation of classical logic. Since no model of classical logic verifies both A and $\neg A$, this instance of the consequence relation trivially holds. It is possible, in principle, to restrict the calculus is such a way that only reasonable literatures $\mathfrak{L}$ can be used inside profiles, but this would needlessly complicate the calculus.

In more standard systems, the previous correspondence might be generalised by requiring that the set $\{(\phi^C, p) \mid (\phi^C, p) \in K_A^t\}$ and the set $\Gamma^A \cup \Delta$ are logically equivalent rather than identical – for some notion of equivalence based on derivability. The

very nature of $\texttt{(un)SecureND}_{grad}$, though, makes its notion of derivability strongly contextual – notice, for instance, that only negation rules enable us to derive one opinion pair by considering only one opinion pair in the profile. As a consequence, there seems to be no obvious way of meaningfully defining a notion of inter-derivability between two opinion pairs without referring to a previously fixed profile. The generality that we usually find in the notion of semantic state is recovered here by considering the logical closure of the semantic states K_A^t when we define the conditions under which a structure $(\{K_j^i\}_{i\in T, j\in\mathbb{A}}, T, \mathbb{A}, \mathfrak{C}, \varepsilon)$ is a suitable model of $\texttt{(un)SecureND}_{grad}$, see Definitions 5.1 and 5.6.

Notice, moreover, that in the following proof, nothing is proved directly of sequents of the form $\mathfrak{L}^{++r(A,(\phi^{C'},p))}; \Gamma^A; \Delta \vdash \mathtt{R}((\phi^{C'},p), \frac{m}{n})$. Indeed, since its derivability occurs as premiss of other operational rules that induce the modification of a profile, deriving such a formula does not guarantee that the context $\mathfrak{L}^{++r(A,(\phi^{C'},p))}; \Gamma^A; \Delta$ indeed corresponds to the state that occurs *after* the update induced by the read operation. We must be able to derive $\mathfrak{L}^{++r(A,(\phi^{C'},p))}; \Gamma^A; \Delta \vdash \mathtt{R}((\phi^{C'},p), \frac{m}{n})$ also before we have updated our context $\mathfrak{L}^{++r(A,(\phi^{C'},p))}; \Gamma^A; \Delta$ since deriving this sequent is a precondition of the operations that can update the context with respect to the operation denoted by $r(A, (\phi^{C'}, p))$. The $\mathtt{R}$ rule is, nevertheless, considered precisely inasmuch as it is indirectly required to derive statements about the results of a reading operation on an opinion state. For similar reasons, the statement does not cover the rules with conclusion of the form $P :: \mathit{Profile}_{temp}$.

Theorem 5.1 (Soundness of the calculus): *For any profile $P = \mathfrak{L}; \Gamma^A; \Delta$ and model $\mathcal{M}$, all following points hold:*

(1) *If $\mathfrak{L}; \Gamma^A; \Delta :: \mathit{Profile}$ is derivable, then there exists a P-state.*
(2) *If $\mathfrak{L}; \Gamma^A; \Delta \vdash (\phi^C, p)$ is derivable, then there exists a P-state and $(\phi^C, p) \in \mathrm{cl}_A(K_A^t)$ holds for any P-state K_A^t.*
(3) *If $\mathfrak{L}; \Gamma^A; \Delta \vdash \mathtt{W}(\phi^C, p)$ is derivable, then there exists a P-state and $\mathtt{W}(\phi^C, p) \in \mathrm{cl}_A(K_A^{t+1})$ holds for any P-state K_A^t.*
(4) *If $\mathfrak{L}; \Gamma^A; \Delta \vdash \mathcal{E}$ is derivable and $\mathcal{E}$ is of the form $\mathtt{R}(\ldots)$, $\mathtt{Tr}(\ldots)$ or $\mathtt{DisTr}(\ldots)$, then*
   - *there exists a P'-state where P' is the context of the bottommost statement of the form $P' :: \mathit{Profile}$ occurring above $\mathfrak{L}; \Gamma^A; \Delta \vdash \mathcal{E}$*
   - *and $\mathcal{E} \in \mathrm{cl}_A(K_A^{t+1})$ holds for any state K_A^{t+1} which is the successor of any P'-state K_A^t of a model $\mathcal{M}$ in which $\mathfrak{C}(0), \ldots, \mathfrak{C}(t) = \mathfrak{L}$.*

Proof: The proof is by induction on the number of rule applications occurring in the derivation of S.

No derivation can contain 0 rule applications. Hence, in the base case, an $::$ rule has been applied. Suppose its conclusion is $; \Gamma^A; :: \mathit{Profile}$. It is clearly always possible to construct a model such that a state K_A^0 of it is a P-state, for $P =; \Gamma^A;$.

We then suppose that the statement holds for all sequents with a derivation containing less than n rule applications and show that it also holds for sequents with

a derivation containing n rule applications. We consider the last rule applied in the derivation of S:

- $\dfrac{\mathfrak{L}; \Gamma^A, (\phi^A, p); \Delta :: \textit{Profile}}{\mathfrak{L}; \Gamma^A, (\phi^A, p); \Delta \vdash (\phi^A, p)}$ StI
  By inductive hypothesis, there exists a P-state K_A^t, for $P = \mathfrak{L}; \Gamma^A, (\phi^A, p); \Delta$. By Definition 5.7, $(\Gamma^A \cup \{(\phi^A, p)\}) \cup \Delta \subseteq K_A^t$ and hence $(\phi^A, p) \in K_A^t \subseteq \mathrm{cl}_A(K_A^t)$, as desired.
- $\dfrac{\mathfrak{L}; \Gamma^A; \Delta, (\phi^C, p) :: \textit{Profile}}{\mathfrak{L}; \Gamma^A; \Delta, (\phi^C, p) \vdash (\phi^C, p)}$ StA
  By inductive hypothesis, there exists a P-state K_A^t, for $P = \mathfrak{L}; \Gamma^A; \Delta, (\phi^C, p)$. By Definition 5.7, $\Gamma^A \cup (\Delta \cup \{(\phi^C, p)\}) \subseteq K_A^t$ and hence $(\phi^C, p) \in K_A^t \subseteq \mathrm{cl}_A(K_A^t)$, as desired.
- $\dfrac{\mathfrak{L}; \Gamma^A; \Delta \vdash (\phi^C, p)}{\mathfrak{L}; \Gamma^A; \Delta \vdash ((\neg\phi^C)^A, 1-p)}$ $\neg I$
  By inductive hypothesis, there exists a P-state, for $P = \mathfrak{L}; \Gamma^A; \Delta$. Let us then consider a generic P-state K_A^t of a generic model $\mathcal{M}$. By inductive hypothesis, $(\phi^C, p) \in \mathrm{cl}_A(K_A^t)$. By Definition 5.6.1., we also have that $((\neg\phi^C)^A, 1-p) \in \mathrm{cl}_A(K_A^t)$, as desired.
- $\dfrac{\mathfrak{L}; \Gamma^A; \Delta \vdash ((\neg\phi^C)^{C'}, p)}{\mathfrak{L}; \Gamma^A; \Delta \vdash (\phi^C, 1-p)}$ $\neg E$
  By inductive hypothesis, there exists a P-state, for $P = \mathfrak{L}; \Gamma^A; \Delta$. Let us then consider a generic P-state K_A^t of a generic model $\mathcal{M}$. By inductive hypothesis, $((\neg\phi^C)^{C'}, p) \in \mathrm{cl}_A(K_A^t)$. By Definition 5.6.2., we also have that $(\phi^C, 1-p) \in \mathrm{cl}_A(K_A^t)$, as desired.
- $\dfrac{\mathfrak{L}; \Gamma^A; \Delta \vdash (\phi^C, p) \quad \mathfrak{L}; \Gamma^A; \Delta \vdash (\psi^{C'}, p')}{\mathfrak{L}; \Gamma^A; \Delta \vdash ((\phi^C \wedge \psi^{C'})^A, \frac{p \cdot p'}{p+p'})}$ $\wedge I$
  By inductive hypothesis, there exists a P-state, for $P = \mathfrak{L}; \Gamma^A; \Delta$. Let us then consider a generic P-state K_A^t of a generic model $\mathcal{M}$. By inductive hypothesis, $(\phi^C, p), (\psi^{C'}, p') \in \mathrm{cl}_A(K_A^t)$. By Definition 5.6.3., we also have that $((\phi^C \wedge \psi^{C'})^A, \frac{p \cdot p'}{p+p'}) \in \mathrm{cl}_A(K_A^t)$, as desired.
- $\dfrac{\mathfrak{L}; \Gamma^A; \Delta \vdash ((\phi^C \wedge \psi^{C'})^{C''}, p) \quad \mathfrak{L}; \Gamma^A; \Delta \vdash (\psi^{C'}, p')}{\mathfrak{L}; \Gamma^A; \Delta \vdash (\phi^C, \frac{p' \cdot p}{p'-p})}$ $\wedge E$
  By inductive hypothesis, there exists a P-state, for $P = \mathfrak{L}; \Gamma^A; \Delta$. Let us then consider a generic P-state K_A^t of a generic model $\mathcal{M}$. By inductive hypothesis, $((\phi^C \wedge \psi^{C'})^{C''}, p), (\psi^{C'}, p') \in \mathrm{cl}_A(K_A^t)$. By Definition 5.6.4., we also have that $(\phi^C, \frac{p' \cdot p}{p'-p}) \in \mathrm{cl}_A(K_A^t)$, as desired.
- $\dfrac{\mathfrak{L}; \Gamma^A; \Delta \vdash ((\phi^{C'} \wedge \psi^C)^{C''}, p) \quad \mathfrak{L}; \Gamma^A; \Delta \vdash (\phi^{C'}, p')}{\mathfrak{L}; \Gamma^A; \Delta \vdash (\psi^C, \frac{p' \cdot p}{p'-p})}$ $\wedge E$
  This case is symmetric to the previous one.
- $\dfrac{\mathfrak{L}; \Gamma^A; \Delta \vdash (\phi^C, p) \quad \mathfrak{L}; \Gamma^A; \Delta \vdash (\psi^{C'}, p')}{\mathfrak{L}; \Gamma^A; \Delta \vdash ((\phi^C \vee \psi^{C'})^A, 1 - \frac{(1-p)\cdot(1-p')}{1-p+1-p'})}$ $\vee I$

By inductive hypothesis, there exists a P-state, for $P = \mathfrak{L}; \Gamma^A; \Delta$. Let us then consider a generic P-state K_A^t of a generic model $\mathcal{M}$. By inductive hypothesis, $(\phi^C, p), (\psi^{C'}, p') \in \mathrm{cl}_A(K_A^t)$. By Definition 5.6.6., we also have that $((\phi^C \vee \psi^{C'})^A, 1 - \frac{(1-p)\cdot(1-p')}{1-p+1-p'}) \in \mathrm{cl}_A(K_A^t)$, as desired.

- $$\frac{\mathfrak{L}; \Gamma^A; \Delta \vdash ((\phi^C \vee \psi^{C'})^{C''}, p) \quad \mathfrak{L}; \Gamma^A; \Delta \vdash (\psi^{C'}, p')}{\mathfrak{L}; \Gamma^A; \Delta \vdash (\phi^C, 1 - \frac{(1-p')\cdot(1-p)}{(1-p')-(1-p)})} \vee E$$

By inductive hypothesis, there exists a P-state, for $P = \mathfrak{L}; \Gamma^A; \Delta$. Let us then consider a generic P-state K_A^t of a generic model $\mathcal{M}$. By inductive hypothesis, $((\phi^C \vee \psi^{C'})^{C''}, p), (\psi^{C'}, p') \in \mathrm{cl}_A(K_A^t)$. By Definition 5.6.7., we also have that $(\phi^C, 1 - \frac{(1-p')\cdot(1-p)}{(1-p')-(1-p)}) \in \mathrm{cl}_A(K_A^t)$, as desired.

- $$\frac{\mathfrak{L}; \Gamma^A; \Delta \vdash ((\phi^{C'} \vee \psi^{C})^{C''}, p) \quad \mathfrak{L}; \Gamma^A; \Delta \vdash (\phi^{C'}, p')}{\mathfrak{L}; \Gamma^A; \Delta \vdash (\psi^{C}, 1 - \frac{(1-p')\cdot(1-p)}{(1-p')-(1-p)})} \vee E$$

This case is symmetric to the previous one.

- $$\frac{\mathfrak{L}; \Gamma^A; \Delta \vdash (\phi^C, p) \quad \mathfrak{L}; \Gamma^A; \Delta \vdash (\psi^{C'}, p')}{\mathfrak{L}; \Gamma^A; \Delta \vdash ((\phi^C \to \psi^{C'})^A, 1 - \frac{p\cdot(1-p')}{p+1-p'})} \to I$$

By inductive hypothesis, there exists a P-state, for $P = \mathfrak{L}; \Gamma^A; \Delta$. Let us then consider a generic P-state K_A^t of a generic model $\mathcal{M}$. By inductive hypothesis, $(\phi^C, p), (\psi^{C'}, p') \in \mathrm{cl}_A(K_A^t)$. By Definition 5.6.9., we also have that $((\phi^C \to \psi^{C'})^A, 1 - \frac{p\cdot(1-p')}{p+1-p'}) \in \mathrm{cl}_A(K_A^t)$, as desired.

- $$\frac{\mathfrak{L}; \Gamma^A; \Delta \vdash ((\phi^{C'} \to \psi^{C})^{C''}, p) \quad \mathfrak{L}; \Gamma^A; \Delta \vdash (\phi^{C'}, p')}{\mathfrak{L}; \Gamma^A; \Delta \vdash (\psi^{C}, \frac{p'\cdot(1-p)}{p})} \to E$$

By inductive hypothesis, there exists a P-state, for $P = \mathfrak{L}; \Gamma^A; \Delta$. Let us then consider a generic P-state K_A^t of a generic model $\mathcal{M}$. By inductive hypothesis, $((\phi^{C'} \to \psi^{C})^{C''}, p), (\phi^{C'}, p') \in \mathrm{cl}_A(K_A^t)$. By Definition 5.6.10., we also have that $(\psi^{C}, \frac{p'\cdot(1-p)}{p}) \in \mathrm{cl}_A(K_A^t)$, as desired.

- $$\frac{\mathfrak{L}; \Gamma^A; \Delta :: \mathit{Profile} \quad p > 0.5}{\mathfrak{L}^{++w(B,(\phi^C,p))}; \Gamma^A; \Delta :: \mathit{Profile}} ++ \qquad \frac{\mathfrak{L}; \Gamma^A; \Delta :: \mathit{Profile} \quad w(B, (\phi^C, p)) \in \mathfrak{L}}{\mathfrak{L}^{++r(B',(\phi^C,p))}; \Gamma^A; \Delta :: \mathit{Profile}} ++$$

where $A \neq B$

Let $P_1 = \mathfrak{L}^{++x(B,(\phi^C,p))}; \Gamma^A; \Delta$, where $x \in \{w, r\}$, and $P_0 = \mathfrak{L}; \Gamma^A; \Delta$. By inductive hypothesis, there exists a P_0-state in some model $\mathcal{M}$. If we consider a model $\mathcal{M}$ such that $\mathfrak{C}(0), \ldots, \mathfrak{C}(t) = \mathfrak{L}^{++x(B,(\phi^C,p))}$ and such that K_A^t is a P_0-state, we clearly have that K_A^{t+1} is a P_1-state.

- $$\frac{\mathfrak{L}; \Gamma^A; \Delta \vdash (\phi^C, p) \quad p > 0.5}{\mathfrak{L}^{++w(A,(\phi^C,p))}; \Gamma^A; \Delta \vdash \mathbb{W}(\phi^C, p)} \mathbb{W}$$

Let $P_1 = \mathfrak{L}^{++w(A,(\phi^C,p))}; \Gamma^A; \Delta$ and $P_0 = \mathfrak{L}; \Gamma^A; \Delta$. By inductive hypothesis, there exists a P_0-state in some model $\mathcal{M}$. If we consider a model $\mathcal{M}$ such that $\mathfrak{C}(0), \ldots, \mathfrak{C}(t) = \mathfrak{L}^{++w(A,(\phi^C,p))}$ and such that K_A^t is a P_0-state, we clearly have that K_A^{t+1} is a P_1-state.

Let us consider now a generic P_1-state K_A^{t+1}. By case *2* of Definition 5.5 and since, by inductive hypothesis, $(\phi^C, p) \in \mathrm{cl}_A(K_A^t)$ for any P_0-state, K_A^{t+1} must have

been constructed as $Up(K_A^t, w(A, (\phi^C, p))) = K_A^t \cup \{\mathtt{W}(\phi^C, p)\}$ for some P_0-state K_A^t. Therefore, as desired, $\mathtt{W}(\phi^C, p) \in \mathrm{cl}_A(K_A^{t+1})$.

- $$\frac{\mathfrak{L}_0; \Gamma^C, \Xi \vdash \mathtt{W}(\phi^{C'}, p) \quad \mathfrak{L}; \Gamma^A; \Delta :: \mathit{Profile}}{\mathfrak{L}^{++r(A,(\phi^{C'},p))}; \Gamma^A; \Delta \vdash \mathtt{R}((\phi^{C'}, p), \frac{m}{n})}\ \mathtt{R}$$

  where $m = \#\mathfrak{L}(w(C, (\phi^{C'}, p)))$, $n = \#\mathfrak{L}(w)$, $\mathfrak{L} = \mathfrak{L}_0^{++\mathfrak{L}_1}$ for some, possibly empty, literature $\mathfrak{L}_1$.

  Let $P_0 = \mathfrak{L}_0; \Gamma^C, \Xi$, $P = \mathfrak{L}; \Gamma^A; \Delta$. By inductive hypothesis, there exist a P-state and a P_0-state. Moreover, again by inductive hypothesis, $\mathtt{W}(\phi^{C'}, p) \in \mathrm{cl}(K_C^{t_0})$ for any P_0-state $K_C^{t_0}$. If we consider a model $\mathcal{M}$ such that $\mathfrak{C}(0), \ldots, \mathfrak{C}(t) = \mathfrak{L}^{++r(A,(\phi^{C'},p))}$ which contains a P-state K_A^t, then it can contain a P_0-state $K_C^{t_0}$ since $\mathfrak{L} = \mathfrak{L}_0^{++\mathfrak{L}_1}$. Let us consider now a generic P-state K_A^t in a model which contains P_0-state $K_C^{t_0}$ and such that $\mathfrak{C}(0), \ldots, \mathfrak{C}(t) = \mathfrak{L}^{++r(A,(\phi^{C'},p))}$. By case *3* of Definition 5.5 and since, by inductive hypothesis, $\mathtt{W}(\phi^{C'}, p) \in \mathrm{cl}(K_C^{t_0})$ for any P_0-state $K_C^{t_0}$, K_A^{t+1} must have been constructed as $Up(K_A^t, r(A, (\phi^{C'}, p)))$ for some P_0-state K_A^t. Therefore, as desired, $\mathtt{R}((\phi^{C'}, p), \frac{m}{n}) \in \mathrm{cl}_A(K_A^{t+1})$. Indeed, since $\mathfrak{L} = \mathfrak{C}(0), \ldots, \mathfrak{C}(t)$, $n = n'$ and $m = m'$.

- $$\frac{\mathfrak{L}; \Gamma^A; \Delta \vdash \mathtt{R}((\phi^C, p), \frac{m}{n}) \quad \mathit{val}^{\perp\!\!\!\perp}(\Gamma^A; \Delta, (\phi^C, p)) \leq \varepsilon_\perp^A}{\mathfrak{L}; \Gamma^A; \Delta \vdash \mathtt{Tr}(\phi^C, p)}\ \mathtt{Tr}$$

  Let $P = \mathfrak{L}; \Gamma^A; \Delta$. By inductive hypothesis, there exists a P'-state – where P' is the context of the bottommost statement of the form P' :: *Profile* occurring above our rule – and $\mathtt{R}((\phi^C, p), \frac{m}{n}) \in \mathrm{cl}_A(K_A^{t+1})$ holds for any state K_A^{t+1} which is the successor of any P'-state K_A^t of a model $\mathcal{M}$ in which $\mathfrak{C}(0), \ldots, \mathfrak{C}(t) = \mathfrak{L}$. Now, since
  - $\mathfrak{C}(t) = r(A, (\phi^{C'}, p))$ since $\mathtt{R}((\phi^C, p), \frac{m}{n}) \in \mathrm{cl}_A(K_A^{t+1})$,
  - $\mathit{val}^{\perp\!\!\!\perp}(\Gamma^A; \Delta, (\phi^C, p)) \leq \varepsilon_\perp^A$,
  - $\mathit{val}^{\perp\!\!\!\perp}(\Gamma^A, \Delta, (\phi^C, p)) = \mathit{val}^{\perp\!\!\!\perp}(K_A^t, (\phi^C, p))$ (see Definitions 3.4 and 5.2) since, by inspection of the rules of the calculus and by inductive hypothesis, $\Gamma^A; \Delta$ are exactly the lists occurring in the context corresponding to K_A^t;

  by cases *(3)(a)–(c)* of Definition 5.5, we can conclude that $\mathtt{Tr}(\phi^C, p) \in \mathrm{cl}_A(K_A^{t+1})$, as desired.

- $$\frac{\mathfrak{L}; \Gamma^A; \Delta \vdash \mathtt{Tr}(\phi^C, p)}{\mathfrak{L}; \Gamma^A; \Delta, (\phi^C, p) :: \mathit{Profile}_{temp}}\ \mathtt{ATr}$$

  There is nothing to prove in this case.

- $$\frac{\mathfrak{L}; \Gamma^A; \Delta \vdash \mathtt{R}((\phi^C, p), \frac{m}{n}) \quad \mathit{val}^{\perp\!\!\!\perp}(\Gamma^A; \Delta, (\phi^C, p)) > \varepsilon_\perp^A \quad \frac{m}{n} \leq \varepsilon_{freq}^A}{\mathfrak{L}; \Gamma^A; \Delta \vdash \mathtt{DisTr}(\phi^C, p)}\ \mathtt{DisTr}$$

  Let $P = \mathfrak{L}; \Gamma^A; \Delta$. By inductive hypothesis, there exists a P'-state – where P' is the context of the bottommost statement of the form P' :: *Profile* occurring above our rule – and $\mathtt{R}((\phi^C, p), \frac{m}{n}) \in \mathrm{cl}_A(K_A^{t+1})$ holds for any state K_A^{t+1} which is the successor of any P'-state K_A^t of a model $\mathcal{M}$ in which $\mathfrak{C}(0), \ldots, \mathfrak{C}(t) = \mathfrak{L}$. Now, since
  - $\mathfrak{C}(t) = r(A, (\phi^{C'}, p))$,
  - $\mathit{val}^{\perp\!\!\!\perp}(\Gamma^A; \Delta, (\phi^C, p)) > \varepsilon_\perp^A$,

- $val^{\perp\!\!\!\perp}(\Gamma^A, \Delta, (\phi^C, p)) = val^{\perp\!\!\!\perp}(K_A^t, (\phi^C, p))$ (see Definitions 3.4 and 5.2) since, by inspection of the rules of the calculus and by inductive hypothesis, $\Gamma^A; \Delta$ are exactly the lists occurring in the context corresponding to K_A^t;

by case *(3)(d)* of Definition 5.5, we can conclude that $\texttt{DisTr}(\phi^C, p) \in \mathrm{cl}_A(K_A^{t+1})$, as desired.

- $$\frac{\mathfrak{L}; \Gamma^A; \Delta \vdash \texttt{R}((\phi^C, p), \frac{m}{n}) \quad val^{\perp\!\!\!\perp}(\Gamma^A; \Delta, (\phi^C, p)) > \varepsilon_\perp^A \quad \frac{m}{n} > \varepsilon_{freq}^A}{\mathfrak{L}; \Gamma'^A; \Delta', (\phi^C, p) :: \mathit{Profile}_{temp}} \texttt{AMis}$$

  where $\Gamma'^A = \Gamma^A \setminus (\Gamma^A \perp\!\!\!\perp_{\varepsilon_\perp^A} (\Delta, (\phi^C, p)))$ and $\Delta' = \Delta \setminus (\Delta \perp\!\!\!\perp_{\varepsilon_\perp^A} (\Gamma^A, (\phi^C, p)))$.
  There is nothing to prove in this case.
- $$\frac{\mathfrak{L}; \Gamma^A; \Delta \vdash \texttt{W}(\phi^C, p)}{\mathfrak{L}; \Gamma^A; \Delta :: \mathit{Profile}} \texttt{::W}$$

  Let $P = \mathfrak{L}; \Gamma^A; \Delta$. By inductive hypothesis, there exists a P-state in some model $\mathcal{M}$, as desired.
- $$\frac{\mathfrak{L}; \Gamma^A; \Delta \vdash \texttt{DisTr}(\phi^C, p)}{\mathfrak{L}; \Gamma^A; \Delta :: \mathit{Profile}} \texttt{::DisTr}$$

  Let $P = \mathfrak{L}; \Gamma^A; \Delta$. By inductive hypothesis, there exists a P-state in some model $\mathcal{M}$, as desired.
- $$\frac{\mathfrak{L}; \Gamma^A; \Delta, (\phi^C, p) :: \mathit{Profile}_{temp} \quad p \leq \varepsilon_{int}^A}{\mathfrak{L}; \Gamma^A, \Delta, (\phi^C, p) :: \mathit{Profile}} \texttt{DisI}$$

  Now, the premiss $\mathfrak{L}; \Gamma^A; \Delta, (\phi^C, p) :: \mathit{Profile}_{temp}$ must be the conclusion either of an $\texttt{ATr}$ or of an $\texttt{AMis}$ rule application.

  In the first case, consider the premiss $\mathfrak{L}; \Gamma^A; \Delta \vdash \texttt{Tr}(\phi^C, p)$ of the $\texttt{ATr}$ rule application. By inductive hypothesis, there exists a P'-state – where P' is the context of the bottommost statement of the form $P' :: \mathit{Profile}$ occurring above the $\texttt{ATr}$ rule – and $\texttt{Tr}(\phi^C, p) \in \mathrm{cl}_A(K_A^{t+1})$ holds for any state K_A^{t+1} which is the successor of any P'-state K_A^t of a model $\mathcal{M}$ in which $\mathfrak{C}(0), \ldots, \mathfrak{C}(t) = \mathfrak{L}$. Now, since we can indeed consider a model in which $\mathfrak{C}(0), \ldots, \mathfrak{C}(t) = \mathfrak{L}$, in it, K_A^{t+1} will be the successor of K_A^t and $\texttt{Tr}(\phi^C, p) \in \mathrm{cl}_A(K_A^{t+1})$. It is clear, considering the premisses of the rule and the thresholds of the model, that K_A^{t+1} can contain $\texttt{Tr}(\phi^C, p)$ only due to case *(3)(a)* of Definition 5.5. Thus, K_A^{t+1} contains also (ϕ^C, p) and, as desired, K_A^{t+1} is also a P-state for $P = \mathfrak{L}; \Gamma^A, \Delta, (\phi^C, p)$.

  In the second case, consider the premiss $\mathfrak{L}; \Gamma^A; \Delta \vdash \texttt{R}((\phi^C, p), \frac{m}{n})$ of the $\texttt{AMis}$ rule application. By inductive hypothesis, there exists a P'-state – where P' is the context of the bottommost statement of the form $P' :: \mathit{Profile}$ occurring above the $\texttt{AMis}$ rule – and $\texttt{R}((\phi^C, p), \frac{m}{n}) \in \mathrm{cl}_A(K_A^{t+1})$ holds for any state K_A^{t+1} which is the successor of any P'-state K_A^t of a model $\mathcal{M}$ in which $\mathfrak{C}(0), \ldots, \mathfrak{C}(t) = \mathfrak{L}$. Now, since we can indeed consider a model in which $\mathfrak{C}(0), \ldots, \mathfrak{C}(t) = \mathfrak{L}$, in it, K_A^{t+1} will be the successor of K_A^t and $\texttt{R}((\phi^C, p), \frac{m}{n}) \in \mathrm{cl}_A(K_A^{t+1})$. Clearly, considering the premisses of the rule and the thresholds of the model, K_A^{t+1} can contain $\texttt{Tr}(\phi^C, p)$ only due to case *(3)(e)* of Definition 5.5. Thus, K_A^{t+1} contains also (ϕ^C, p) and, as desired, K_A^{t+1} is also a P-state for $P = \mathfrak{L}; \Gamma^A, \Delta, (\phi^C, p)$.
- $$\frac{\mathfrak{L}; \Gamma^A; \Delta, (\phi^C, p) :: \mathit{Profile}_{temp} \quad (\phi^A, q) \notin \Delta \quad p > \varepsilon_{int}^A}{\mathfrak{L}; \Gamma^A, \Delta, (\phi^C, p), (\phi^A, p) :: \mathit{Profile}} \texttt{I}$$

  where ϕ^A contains at least one agent name different from A.

Now, the premiss $\mathfrak{L}; \Gamma^A; \Delta, (\phi^C, p) :: \mathit{Profile}_{temp}$ must be the conclusion either of an $\mathtt{ATr}$ or of an $\mathtt{AMis}$ rule application.

In the first case, consider the premiss $\mathfrak{L}; \Gamma^A; \Delta \vdash \mathtt{Tr}(\phi^C, p)$ of the $\mathtt{ATr}$ rule application. By inductive hypothesis, there exists a P'-state – where P' is the context of the bottommost statement of the form $P' :: \mathit{Profile}$ occurring above the $\mathtt{ATr}$ rule – and $\mathtt{Tr}(\phi^C, p) \in \mathrm{cl}_A(K_A^{t+1})$ holds for any state K_A^{t+1} which is the successor of any P'-state K_A^t of a model $\mathcal{M}$ in which $\mathfrak{C}(0), \ldots, \mathfrak{C}(t) = \mathfrak{L}$. Now, since we can indeed consider a model in which $\mathfrak{C}(0), \ldots, \mathfrak{C}(t) = \mathfrak{L}$, in it, K_A^{t+1} will be the successor of K_A^t and $\mathtt{Tr}(\phi^C, p) \in \mathrm{cl}_A(K_A^{t+1})$. It is clear, considering the premisses of the rule and the thresholds of the model, that K_A^{t+1} can contain $\mathtt{Tr}(\phi^C, p)$ only due to case *(3)(b)* of Definition 5.5. Thus, K_A^{t+1} contains also $(\phi^C, p), (\phi^A, p)$ and, as desired, K_A^{t+1} is also a P-state for $P = \mathfrak{L}; \Gamma^A, \Delta, (\phi^C, p), (\phi^A, p)$.

In the second case, consider the premiss $\mathfrak{L}; \Gamma^A; \Delta \vdash \mathtt{R}((\phi^C, p), \frac{m}{n})$ of the $\mathtt{AMis}$ rule application. By inductive hypothesis, there exists a P'-state – where P' is the context of the bottommost statement of the form $P' :: \mathit{Profile}$ occurring above the $\mathtt{AMis}$ rule – and $\mathtt{R}((\phi^C, p), \frac{m}{n}) \in \mathrm{cl}_A(K_A^{t+1})$ holds for any state K_A^{t+1} which is the successor of any P'-state K_A^t of a model $\mathcal{M}$ in which $\mathfrak{C}(0), \ldots, \mathfrak{C}(t) = \mathfrak{L}$. Now, since we can indeed consider a model in which $\mathfrak{C}(0), \ldots, \mathfrak{C}(t) = \mathfrak{L}$, in it, K_A^{t+1} will be the successor of K_A^t and $\mathtt{R}((\phi^C, p), \frac{m}{n}) \in \mathrm{cl}_A(K_A^{t+1})$. It is clear, considering the premisses of the rule and the thresholds of the model, that K_A^{t+1} can contain $\mathtt{Tr}(\phi^C, p)$ only due to case *(3)(f)* of Definition 5.5. Thus, K_A^{t+1} contains also $(\phi^C, p), (\phi^A, p)$ and, as desired, K_A^{t+1} is also a P-state for $P = \mathfrak{L}; \Gamma^A, \Delta, (\phi^C, p), (\phi^A, p)$.

- $$\frac{\mathfrak{L}; \Gamma^A; \Delta, (\phi^C, p) :: \mathit{Profile}_{temp} \quad (\phi^A, q) \in \Delta \quad q < p > \varepsilon^A_{int}}{\mathfrak{L}; \Gamma'^A, \Delta', (\phi^C, p), (\phi^A, p) :: \mathit{Profile}} \ \mathtt{I}$$

  where $\Delta' = \Delta \setminus \{(\phi^A, q)\}$ and ϕ^A contains at least one agent name different from A.

  Now, the premiss $\mathfrak{L}; \Gamma^A; \Delta, (\phi^C, p) :: \mathit{Profile}_{temp}$ must be the conclusion either of an $\mathtt{ATr}$ or of an $\mathtt{AMis}$ rule application.

  In the first case, consider the premiss $\mathfrak{L}; \Gamma^A; \Delta \vdash \mathtt{Tr}(\phi^C, p)$ of the $\mathtt{ATr}$ rule application. By inductive hypothesis, there exists a P'-state – where P' is the context of the bottommost statement of the form $P' :: \mathit{Profile}$ occurring above the $\mathtt{ATr}$ rule – and $\mathtt{Tr}(\phi^C, p) \in \mathrm{cl}_A(K_A^{t+1})$ holds for any state K_A^{t+1} which is the successor of any P'-state K_A^t of a model $\mathcal{M}$ in which $\mathfrak{C}(0), \ldots, \mathfrak{C}(t) = \mathfrak{L}$. Now, since we can indeed consider a model in which $\mathfrak{C}(0), \ldots, \mathfrak{C}(t) = \mathfrak{L}$, in it, K_A^{t+1} will be the successor of K_A^t and $\mathtt{Tr}(\phi^C, p) \in \mathrm{cl}_A(K_A^{t+1})$. Clearly, considering the premisses of the rule and the thresholds of the model, K_A^{t+1} can contain $\mathtt{Tr}(\phi^C, p)$ only due to case *(3)(c)* of Definition 5.5. Thus, K_A^{t+1} contains also $(\phi^C, p), (\phi^A, p)$ and, as desired, K_A^{t+1} is also a P-state for $P = \mathfrak{L}; \Gamma^A, \Delta, (\phi^C, p), (\phi^A, p)$.

  In the second case, consider the premiss $\mathfrak{L}; \Gamma^A; \Delta \vdash \mathtt{R}((\phi^C, p), \frac{m}{n})$ of the $\mathtt{AMis}$ rule application. By inductive hypothesis, there exists a P'-state – where P' is the context of the bottommost statement of the form $P' :: \mathit{Profile}$ occurring above the $\mathtt{AMis}$ rule – and $\mathtt{R}((\phi^C, p), \frac{m}{n}) \in \mathrm{cl}_A(K_A^{t+1})$ holds for any state K_A^{t+1} which is the successor of any P'-state K_A^t of a model $\mathcal{M}$ in which $\mathfrak{C}(0), \ldots, \mathfrak{C}(t) = \mathfrak{L}$. Now, since we can indeed consider a model in which $\mathfrak{C}(0), \ldots, \mathfrak{C}(t) = \mathfrak{L}$, in it, K_A^{t+1} will be the

successor of K_A^t and $\mathtt{R}((\phi^C, p), \frac{m}{n}) \in \mathrm{cl}_A(K_A^{t+1})$. It is clear, considering the premisses of the rule and the thresholds of the model, that K_A^{t+1} can contain $\mathtt{Tr}(\phi^C, p)$ only due to case *(3)(f)* of Definition 5.5. Thus, K_A^{t+1} contains also $(\phi^C, p)(\phi^A, p)$ and, as desired, K_A^{t+1} is also a P-state for $P = \mathfrak{L}; \Gamma^A, \Delta, (\phi^C, p), (\phi^A, p)$ ■.

We now need a measure of the complexity of a derivation of $P \vdash \mathcal{E}$ in case $\mathcal{E} \in \mathrm{cl}(K_A^t)$ and P is a K_A^t-profile. This will be used in the proof of Theorem 5.2. Intuitively, the age of an expression is a value that indicates how many steps it takes (in terms of applications of the update functions *Up* and ded_A introduced in Definitions 5.5 and 5.6) to justify that the expression $\mathcal{E}$ is contained in the closure of the state K_A^t. The condition 'unless another case applies to $\mathcal{E}$ and yields a smaller value' is meant to guarantee that the number of steps considered is minimal and hence that the age of the expression $\mathcal{E}$ is unique.

Before defining the age of an expression, we introduce some useful notation.

Definition 5.8 (Operations with pairs and order): Addition of pairs of numbers is defined element-wise: $(n, m) + (n', m') = (n + n', m + m')$. Projections of pairs are defined as usual: $\mathrm{fst}(n, m) = n$ and $\mathrm{snd}(n, m) = m$. We order pairs of numbers lexicographically.

Definition 5.9 (Age of an expression): For any model $\mathcal{M} = (\{K_j^i\}_{i \in T, j \in \mathbb{A}}, T, \mathbb{A}, \mathfrak{C}, \varepsilon)$, state K_A^t in $\mathcal{M}$ and expression $\mathcal{E}$ such that $\mathcal{E} \in \mathrm{cl}(K_A^t)$, the measure $\mathrm{age}(\mathcal{E}, K_A^t)$ of $\mathcal{E}$ with respect to K_A^t is inductively defined as follows:

- if $\mathcal{E} \in K_A^t$, then $\mathrm{age}(\mathcal{E}, K_A^t) = (t, 0)$;
- if $\mathcal{E} \notin K_A^t$ and $\mathcal{E} = ((\neg\phi^C)^A, 1 - p)$, then $\mathrm{age}(\mathcal{E}, K_A^t) = \mathrm{age}((\phi^C, p), K_A^t) + (0, 1)$, unless another case applies to $\mathcal{E}$ and yields a smaller value;
- if $\mathcal{E} \notin K_A^t$, $\mathcal{E} = (\phi^C, 1 - p)$ and either $C = C' = A$ or $C \neq A$, then $\mathrm{age}(\mathcal{E}, K_A^t) = \mathrm{age}(((\neg\phi^C)^{C'}, p), K_A^t) + (0, 1)$, unless another case applies to $\mathcal{E}$ and yields a smaller value;
- if $\mathcal{E} \notin K_A^t$ and $\mathcal{E} = ((\phi^C \wedge \psi^{C'})^A, \frac{p \cdot p'}{p + p'})$ where $p + p' \neq 0$, then $\mathrm{age}(\mathcal{E}, K_A^t) = (t, n + m + 1)$ where $n = \mathrm{snd}(\mathrm{age}((\phi^C, p), K_A^t))$ and $m = \mathrm{snd}(\mathrm{age}((\psi^{C'}, p'), K_A^t))$ unless another case applies to $\mathcal{E}$ and yields a smaller value;
- if $\mathcal{E} \notin K_A^t$, $\mathcal{E} = (\phi^C, \frac{p' \cdot p}{p' - p})$ where $p' - p > 0$, $p' \cdot p \leq p' - p$ and either $C = C' = C'' = A$ or $C \neq A$, then $\mathrm{age}(\mathcal{E}, K_A^t) = (t, n + m + 1)$ where $n = \mathrm{snd}(\mathrm{age}(((\phi^C \wedge \psi^{C'})^{C''}, p), K_A^t))$ and $m = \mathrm{snd}(\mathrm{age}((\psi^{C'}, p'), K_A^t))$, unless another case applies to $\mathcal{E}$ and yields a smaller value;
- if $\mathcal{E} \notin K_A^t$, $\mathcal{E} = (\psi^{C'}, \frac{p' \cdot p}{p' - p})$ where $p' - p > 0$, $p' \cdot p \leq p' - p$ and either $C = C' = C'' = A$ or $C \neq A$, then $\mathrm{age}(\mathcal{E}, K_A^t) = (t, n + m + 1)$ where $n = \mathrm{snd}(\mathrm{age}(((\phi^C \wedge \psi^{C'})^{C''}, p), K_A^t))$ and $m = \mathrm{snd}(\mathrm{age}((\phi^C, p'), K_A^t))$, unless another case applies to $\mathcal{E}$ and yields a smaller value;
- if $\mathcal{E} \notin K_A^t$ and $\mathcal{E} = ((\phi^C \vee \psi^{C'})^A, 1 - \frac{(1-p) \cdot (1-p')}{1-p+1-p'})$ where $1 - p + 1 - p' \neq 0$, then $\mathrm{age}(\mathcal{E}, K_A^t) = (t, n + m + 1)$ where $n = \mathrm{snd}(\mathrm{age}((\phi^C, p), K_A^t))$ and $m =$

snd(age$((\psi^{C'}, p'), K_A^t)$), unless another case applies to $\mathcal{E}$ and yields a smaller value;

- if $\mathcal{E} \notin K_A^t$, $\mathcal{E} = (\phi^C, 1 - \frac{(1-p')\cdot(1-p)}{(1-p')-(1-p)})$ where $(1-p') - (1-p) > 0$, $(1-p') \cdot (1-p) \leq (1-p') - (1-p)$ and either $C = C' = C'' = A$ or $C \neq A$, then age$(\mathcal{E}, K_A^t) = (t, n+m+1)$ where $n =$ snd(age$(((\phi^C \vee \psi^{C'})^{C''}, p), K_A^t)$) and $m =$ snd (age$((\psi^{C'}, p'), K_A^t)$), unless another case applies to $\mathcal{E}$ and yields a smaller value;
- if $\mathcal{E} \notin K_A^t$, $\mathcal{E} = (\psi^{C'}, 1 - \frac{(1-p')\cdot(1-p)}{(1-p')-(1-p)})$ where $(1-p') - (1-p) > 0$, $(1-p') \cdot (1-p) \leq (1-p') - (1-p)$ and either $C = C' = C'' = A$ or $C' \neq A$, then age$(\mathcal{E}, K_A^t) = (t, n+m+1)$ where $n =$ snd(age$(((\phi^C \vee \psi^{C'})^{C''}, p), K_A^t)$) and $m =$ snd(age$((\phi^C, p'), K_A^t)$), unless another case applies to $\mathcal{E}$ and yields a smaller value;
- if $\mathcal{E} \notin K_A^t, \mathcal{E} = ((\phi^C \to \psi^{C'})^A, 1 - \frac{p\cdot(1-p')}{p+1-p'})$ and $p + 1 - p' \neq 0$, then age$(\mathcal{E}, K_A^t) = (t, n+m+1)$ where $n =$ snd(age$((\phi^C, p), K_A^t)$) and $m =$ snd(age$((\psi^{C'}, p'), K_A^t)$), unless another case applies to $\mathcal{E}$ and yields a smaller value;
- if $\mathcal{E} \notin K_A^t$, $\mathcal{E} = (\psi^{C'}, 1 - \frac{p'\cdot(1-p)}{p'-(1-p)})$ where $p' - (1-p) > 0$, $p' \cdot (1-p) \leq p' - (1-p)$ and either $C = C' = C'' = A$ or $C' \neq A$, then age$(\mathcal{E}, K_A^t) = (t, n+m+1)$ where $n =$ snd(age$(((\phi^C \to \psi^{C'})^{C''}, p), K_A^t)$) and $m =$ snd(age$((\phi^C, p'), K_A^t)$), unless another case applies to $\mathcal{E}$ and yields a smaller value.

Intuitively, the derivation will be constructed by using rules that correspond to the steps of the construction of the model (in terms of applications of the update function *Up* introduced in Definition 5.5 and of the closure operator cl introduced in Definition 5.6) required to justify that $\mathcal{E} \in \mathrm{cl}(K_A^t)$. The fact that the construction of the derivation always terminates in a finite number of steps is proved by induction on the age of $\mathcal{E}$ with respect to K_A^t. The fact that age$(\mathcal{E}, K_A^t)$ is used as inductive measure in this proof also explains why the first element of a pair which is the value of a function application age$(\mathcal{E}, K_A^t)$ is t even though t is already evident from the notation age$(\mathcal{E}, K_A^t)$. Indeed, the proof requires the existence of an order defined on the set of these pairs – that is, the set of *values* of the function applications age$(\mathcal{E}, K_A^t)$ – and t is a key element in the definition of this order.

The way we map states in a model to the corresponding sequent contexts is slightly more complicated than the way we map sequent contexts to states in the model. Indeed, the change that leads from a state in a model to its successor state can correspond to several operational rule applications that gradually modify the context of the sequents involved. The change from one profile to the next one (which actually corresponds to the change from one state to its successor) is represented, in a derivation, by the sequence of rule applications that enable us to derive the conclusion P :: *Profile* of an `I` or `DisI` rule application from the premiss P' :: *Profile* of a `R` rule application, possibly through applications of `Tr`, `DisTr`, `ATr`, `AMis`. Schematically, if a state K_A^t corresponds to the profile P' occurring in a derivation with the following form:

$$\dfrac{\dfrac{\dots \quad P' :: \textit{Profile}}{\vdots \ \texttt{ATr / AMis / Tr / DisTr}}\ \texttt{R}}{P :: \textit{Profile}}\ \texttt{I/DisI}$$

then K_A^{t+1} will correspond to the profile P. In Theorem 5.1, this issue was dealt with by handling in a special way all expressions of the form $\mathtt{R}(\ldots)$, $\mathtt{Tr}(\ldots)$ and $\mathtt{DisTr}(\ldots)$. Also in the following theorem, we treat opinions and operational formulae differently. That is, we treat opinions by a, more or less, traditional completeness statement: there exists a derivation of $\mathfrak{L}; \Gamma^A; \Delta \vdash (\phi^C, p)$ for a suitable context $\mathfrak{L}; \Gamma^A; \Delta$ in case $(\phi^C, p) \in \mathrm{cl}_A(K_A^t)$ holds. As for the operational rules that are added to the state of A at time t, they can be derived, but not with a sequent context that exactly corresponds to the semantic state K_A^t. Hence, in case $\mathcal{E}$ is an operational formula that was added to the state of A at t for the first time – formally, $\mathcal{E} \in K_A^t \setminus K_A^{t-1}$ – instead of finding complex ways of constraining the context of $\mathfrak{L}; \ldots \vdash \mathcal{E}$ and then requiring that $\mathfrak{L}; \ldots \vdash \mathcal{E}$ is derivable, we simply require that our derivation of $\mathfrak{L}; \Gamma^A; \Delta \vdash (\phi^C, p)$ contains a derivation of $\mathfrak{L}; \ldots \vdash \mathcal{E}$ for any context $\mathfrak{L}; \ldots$. In other terms, we use the fact that $\mathfrak{L}; \ldots \vdash \mathcal{E}$ has been derived a few steps above $\mathfrak{L}; \Gamma^A; \Delta \vdash (\phi^C, p)$ in order to guarantee that the context $\mathfrak{L}; \ldots$ represents an intermediate step between the context that corresponds to K_A^{t-1} and the one that corresponds to K_A^t. By doing this, we are able to handle the fact that a single update step of a state in a model corresponds to several inferential steps updating the corresponding context.

Theorem 5.2 (Completeness of the calculus): *For any model $\mathcal{M}$, state K_A^t of $\mathcal{M}$, and profile $\mathfrak{L}; \Gamma^A; \Delta$, if $(\phi^C, p) \in \mathrm{cl}_A(K_A^t)$ and $\mathfrak{L}; \Gamma^A; \Delta$ is a K_A^t-profile, then $\mathfrak{L}; \Gamma^A; \Delta \vdash (\phi^C, p)$ is derivable by a derivation that contains a derivation of a sequent of the form $\mathfrak{L}; \ldots \vdash \mathcal{E}$ for any operational formula $\mathcal{E} \in K_A^t \setminus K_A^{t-1}$.*

Proof: We prove the following, stronger statement.

For any model $\mathcal{M}$, state K_A^t of $\mathcal{M}$, and profile $\mathfrak{L}; \Gamma^A; \Delta$, the following hold:

(1) If $(\phi^C, p) \in \mathrm{cl}_A(K_A^t)$ and $\mathfrak{L}; \Gamma^A; \Delta$ is a K_A^t-profile, then $\mathfrak{L}; \Gamma^A; \Delta \vdash (\phi^C, p)$ is derivable.

(2) The derivation of $\mathfrak{L}; \Gamma^A; \Delta \vdash (\phi^C, p)$ contains a derivation of a sequent of the form $\mathfrak{L}; \cdots \vdash \mathcal{E}$ for any operational formula $\mathcal{E} \in K_A^t \setminus K_A^{t-1}$.

(3) For any $(\phi^C, p) \in K_A^t$ and for any K_A^t-profile $\mathfrak{L}; \Gamma^A; \Delta$, either

$$\Delta = \Delta', (\phi^C, p)$$

and $\mathfrak{L}; \Gamma^A; \Delta', (\phi^C, p) ::$ *Profile* is derivable or

$$\Gamma^A = \Gamma'^A, (\phi^C, p)$$

and $\mathfrak{L}; \Gamma'^A, (\phi^C, p); \Delta ::$ *Profile* is derivable.

■

The proof is by induction on $\mathrm{age}((\phi^C, p), K_A^t)$.

In the base case, $t = 0$ and $K_A^t = K_A^0$. Then $(\phi^C, p) = (a^A, p) \in K_A^0$ and K_A^0-profiles are of the form $; \Gamma^A;$. If this is the case, $; \Gamma^A; \vdash (a^C, p)$ can be proved by an application of $::$ with conclusion $; \Gamma^A; ::$ *Profile* and one of $\mathtt{StI}$. Since $t = 0$, no other cases are possible.

Suppose now that, for any expression $\mathcal{E}' = (\phi'^{D'}, q')$ and state $K_C^{t'}$ such that $\mathrm{age}(\mathcal{E}', K_C^{t'}) < (n, m)$, the statement holds. We show that it also holds for any expression $\mathcal{E} = (\phi^D, q)$ such that $\mathrm{age}(\mathcal{E}, K_A^t) = (n, m)$.

Now, either $\mathcal{E} \in K_A^t$ or $\mathcal{E} \notin K_A^t$.

- If $\mathcal{E} \notin K_A^t$, we reason by cases on the form of $\mathcal{E}$. The age of $\mathcal{E}$ is computed according to exactly one case of Definition 5.9. The application of this case to $\mathcal{E}$ univocally corresponds to an application of a logical introduction or elimination rule of the calculus. The rule application will be of one of the following forms:

$$\frac{\mathfrak{L}; \Gamma^A; \Delta \vdash \mathcal{E}'}{\mathfrak{L}; \Gamma^A; \Delta \vdash \mathcal{E}} \qquad \frac{\mathfrak{L}; \Gamma^A; \Delta \vdash \mathcal{E}' \quad \mathfrak{L}; \Gamma^A; \Delta \vdash \mathcal{E}''}{\mathfrak{L}; \Gamma^A; \Delta \vdash \mathcal{E}}$$

  where $\mathfrak{L}; \Gamma^A; \Delta$ is a K_A^t-profile and both $\mathrm{age}(\mathcal{E}', K_A^t) < \mathrm{age}(\mathcal{E}, K_A^t)$ and $\mathrm{age}(\mathcal{E}'', K_A^t) < \mathrm{age}(\mathcal{E}, K_A^t)$ hold by Definition 5.9. Hence, by inductive hypothesis, all premisses of the rule application are derivable. But this means, as desired, that the sequent $\mathfrak{L}; \Gamma^A; \Delta \vdash \mathcal{E}$ is derivable as well.
- Suppose then that $\mathcal{E} \in K_A^t$. If this is the case, by Definition 5.9, $\mathrm{age}(\mathcal{E}, K_A^t) = (t, 0)$. If $t = 0$, we simply are in the base case. Otherwise, let us consider how K_A^t has been constructed by applying the update function *Up* to K_A^{t-1} according to Definition 5.5. We reason by cases on the value of $\mathfrak{C}(t-1)$ where $\mathfrak{C}$ is the action choice function of model $\mathcal{M}$.

  (1) $\mathfrak{C}(t-1) = w(B, (\phi^C, p))$ or $\mathfrak{C}(t-1) = r(B, (\phi^C, p))$ with $B \neq A$ and

  $$K_A^t = Up(K_A^{t-1}, w(B, (\phi^C, p))) = Up(K_A^{t-1}, r(B, (\phi^C, p))) = K_A^{t-1}$$

  (Definition 5.5, case *1*).

  As for the first point of the statement, by inductive hypothesis, there is a derivation of $\mathfrak{L}_0, \Gamma^A, \Delta \vdash \mathcal{E}$ for $\mathfrak{L} = \mathfrak{L}_0^{++w(B,(\phi^C,p))}$ or $\mathfrak{L} = \mathfrak{L}_0^{++r(B,(\phi^C,p))}$. Indeed, $\mathfrak{L}_0, \Gamma^A, \Delta$ is a K_A^{t-1}-profile and $\mathrm{age}(\mathcal{E}, K_A^{t-1}) < \mathrm{age}(\mathcal{E}, K_A^t)$. By inspection of the rules of the calculus, it is clear that such a derivation must contain a derivation of $\mathfrak{L}_0, \Gamma^A, \Delta :: \mathit{Profile}$. An application of the suitable $++$ rule, then, yields the desired derivation.

  As for the second point of the statement, no formula is contained in $K_A^t \setminus K_A^{t-1}$ and hence there is nothing to prove.

  As for the third point of the statement, it is enough to apply a `StI` or `StA` rule to the conclusion of $++$.

  (2) $\mathfrak{C}(t-1) = w(A, (\phi^C, p))$ and

  $$Up(K_A^{t-1}, w(A, (\phi^C, p))) = K_A^{t-1} \cup \{\mathtt{W}(\phi^C, p)\}$$

  (Definition 5.5, case *2*).

  By inductive hypothesis, there is a derivation of $\mathfrak{L}_0, \Gamma^A, \Delta \vdash (\phi^C, p)$. Indeed, $\mathfrak{L}_0, \Gamma^A, \Delta$ is a K_A^{t-1}-profile and $\mathrm{age}((\phi^C, p), K_A^{t-1}) < \mathrm{age}((\phi^C, p), K_A^{t-1})$. Moreover, because of how K_A^t is defined, we have that $\mathcal{E} \in K_A^{t-1}$. Hence, by inductive hypothesis, $\mathfrak{L}_0, \Gamma^A, \Delta \vdash \mathcal{E}$ is derivable for any K_A^{t-1}-profile $\mathfrak{L}_0, \Gamma^A, \Delta$. We can then construct the following derivation:

$$\dfrac{\dfrac{\dfrac{\begin{array}{c}\vdots\\ \mathfrak{L}_0, \Gamma^A, \Delta \vdash (\phi^C, p) \quad p > 0.5\end{array}}{\mathfrak{L}; \Gamma^A, \Delta \vdash \mathtt{W}(\phi^C, p)}\ \mathtt{W}}{\mathfrak{L}; \Gamma^A; \Delta :: \mathit{Profile}}\ ::\mathtt{W}}{\mathfrak{L}; \Gamma^A; \Delta \vdash \mathcal{E}}\ \mathtt{StA/StI}$$

Indeed, since $\mathcal{E}$ is not an operational formula, if $\mathfrak{L}_0, \Gamma^A, \Delta \vdash \mathcal{E}$ is derivable by a $\mathtt{StI}$ or $\mathtt{StA}$ rule then so is $\mathfrak{L}; \Gamma^A; \Delta \vdash \mathcal{E}$.

This derivation shows that all three points of the statement hold.

(3) $\mathfrak{C}(t-1) = r(A, (\phi^C, p))), val^{\perp\!\!\!\perp}(K_A^{t-1}, (\phi^C, p)) \leq \varepsilon_\perp^A, p \leq \varepsilon_{int}^A$ and

$$Up(K_A^{t-1}, r(A, (\phi^C, p))) = K_A^{t-1} \cup \left\{ \mathtt{R}\left((\phi^C, p), \frac{m}{n}\right), \mathtt{Tr}(\phi^C, p), (\phi^C, p) \right\}$$

(Definition 5.5, case *(3)(a)*).

According to Definition 5.1, condition *2*, there exists a time point $t' < t$ such that $\mathfrak{C}(t') = w(A', (\phi^C, p))$ for some agent A'. Hence, by inductive hypothesis (or by $::$ and $++$ applications if $\Delta = \emptyset$), both premisses of the $\mathtt{R}$ rule application displayed below are derivable. Therefore, we can construct the following derivation in which $\mathfrak{L} = \mathfrak{L}_0^{++r(A,(\phi^C,p))}$ and $\mathcal{E}$ can either be (ϕ^C, p) or any element of Γ^A, Δ:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\vdots \qquad \vdots}{\mathfrak{L}_1; \Gamma^{A'}; \Xi \vdash \mathtt{W}(\phi^C, p) \quad \mathfrak{L}_0; \Gamma^A; \Delta :: \mathit{Profile}}}{\mathfrak{L}; \Gamma^A; \Delta \vdash \mathtt{R}((\phi^C, p), \frac{m}{n})}\,\mathtt{R} \qquad val^{\perp\!\!\!\perp}(\Gamma^A; \Delta, (\phi^C, p)) \leq \varepsilon_\perp^A}{\mathfrak{L}; \Gamma^A; \Delta \vdash \mathtt{Tr}(\phi^C, p)}\,\mathtt{Tr}}{\mathfrak{L}; \Gamma^A; \Delta, (\phi^C, p) :: \mathit{Profile}_{temp}}\,\mathtt{ATr} \qquad p \leq \varepsilon_{int}^A}{\mathfrak{L}; \Gamma^A; \Delta, (\phi^C, p) :: \mathit{Profile}}\,\mathtt{DisI}}{\mathfrak{L}; \Gamma^A; \Delta, (\phi^C, p) \vdash \mathcal{E}}\,\mathtt{StI/StA}$$

Indeed, by Definition 5.7, $\mathfrak{L} = \mathfrak{C}(0), \ldots, \mathfrak{C}(t-1)$ holds and hence $\frac{n}{m}$ will be the same number both according to Definition 5.5 and according to the definition of the $\mathtt{R}$ rule.

This shows that all three points of the statement hold.

(4) $\mathfrak{C}(t-1) = r(A, (\phi^C, p))), val^{\perp\!\!\!\perp}(K_A^{t-1}, (\phi^C, p)) \leq \varepsilon_\perp^A, p > \varepsilon_{int}^A$ and

$$Up(K_A^{t-1}, r(A, (\phi^C, p))) = K_A^{t-1} \cup \left\{ \mathtt{R}\left((\phi^C, p), \frac{m}{n}\right), \mathtt{Tr}(\phi^C, p), (\phi^C, p) \right\}$$

for any q such that $p \leq q$ (Definition 5.5, case *(3)(b)*).

This case is completely analogous to the previous one except the fact that the second version of the $\mathtt{DisI}$ rule is applied.

(5) $\mathfrak{C}(t-1) = r(A, (\phi^C, p))), val^{\perp\!\!\!\perp}(K_A^{t-1}, (\phi^C, p)) \leq \varepsilon_\perp^A, p > \varepsilon_{int}^A$ and

$$\begin{aligned} &Up(K_A^{t-1}, r(A, (\phi^C, p))) \\ &\quad = (K_A^{t-1} \setminus \{(\phi^A, q)\}) \cup \left\{ \mathtt{R}\left((\phi^C, p), \frac{m}{n}\right), \mathtt{Tr}(\phi^C, p), (\phi^C, p), (\phi^A, p) \right\} \end{aligned}$$

where $q < p$ and ϕ^A contains at least one agent name different from A (Definition 5.5, case *3(c)*).

This case is completely analogous to the previous two cases except the fact that the $\mathtt{I}$ rule is applied instead of the $\mathtt{DisI}$ rule.

(6) $\mathfrak{C}(t-1) = r(A,(\phi^C,p))), val^{\perp\!\!\!\perp}(K_A^{t-1},(\phi^C,p)) > \varepsilon_\perp^A, \frac{m}{n} \leq \varepsilon_{freq}^A$ and

$$Up(K_A^{t-1}, r(A,(\phi^C,p))) = K_A^{t-1} \cup \left\{ \mathtt{R}\left((\phi^C,p), \frac{m}{n}\right), \mathtt{DisTr}(\phi^C,p) \right\}$$

(Definition 5.5, case *3(d)*).

According to Definition 5.1, condition *2*, there exists a time point $t' < t$ such that $\mathfrak{C}(t') = w(A',(\phi^C,p))$ for some agent A'. Hence, by inductive hypothesis (or by $::$ and $++$ applications if $\Delta = \emptyset$), both premisses of the $\mathtt{R}$ rule application displayed below are derivable. Therefore, we can construct the following derivation in which $\mathfrak{L} = \mathfrak{L}_0^{++r(A,(\phi^C,p))}$ and $\mathcal{E}$ can either be (ϕ^C,p) or any element of Γ^A, Δ:

$$\dfrac{\dfrac{\dfrac{\dfrac{\vdots}{\mathfrak{L}_1;\Gamma^{A'};\Xi \vdash \mathtt{W}(\phi^C,p)} \quad \dfrac{\vdots}{\mathfrak{L}_0;\Gamma^A;\Delta :: \mathit{Profile}}}{\mathfrak{L};\Gamma^A;\Delta \vdash \mathtt{R}((\phi^C,p),\frac{m}{n})}\,\mathtt{R} \quad val^{\perp\!\!\!\perp}(\Gamma^A;\Delta,(\phi^C,p)) > \varepsilon_\perp^A \quad \frac{m}{n} \leq \varepsilon_{freq}^A}{\mathfrak{L};\Gamma^A;\Delta \vdash \mathtt{DisTr}(\phi^C,p)}\,\mathtt{DisTr}}{\mathfrak{L};\Gamma^A;\Delta :: \mathit{Profile}}\,::\mathtt{DisTr}}{\mathfrak{L};\Gamma^A;\Delta \vdash \mathcal{E}}\,\mathtt{StI/StA}$$

Indeed, by Definition 5.7, $\mathfrak{L} = \mathfrak{C}(0),\ldots,\mathfrak{C}(t-1)$ holds and hence $\frac{n}{m}$ will be the same number both according to Definition 5.5 and according to the definition of the $\mathtt{R}$ rule.

This shows that all three points of the statement hold.

(7) $\mathfrak{C}(t-1) = r(A,(\phi^C,p))), val^{\perp\!\!\!\perp}(X,(\phi^C,p)) > \varepsilon_\perp^A, \frac{m}{n} > \varepsilon_{freq}^A, p \leq \varepsilon_{int}^A$ and

$$\begin{aligned} &Up(K_A^{t-1}, r(A,(\phi^C,p))) \\ &\quad = (K_A^{t-1} \setminus (K_A^{t-1} \perp\!\!\!\perp_{\varepsilon_\perp^A} (\phi^C,p))) \cup \left\{ \mathtt{R}((\phi^C,p),\frac{m}{n}),(\phi^C,p) \right\} \end{aligned}$$

(Definition 5.5, case *3(e)*).

According to Definition 5.1, condition *2*, there exists a time point $t' < t$ such that $\mathfrak{C}(t') = w(A',(\phi^C,p))$ for some agent A'. Hence, by inductive hypothesis (or by $::$ and $++$ applications if $\Delta = \emptyset$), both premisses of the $\mathtt{R}$ rule application displayed below are derivable. Therefore, we can construct the following derivation in which $\mathfrak{L} = \mathfrak{L}_0^{++r(A,(\phi^C,p))}$ and $\mathcal{E}$ can either be (ϕ^C,p) or any element of Γ^A, Δ:

$$\dfrac{\dfrac{\dfrac{\dfrac{\vdots}{\mathfrak{L}_1;\Gamma^{A'};\Xi \vdash \mathtt{W}(\phi^C,p)} \quad \dfrac{\vdots}{\mathfrak{L}_0;\Gamma^A;\Delta :: \mathit{Profile}}}{\mathfrak{L};\Gamma^A;\Delta \vdash \mathtt{R}((\phi^C,p),\frac{m}{n})}\,\mathtt{R} \quad val^{\perp\!\!\!\perp}(\Gamma^A;\Delta,(\phi^C,p)) > \varepsilon_\perp^A \quad \frac{m}{n} > \varepsilon_{freq}^A}{\mathfrak{L};\Gamma'^A;\Delta',(\phi^C,p) :: \mathit{Profile}_{temp}}\,\mathtt{AMis} \quad p \leq \varepsilon_{int}^A}{\mathfrak{L};\Gamma'^A;\Delta',(\phi^C,p) :: \mathit{Profile}}\,\mathtt{DisI}}{\mathfrak{L};\Gamma'^A;\Delta',(\phi^C,p) \vdash \mathcal{E}}\,\mathtt{StA}$$

Indeed, by Definition 5.7, $\mathfrak{L} = \mathfrak{C}(0),\ldots,\mathfrak{C}(t-1)$ holds and hence $\frac{n}{m}$ will be the same number both according to Definition 5.5 and according to the definition of the $\mathtt{R}$ rule.

This shows that all three points of the statement hold.

(8) $\mathfrak{C}(t-1) = r(A, (\phi^C, p))), val^{\perp\!\!\!\perp}(K_A^{t-1}, (\phi^C, p)) > \varepsilon_\perp^A, \frac{m}{n} > \varepsilon_{freq}^A, p > \varepsilon_{int}^A, (\phi^A, q) \in K_A^{t-1}$ with $p \leq q$ and

$$Up(K_A^{t-1}, r(A, (\phi^C, p))) = (K_A^{t-1} \setminus (K_A^{t-1} \perp\!\!\!\perp_{\varepsilon_\perp^A} (\phi^C, p))) \cup \left\{ \mathrm{R}\left((\phi^C, p), \frac{m}{n}\right), (\phi^C, p) \right\}$$

(Definition 5.5, case *3(g)*).

This case is completely analogous to the previous one except the fact that the second version of the `DisI` rule is applied instead of its first version.

(9) $\mathfrak{C}(t-1) = r(A, (\phi^C, p))), val^{\perp\!\!\!\perp}(K_A^{t-1}, (\phi^C, p)) > \varepsilon_\perp^A, \frac{m}{n} > \varepsilon_{freq}^A, p > \varepsilon_{int}^A, (\phi^A, q) \notin K_A^{t-1}$ for any q, and

$$Up(K_A^{t-1}, r(A, (\phi^C, p))) = (K_A^{t-1} \setminus (K_A^{t-1} \perp\!\!\!\perp_{\varepsilon_\perp^A} (\phi^C, p))) \cup \left\{ \mathrm{R}\left((\phi^C, p), \frac{m}{n}\right), (\phi^C, p), (\phi^A, p) \right\}$$

(Definition 5.5, case *3(f)*).

This case is completely analogous to the previous two cases except the fact that the first version of the `I` rule is applied instead of the `DisI` rule.

(10) $\mathfrak{C}(t-1) = r(A, (\phi^C, p))), val^{\perp\!\!\!\perp}(K_A^{t-1}, (\phi^C, p)) > \varepsilon_\perp^A, \frac{m}{n} > \varepsilon_{freq}^A, (\phi^A, q) \in K_A^{t-1}$, $p > \varepsilon_{int}^A$ and

$$\begin{aligned} Up(K_A^{t-1}, r(A, (\phi^C, p))) &= ((K_A^{t-1} \setminus (K_A^{t-1} \perp\!\!\!\perp_{\varepsilon_\perp^A} (\phi^C, p))) \setminus \{(\phi^A, q)\}) \\ &\quad \cup \left\{ \mathrm{R}\left((\phi^C, p), \frac{m}{n}\right), (\phi^C, p), (\phi^A, p) \right\} \end{aligned}$$

where $q < p$ and ϕ^A contains at least one agent name different from A (Definition 5.5, case *3(h)*).

This case is completely analogous to the previous one except the fact that the second version of the `I` rule is applied instead of its first version.

6. Semantical examples

We now show an example of a model formalising a complex interaction including the one presented in Example 3.1. The simple exchange between Amira and Balthazar was actually taken from a real public debate among medical experts that happened during the SARS-CoV-2 pandemic. In particular we refer to the dataset created for (Primiero et al., 2023), with 90 articles selected from 12 different newspapers reporting the debate among Italian medical experts on SARS-CoV-2. The list of articles and related metadata can be found at https://docs.google.com/spreadsheets/d/1txVJsm0y8AkjlfFj1E9EOwVP78VUY3f5yk30U04shII. In the present context, we refer to a subset of the original dataset, related

Table 1. Summary of positions and thresholds for each of the medical experts considered.

Agent	Public declaration	Statement	$\varepsilon_{\perp}$	ε_{int}	ε_{freq}
A	'The virus clinically no longer exists'	$(\neg\phi, 0.9)$	0.6	0.7	0.4
B	'What we see is just the tip of the iceberg'	$(\phi, 0.8)$	0.5	0.8	0.8
C	'The enemy seems less aggressive, but it takes time to have certain results'	$(\neg\phi, 0.6)$	0.3	1	0.6
D	'Covid could be different, it no longer has the same firepower'	$(\neg\phi, 0.7)$	0.8	0.8	0.6

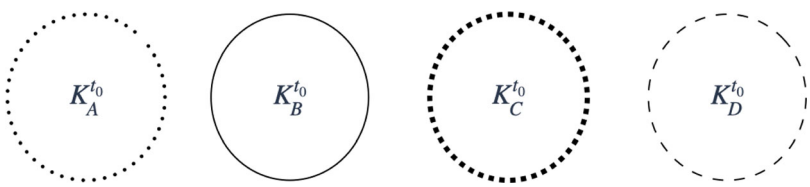
to the period 06.03.20–14.07.20, where Italian medical experts where exchanging their opinions concerning the statement

$$\phi = \text{'the situation concerning SARS-CoV-2 is critical'}.$$

For the purpose of this example we select 4 agents and, as shown in Table 1, we associate to each of them a value *p* for their conviction on ϕ, or its negation, determined on the basis of their public declaration as reported in the selected articles. In particular: for agent *A*, who declared 'it no longer exists', we interpret it as a very strong conviction, also traceable in *B*'s statement 'is just the tip of the iceberg', albeit to a slightly less degree; on the other hand, the use of the term 'seems' led us to lower the value of *C*'s conviction, while for *D*, despite the conditional form used, the subsequent statement leads to a balance of the value around 0.7. Moreover, we give a value to each of the thresholds involved in the evaluation procedure, choosing values that best fits the real trend of the debate.

Actions performed by the agents

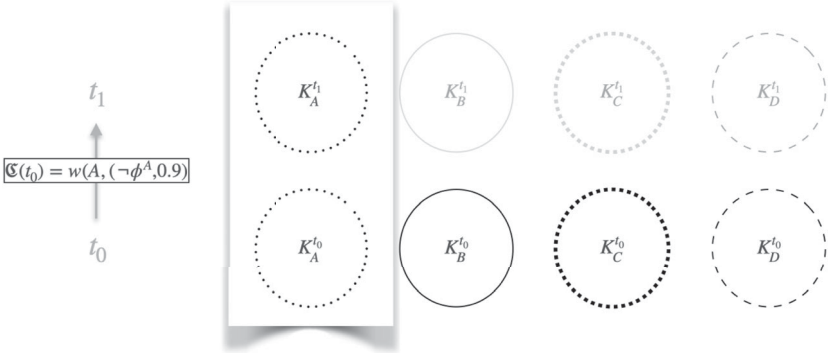
The four medical experts (A, B, C, D) are discussing the topic ϕ. Each agent is represented by an opinion set $K_S^{t_i}$:



The model's evolution depends on the actions that the agents choose to perform as extracted from the dataset. Agent *A* expresses her opinion. Formally this translates into the following expression,

$$\mathfrak{C}(t_0) = w(A, (\neg\phi^A, 0.9))$$

where the function $\mathfrak{C}$, applied to the time instant t_0, indicates the name of the action and that of the agent who chooses to perform it and brings the model to time t_1. In this case, agent *A* chooses to write $\neg\phi$ with a value of 0.9.



This choice involves a modification of the set of opinions $K_A^{t_0}$ and the consequent transition to the set of opinions $K_A^{t_1}$. On the other hand, for the other agents the set of opinions at time t_1 remains the same as at time t_0. Each of the other sets will only change when the relative agent will take an action.

The action that brings us to time t_2 is the following:

$$\mathfrak{C}(t_1) = w(B, (\phi^B, 0.8))$$

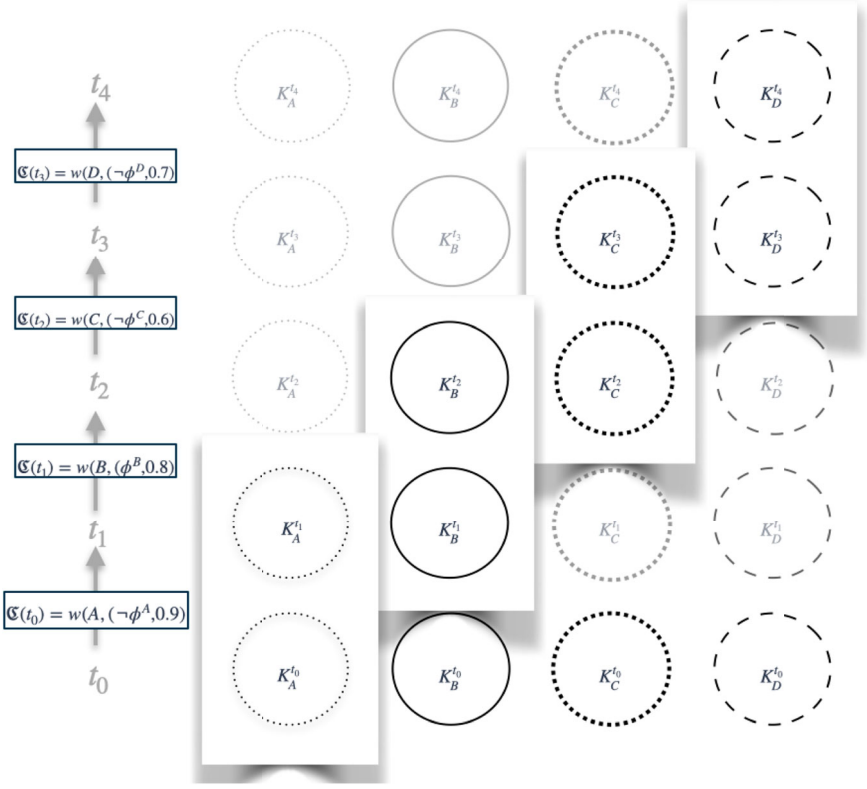
where the function application $\mathfrak{C}(t_1)$ indicates that agent B chooses to write ϕ with a value of 0.8.

This choice involves a modification of the set of opinions $K_B^{t_1}$ and the consequent transition to the set of opinions $K_B^{t_2}$ while, again, for the other agents the set of opinions doesn't change. Similarly, the actions

$$\mathfrak{C}(t_2) = w(C, (\neg\phi^C, 0.6))$$

$$\mathfrak{C}(t_3) = w(D, (\neg\phi^D, 0.7)$$

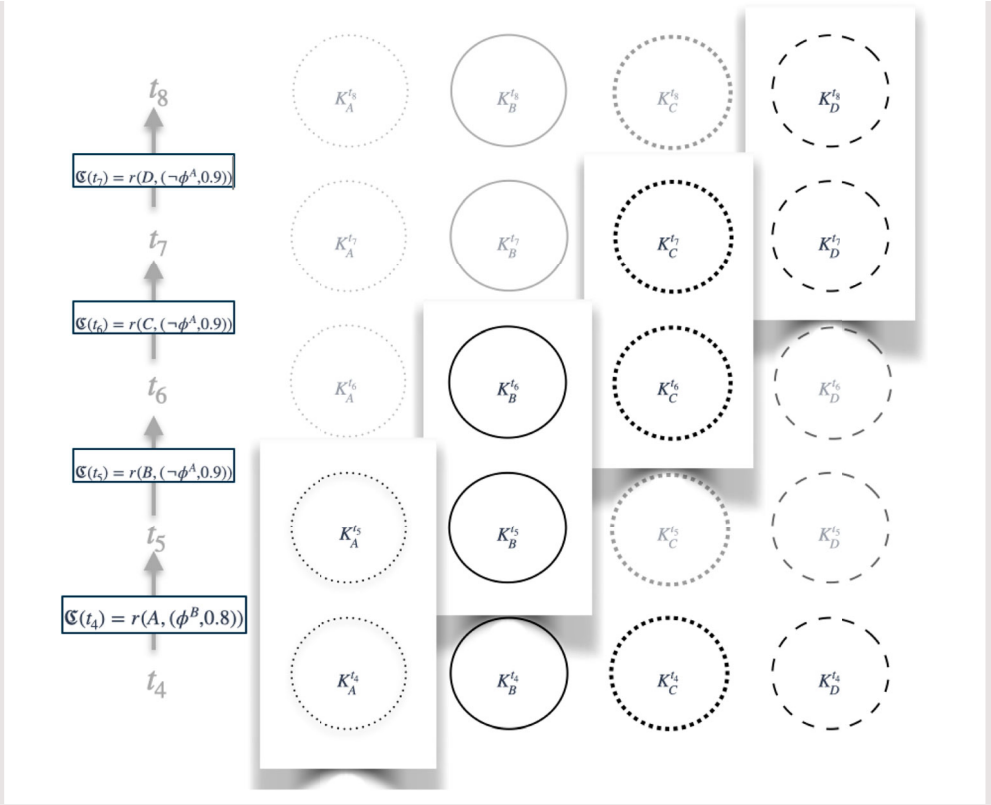
performed by agents C and D, respectively at time t_2 and t_3 make the states evolve:



Thus, at state 4, each agent has expressed her opinion by writing it. Subsequently, with actions

$$\mathfrak{C}(t_4) = r(A, (\phi^B, 0.8))$$
$$\mathfrak{C}(t_5) = r(B, (\neg\phi^A, 0.9))$$
$$\mathfrak{C}(t_6) = r(C, (\neg\phi^A, 0.9))$$
$$\mathfrak{C}(t_7) = r(D, (\neg\phi^A, 0.9))$$

the agents choose to read the opinions written by others, making the states evolve as follows:



In the following, we show how the choices taken by the agents in the example just analysed affect *K*, i.e. the set of expressions representing the opinion state of agents. In particular, we show how the update function *Up* modifies the set *K* at each time.

Operation of Update with respect to K over time

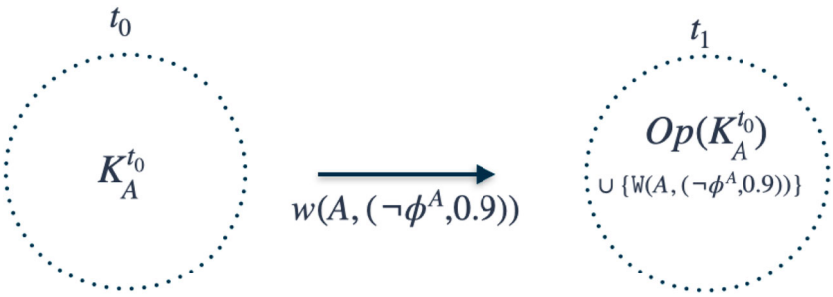
Consider the first action of the model that brings us to time t_1

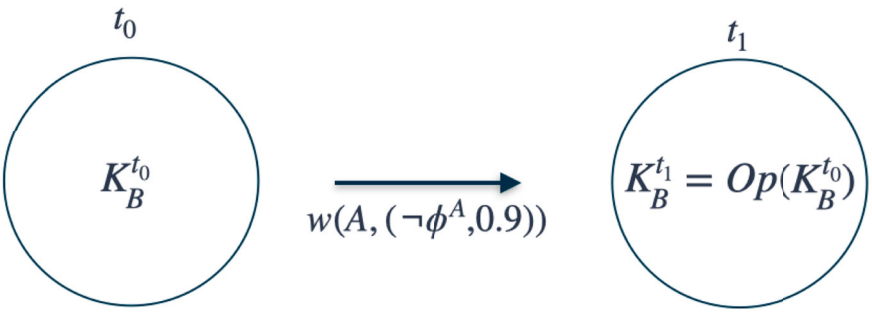
$$\mathfrak{C}(t_0) = w(A, (\neg\phi^A, 0.9))$$

performed by agent *A*. According to Definition 5.5, the behaviour of the update function from time t_0 to time t_1 is the following for each agent:

$$K_A^{t_1} = Up(K_A^{t_0}, \mathfrak{C}(t_0)) = Up(K_A^{t_0}, w(A, (\neg\phi^A, 0.9)) = Op(K_A^{t_0}) \cup \{\mathbb{W}(A, (\neg\phi^A, 0.9))\}$$
$$K_B^{t_1} = Up(K_B^{t_0}, \mathfrak{C}(t_0)) = Up(K_B^{t_0}, w(A, (\neg\phi^A, 0.9))) = Op(K_B^{t_0})$$
$$K_C^{t_1} = Up(K_C^{t_0}, \mathfrak{C}(t_0)) = Up(K_C^{t_0}, w(A, (\neg\phi^A, 0.9))) = Op(K_C^{t_0})$$
$$K_D^{t_1} = Up(K_D^{t_0}, \mathfrak{C}(t_0)) = Up(K_D^{t_0}, w(A, (\neg\phi^A, 0.9))) = Op(K_D^{t_0})$$

In particular, the Update function extends the opinion set of A, the agent who performed the action, while leaving the others unchanged, e.g. for agent B:





Similarly, the update function will respectively modify the set of B in the transition to t_2, that of C in the transition to t_3 and that of D in the transition to t_4, leaving the others always unchanged.

Hence, at every instant in time the opinion sets of the agents who do not act remain unchanged, and for each state let us focus on the functioning of the update function on the set of the agent who reads the information. For the purposes of the example, in what follows we consider that the agents' opinion sets contain only the formula that each of them wrote so far and nothing else.

The action bringing us from t_4 to t_5 is

$$\mathfrak{C}(t_4) = r(A, (\phi^B, 0.8))$$

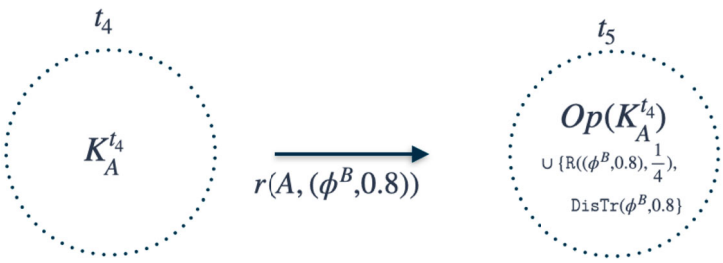
performed by agent A. Informally, we can say that agent A reads the message and rejects it: it is an opinion that contrasts with her own, and which moreover has been produced by other agents too few times to be influenced by it. Technically this is reflected in the following way. Given that:

- $val^{\perp\!\!\!\perp}(K_A^t, (\phi^B, 0.8)) > \varepsilon_{\perp}^A : 0.72 > 0.6$
- $\frac{m}{n} \leq \varepsilon_{freq}^A : 0.25 \leq 0.4$

and following Definition 5.5, the behaviour of the update function from time t_4 to time t_5, with respect to K_A is

$$K_A^{t_5} = Up(K_A^{t_4}, \mathfrak{C}(t_4)) = Up(K_A^{t_4}, r(A, (\phi^B, 0.8)) =$$
$$= Op(K_A^{t_4}) \cup \left\{ \mathtt{R}\left((\phi^B, 0.8), \tfrac{1}{4}\right), \mathtt{DisTr}(\phi^B, 0.8) \right\}.$$

The opinion set at time t_4 is expanded at time t_5 through the addition of formulae $\mathtt{R}((\phi^B, 0.8), \frac{1}{4})$ and $\mathtt{DisTr}(\phi^B, 0.8)$:



The action bringing us from t_5 to t_6 is

$$\mathfrak{C}(t_5) = r(B, (\neg\phi^A, 0.9))$$

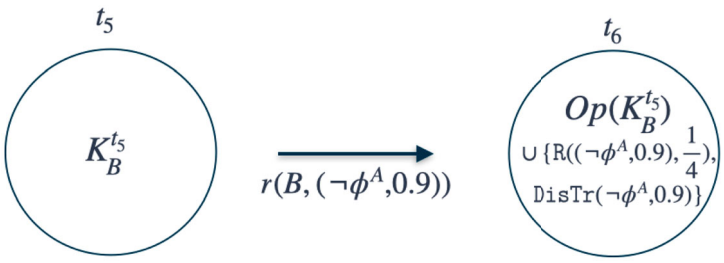
performed by agent B. Informally, we can say, again, that agent B reads the message and rejects it: it is an opinion that contrasts with her own, which moreover has been produced too few times to be influenced by it. Technically this is reflected in the following way. Given that:

- $val^{\perp\!\!\!\perp}(K_B^t, (\neg\phi^A, 0.9)) > \varepsilon_\perp^B : 0.72 > 0.5$
- $\frac{m}{n} \leq \varepsilon_{freq}^A : 0.25 \leq 0.8$

and following Definition 5.5, the behaviour of the update function from time t_5 to time t_6, with respect to K_B is

$$K_B^{t_6} = Up(K_B^{t_5}, \mathfrak{C}(t_5)) = Up(K_B^{t_5}, r(B, (\neg\phi^A, 0.9)) =$$
$$= Op(K_B^{t_5}) \cup \left\{ \mathtt{R}\left((\neg\phi^A, 0.9), \tfrac{1}{4}\right), \mathtt{DisTr}(\phi^A, 0.9) \right\}.$$

The opinion set at time t_5 is expanded at time t_6 through the addition of formulae $\mathtt{R}((\neg\phi^A, 0.9), \frac{1}{4})$ and $\mathtt{DisTr}(\phi^A, 0.9)$:



The action bringing us from t_6 to t_7 is

$$\mathfrak{C}(t_6) = r(C, (\neg\phi^A, 0.9))$$

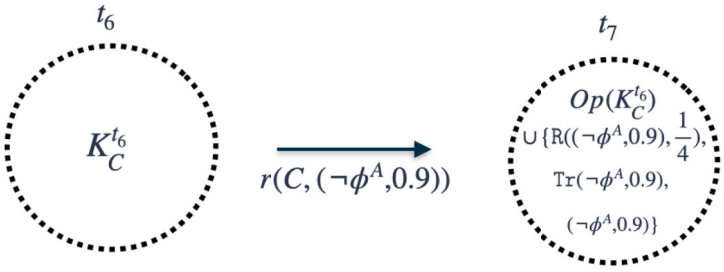
performed by agent C. Informally, agent C reads the message and accepts it, as it is in line with her previous opinions. The conviction with which A writes the message is not, however, strong enough for C to decide to make it her own opinion: agent C is therefore limited to accept the fact that A supports ϕ. Technically this is reflected in the following way. Given that:

- $val^{\perp\!\!\!\perp}(K_C^t, (\neg\phi^A, 0.9)) \leq \varepsilon_\perp^C : 0 \leq 0.3$
- $p \leq \varepsilon_{int}^C : 0.9 \leq 1$

and following Definition 5.5, the behaviour of the update function from time t_6 to time t_7, with respect to K_C is

$$\begin{aligned} K_C^{t_7} &= Up(K_C^{t_6}, \mathfrak{C}(t_6)) = Up(K_C^{t_6}, r(C, (\neg\phi^A, 0.9)) = \\ &= Op(K_C^{t_6}) \cup \left\{\mathtt{R}\left((\neg\phi^A, 0.9), \tfrac{1}{4}\right), \mathtt{Tr}(\neg\phi^A, 0.9), (\neg\phi^A, 0.9)\right\} \end{aligned}$$

The opinion set at time t_6 is expanded at time t_7 through the addition of formulae $\mathtt{R}((\neg\phi^A, 0.9), \frac{1}{4})$, $\mathtt{Tr}(\phi^A, 0.9)$ and $(\neg\phi^A, 0.9)$:



The action bringing us from t_7 to t_8 is

$$\mathfrak{C}(t_7) = r(D, (\neg\phi^A, 0.9))$$

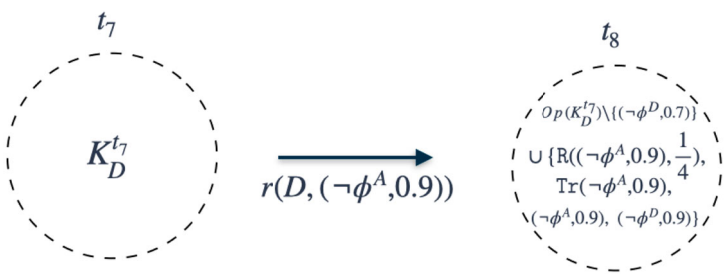
performed by agent D. Agent D reads the message and accepts it, as it is in line with her previous opinions. This time, however, the conviction with which A writes the message is strong enough for D to decide to make it her own opinion: therefore, agent D not only accepts the fact that A supports ϕ, but also adds her signature to it (subscribes to it). Technically this is reflected in the following way. Given that:

- $val^{\perp\!\!\!\perp}(K_D^t, (\neg\phi^A, 0.9)) \leq \varepsilon_{\perp}^D : 0 \leq 0.8$
- $q < p > \varepsilon_{int}^D : 0.7 \leq 0.9 > 0.8$

and following Definition 5.5, the behaviour of the update function from time t_7 to time t_8, with respect to K_D is

$$\begin{aligned} K_D^{t_8} &= Up(K_D^{t_7}, \mathfrak{C}(t_7)) = Up(K_D^{t_7}, r(D, (\neg\phi^A, 0.9)) = \\ &= Op(K_D^{t_7}) \setminus \{(\neg\phi^D, 0.7)\}) \\ &\quad \cup \left\{ \mathtt{R}\left((\neg\phi^A, 0.9), \tfrac{1}{4}\right), \mathtt{Tr}(\neg\phi^A, 0.9), (\neg\phi^A, 0.9), (\neg\phi^D, 0.9) \right\} \end{aligned}$$

The opinion set at time t_7 is expanded at time t_8 through the addition of formulae $\mathtt{R}((\neg\phi^A, 0.9), \frac{1}{4})$, $\mathtt{Tr}(\phi^A, 0.9)$, $(\neg\phi^A, 0.9)$ and $(\neg\phi^D, 0.9)$:



7. Future work

We have presented the logic $(\mathtt{un})\mathtt{SecureND}_{grad}$ to model both proof-theoretically and semantically exchanges of information affected by uncertainty and whose acceptance is regulated by trust and negative trust properties on communication relations. The introduction of the system $(\mathtt{un})\mathtt{SecureND}_{grad}$ opens several future work directions.

First of all, a natural extension of the present work concerns the communication strategies of agents. Indeed, in $(\mathtt{un})\mathtt{SecureND}_{grad}$ the choice of an agent whether to divulge one of her opinions or to consider an opinion which is accessible to her is completely left undetermined. In other words, the communication attitudes and goals of an agent are simply determined by modelling choices. These choices, in turn, are not made formally explicit or internalised in the system. A very interesting

possibility – and a very promising one as far as applications are concerned – is the implementation of mechanisms to explicitly determine agents' communication attitudes. In other terms, one could specify general rules that constrain the temporal evolution of models (semantically) and the applicability of certain rules (in the proof system). The introduction of mechanisms of this kind would enable us to formally define behavioural guidelines determining the choices of agents concerning what to read and what to write at any point in time. For instance, an agent devoted to propagandistic aims could always stick to her opinions without considering those divulged by other agents, and could divulge only opinions with high conviction degrees. A balanced agent, on the other hand, could spend an equal amount of time to divulge her opinions and to consider the opinions that other agents divulged. Such an extension of the system would make it an excellent tool for computational simulations concerning the behaviour and evolution of complex communication environments such as social media and digital environments in which communications happen in a structured way, e.g. commercial websites in which users can divulge their preferences and opinions on products and topics. A strategy in this direction was implemented by rules in Prandi and Primiero (2022) for $\texttt{(un)SecureND}$.

Along the same lines, it would be extremely interesting to implement in $\texttt{(un)SecureND}_{grad}$ a mechanism that models how much influence an agent has on other agents. This could be simply done by exploiting the value $\frac{m}{n}$ occurring in expressions of the form $\texttt{R}((\phi^B, p), \frac{m}{n})$. This expression, indeed, indicates that an agent A is considering (*reading*) the opinion (ϕ^B, p). The greater the value $\frac{m}{n}$, the easier for A to accept the opinion (ϕ^B, p) even though it partially conflicts with A's present convictions. Now, a very promising way to model the influence agents have on each other is to attribute a weight to each ordered pair of agents and use it in the computation of $\frac{m}{n}$ as follows: when agent A considers an opinion (ϕ^B, p), instead of using the value $\frac{m}{n}$ we use $\frac{m}{n} \cdot w$, where w is the weight of the influence that B has on A. Thus it would be possible to capture different kinds of relationships between agents and thus also a greater variety of types of agents.

Another way to model the mutual influence between agents is to define a trustworthiness ranking among them to establish when to accept or reject incoming contradictory information. An algorithm computing such ranking based on three parameters (knowledgeability, reputation and popularity) has been presented in Ceolin and Primiero (2019), Ceolin et al. (2021) and Primiero et al. (2023) using as underlying logic $\texttt{(un)SecureND}$ from Primiero (2020). Future work shall focus on refining this trustworthiness ranking algorithm introducing the account of uncertainty modelled by $\texttt{(un)SecureND}_{grad}$.

Finally, there are several phenomena of interest that might constitute perfect case studies to be analysed by $\texttt{(un)SecureND}_{grad}$. As mentioned in Section 2, the system could be employed as an enriched version of a communication network aimed at studying the way opinions and ideas spread. Particular phenomena that could be analysed are, for instance, the effects of filter bubbles in social networks, the effectiveness of propaganda strategies, the way different initial conditions influence the evolution of a social network as far as opinion diffusion and acceptance is concerned, and related equilibria issues.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

All authors thankfully acknowledge the support of the Italian Ministry of University and Research through the projects PRIN2020 n. 2020SSKZ7R (BRIO–Bias, Risk and Opacity in AI), PRIN2022 n.20223E8Y4X (SMARTEST – Simulation of Probabilistic Systems for the Age of the Digital Twin) and 'Departments of Excellence 2023–2027' awarded to the Department of Philosophy 'Piero Martinetti'.

ORCID

Francesca Doneda http://orcid.org/0009-0008-4682-225X
Francesco A. Genco https://orcid.org/0000-0001-7415-5839
Giuseppe Primiero https://orcid.org/0000-0003-3264-7100

References

Aldini, A., Curzi, G., Graziani, P., & Tagliaferri, M. (2021). Trust evidence logic. In J. Vejnarová & N. Wilson (Eds.), *Symbolic and quantitative approaches to reasoning with uncertainty. ECSQARU 2021. Lecture notes in computer science* (Vol. 12897). Springer.

Allport, F. H., & Lepkin, M. (1945). Wartime rumors of waste and special privilege: Why some people believe them. *The Journal of Abnormal and Social Psychology*, *40*(1), 3–36. https://doi.org/10.1037/h0058110

Austin, J. L. (1975). *How to do things with words*. Harvard University Press.

Begg, I. M., Anas, A., & Farinacci, S. (1992). Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General*, *121*(4), 446–458. https://doi.org/10.1037/0096-3445.121.4.446

Bell, D., & LaPadula, L. (1973). *Secure computer systems: Mathematical foundations* (Technical Report MTR-2547 1).

Bentahar, J., Drawel, N., & Sadiki, A. (2022). Quantitative group trust: A two-stage verification approach. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022)*, Online, May 9–13, 2022 (pp. 100–108). International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS).

Bertino, E., Bonatti, P. A., & Ferrari, E. (2001, August). TRBAC: A temporal role-based access control model. *ACM Transactions on Information and System Security*, *4*(3), 191–233. https://doi.org/10.1145/501978.501979

Castelfranchi, C., & Falcone, R. (2010). *Trust theory: A socio-cognitive and computational model*. John Wiley & Sons Ltd.

Ceolin, D., Doneda, F., & Primiero, G. (2021). Computable trustworthiness ranking of medical experts in Italy during the SARS-CoV-19 pandemic. In *Proceedings of the Conference on Information Technology for Social Good (GoodIT '21)* (pp. 271–276). Association for Computing Machinery. https://doi.org/10.1145/3462203.3475907.

Ceolin, D., & Primiero, G. (2019). A granular approach to source trustworthiness for negative trust assessment. In *Proceedings of the IFIP International Conference on Trust Management* (pp. 108–121). Springer International Publishing. https://doi.org/10.1007/978-3-030-33716-2_9

Chakraborty, S., & Ray, I. (2006). TrustBAC: Integrating trust relationships into the RBAC model for access control in open systems. In *Proceedings of the Eleventh ACM Symposium on Access Control Models and Technologies* (pp. 49–58). Association for Computing Machinery. https://doi.org/10.1145/1133058.1133067.

Chandran, S. M., & Joshi, J. B. D. (2005). LoT-RBAC: A location and time-based RBAC model. In A. H. H. Ngu, M. Kitsuregawa, E. J. Neuhold, J Y. Chungand, & Q. Z. Sheng (Eds.), *Web information systems engineering – WISE 2005* (pp. 361–375). Springer.

Dastani, M., Herzig, A., Hulstijn, J., & Van Der Torre, L. (2005). Inferring trust. In *Proceedings of the 6th International Workshop on Computational Logic in Multi-Agent Systems (CLIMA)* (pp. 144–160). Springer.

Demolombe, R. (2004). Reasoning about trust: A formal logical framework. In C. Jensen, S. Posladand, & T. Dimitrakos (Eds.), *Trust Management* (pp. 291–303). Springer.

Demolombe, R. (2009). Graded trust. In *AAMAS Trust* (pp. 1–12). International Foundation for Autonomous Agents and MultiAgent Systems (IFAAMAS).

Demolombe, R. (2011). Transitivity and propagation of trust in information sources: An analysis in modal logic. In J. Leite, P. Torroni, T. Agotnes, G. Boellaand, & L. van der Torre (Eds.), *Computational logic in multi-agent systems* (pp. 13–28). Springer.

Demolombe, R. (2017). Reasoning about trust and aboutness in the context of communication. *Journal of Applied Non-Classical Logics*, *27*(3–4), 292–303. https://doi.org/10.1080/11663081.2017.1420316

Demolombe, R., & Liau, C. J. (2000). A logic of graded trust and belief fusion. In *Proceedings of the workshop on Deception, Fraud, and Trust in Agent Societies held during the Autonomous Agents Conference: Trust in Cyber-societies, Integrating the Human and Artificial Perspectives* (pp. 13–25). Springer-Verlag.

Dubois, D., & Prade, H. (2013). Belief revision with uncertain inputs in the possibilistic setting. *arXiv preprint arXiv:1302.3575*.

Falcone, R., Pezzulo, G., & Castelfranchi, C. (2003). A fuzzy approach to a belief-based trust computation. In R. Falcone, S. Barber, L. Korba, & M. Singh (Eds.), *Trust, reputation, and security: Theories and practice. TRUST 2002. Lecture notes in computer science* (Vol. 2631). Springer. https://doi.org/10.1007/3-540-36609-1_7

Ferraiolo, D. F., Sandhu, R., Gavrila, S., Kuhn, D. R., & Chandramouli, R. (2001, August). Proposed NIST standard for role-based access control. *ACM Transactions on Information and System Security*, *4*(3), 224–274. https://doi.org/10.1145/501978.501980

Genco, F. A. (2024). A logic of knowledge and justifications, with an application to computational trust. *Studia Logica*. https://doi.org/10.1007/s11225-024-10165-7

Granovetter, M. (1978). Threshold models of collective behavior. *American Journal of Sociology*, *83*(6), 1420–1443. https://doi.org/10.1086/226707

Herzig, A., Lorini, E., Hübner, J. F., & Vercouter, L. (2010). A logic of trust and reputation. *Logic Journal of the IGPL*, *18*(1), 214–244. https://doi.org/10.1093/jigpal/jzp077

Holdcroft, D. (1979). Assertive acts, context, and evidence. *Journal of Pragmatics*, *3*(5), 473–488. https://doi.org/10.1016/0378-2166(79)90023-7

Hunter, A. (2022). Belief revision with dishonest reports. In *AI 2022: Advances in Artificial Intelligence: 35th Australasian Joint Conference, AI 2022*, Perth, WA, Australia, December 5–8, 2022, Proceedings (pp. 397–410). Springer-Verlag. https://doi.org/10.1007/978-3-031-22695-3_28

Kephart, J., & White, S. (1991). Directed-graph epidemiological models of computer viruses. In *Proceedings of the 1991 IEEE Computer Society Symposium on Research in Security and Privacy* (pp. 343–359). IEEE.

Kermack, W., McKendrick, A., & Walker, G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, *115*(772), 700–721.

Kohlas, R., Jonczy, J., & Haenni, R. (2008). A trust evaluation method based on logic and probability theory. In Y. Karabulut, J. Mitchell, P. Herrmannand, & C. D. Jensen (Eds.), *Trust management II* (pp. 17–32). Springer.

Leturc, C., & Bonnet, G. (2018). A normal modal logic for trust in the sincerity. In *17th International Conference on Autonomous Agents and Multiagent Systems*. International Foundation for Autonomous Agents and MultiAgent Systems (IFAAMAS).

Liau, C. J. (2003). Belief, information acquisition, and trust in multi-agent systems—A modal logic formulation. *Artificial Intelligence*, *149*(1), 31–60. https://doi.org/10.1016/S0004-3702(03)00063-8

Lorini, E., & Demolombe, R. (2009). From trust in information sources to trust in communication systems: An analysis in modal logic. In J. J. C. Meyer, & J. Broersen (Eds.), *Knowledge representation for agents and multi-agent systems* (pp. 81–98). Springer.

Lorini, E., Jiang, G., & Perrussel, L. (2014). Trust-based belief change. In *ECAI 2014* (pp. 549–554). IOS Press.

Mui, L. (2002). *Computational models of trust and reputation: Agents, evolutionary games, and social networks* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Muller, G., & Vercouter, L. (2003). Decentralized monitoring of agent communications with a reputation model. In *Trusting Agents for Trusting Electronic Societies, Theory and Applications in HCI and E-Commerce. Lecture Notes in Computer Science.* (pp. 144–161). Springer.

Namatame, A., & Chen, S. H. (2016). *Agent-based modeling and network dynamics*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198708285.001.0001

Oleshchuk, V. (2012). Trust-aware RBAC. In I. Kotenko, & V. Skormin (Eds.), *Computer network security* (pp. 97–107). Springer.

Pennycook, G., & Rand, D. G. (2020). Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of Personality*, *88*(2), 185–200. https://doi.org/10.1111/jopy.v88.2

Prandi, L., & Primiero, G. (2020). Effects of misinformation diffusion during a pandemic. *Applied Network Science*, *5*(1), 82. https://doi.org/10.1007/s41109-020-00327-6

Prandi, L., & Primiero, G. (2022). A logic for biassed information diffusion by paranoid agents in social networks. *Journal of Logic and Computation*, *32*(6), 1292–1315. https://doi.org/10.1093/logcom/exac020

Prawitz, D. (1971). Ideas and results in proof theory. In *Proceedings of the Second Scandinavian Logic Symposium* (Vol. 63, pp. 235–307). Elsevier.

Prawitz, D. (2006). *Natural deduction: A proof-theoretical study*. Courier Dover Publications.

Primiero, G. (2016). A calculus for distrust and mistrust. In S. M. Habib, J. Vassileva, S. Mauwand, & M. Mühlhäuser (Eds.), *Trust Management X – 10th IFIP WG 11.11 International Conference, IFIPTM 2016, Darmstadt, Germany, July 18–22, 2016, Proceedings* (Vol. 473, pp. 183–190). Springer. https://doi.org/10.1007/978-3-319-41354-9_15

Primiero, G. (2020). A logic of negative trust. *Journal of Applied Non-Classical Logics*, *30*(3), 193–222. https://doi.org/11663081.2020.1789404

Primiero, G., Bottone, M., Raimondi, F., & Tagliabue, J. (2016). Contradictory information flow in networks with trust and distrust. In H. Cherifi, S. Gaito, W. Quattrociocchi- and A. Sala (Eds.), *Complex Networks & Their Applications V – Proceedings of the 5th International Workshop on Complex Networks and their Applications (COMPLEX NETWORKS 2016), Milan, Italy, November 30 – December 2, 2016* (Vol. 693, pp. 361–372). Springer. https://doi.org/10.1007/978-3-319-50901-3_29

Primiero, G., Ceolin, D., & Doneda, F. (2023). A computational model for assessing experts' trustworthiness. *Journal of Experimental & Theoretical Artificial Intelligence*, *37*(1), 131–162. https://doi.org/10.1080/0952813X.2023.2183272

Primiero, G., & Kosolosky, L. (2016). The semantics of untrustworthiness. *Topoi*, *35*(1), 253–266. https://doi.org/10.1007/s11245-013-9227-2

Primiero, G., Martorana, A., & Tagliabue, J. (2018). Simulation of a trust and reputation based mitigation protocol for a black hole style attack on VANETs. In 2018 *IEEE European Symposium on Security and Privacy Workshops, Euros&P Workshops 2018, London, United Kingdom, April 23–27, 2018* (pp. 127–135). IEEE. https://doi.org/10.1109/EuroSPW.2018.00025

Primiero, G., Raimondi, F., Bottone, M., & Tagliabue, J. (2017). Trust and distrust in contradictory information transmission. *Applied Network Science*, *2*(1), 12. https://doi.org/10.1007/s41109-017-0029-0

Primiero, G., Tuci, E., Tagliabue, J., & Ferrante, E. (2018). Swarm attack: A self-organized model to recover from malicious communication manipulation in a swarm of simple simulated agents.

In M. Dorigo, M. Birattari, C. Blum, A. L. Christensen, A. Reinaand V. Trianni (Eds.), *Swarm Intelligence – 11th International Conference, ANTS 2018, Rome, Italy, October 29–31, 2018, Proceedings* (Vol. 11172, pp. 213–224). Springer. https://doi.org/10.1007/978-3-030-00533-7_17

Ries, S., Habib, S. M., Mühlhäuser, M., & Varadharajan, V. (2011). CertainLogic: A logic for modeling trust and uncertainty. In J. M. McCune, B. Balacheff, A. Perrig, A R. Sadeghi, A. Sasseand, & Y. Beres (Eds.), *Trust and trustworthy computing* (pp. 254–261). Springer.

Rodenhäuser, B. (2014). *A matter of trust: Dynamic attitudes in epistemic logic*. University of Amsterdam.

Rousseau, D., Sitkin, S., Burt, R., & Camerer, C. (1998, July). Introduction to special topic forum: Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, *23*(3), 393–404. http://www.jstor.org/stable/259285

Schelling, T. C. (1978). *Micromotives and macrobehavior*. W. W. Norton & Company.

Sensoy, M., De Mel, G., Fokoue, A., Norman, T. J., Pan, J. Z., Tang, Y., Oren, N., Sycara, K., Kaplan, L., & Pham, T. (2013). Reasoning with uncertain information and trust. In *Proceedings of SPIE 0277-786X* (Vol. 8758, pp. 99–110). SPIE.

Sherchan, W., Nepal, S., & Paris, C. (2013). A survey of trust in social networks. *ACM Computing Surveys (CSUR)*, *45*(4), 1–33. https://doi.org/10.1145/2501654.2501661

Van Ditmarsch, H., van Der Hoek, W., & Kooi, B. (2007). *Dynamic epistemic logic* (Vol. 337). Springer Science & Business Media.

Van Mieghem, P., Omic, J., & Kooij, R. (2009). Virus spread in networks. *IEEE/ACM Transactions on Networking*, *17*(1), 1–14. https://doi.org/10.1109/TNET.2008.925623

Watts, D. J. (2002). A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, *99*(9), 5766–5771. https://doi.org/10.1073/pnas.082090499

Whittlesea, B. W. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(6), 1235–1253.

Yasser, A., & Ismail, H. O. (2021). Trust is all you need: From belief revision to information revision. In *Logics in Artificial Intelligence: 17th European Conference, JELIA 2021*, Virtual Event, May 17–20, 2021, Proceedings (pp. 50–65). Springer-Verlag. https://doi.org/10.1007/978-3-030-75775-5_5