Definition and properties of Belief Dynamics in agents reasoning with argumentation

Matteo Campanelli

January 22, 2014

Contents

1 Introduction

In communication settings, argumentative agents use the content of information to draw conclusions and make decisions, however awareness of other non-logical features of the information they acquired, such as the context in which they received it, may enrich their reasoning capabilities.

Information always comes together with meta-information [Bre01]: when exchanging information, agents have available certain bits of meta-information attached to the informational content itself. Some of this meta-information derives directly from the source of the information (is the source reliable?), some others can be specific to the content of the information itself (how relevant is this information for the agent's purposes?), in some cases it can depend on the specific setting and on other information (how connected was this information to the rest?). From now on, I will call this meta-information *context* of the information.

Agents use the information they acquire to draw conclusions on their state of affairs and on their decision. Hence the context attached to this information can be used to affect the way agents reason.

Argumentation is a reasoning framework that allows to deal with incomplete and conflicting information [CRL00, Pra10, AC98].

Agent dialogues [MP09] are a communication framework by which agents may exchange information in the form of arguments. If context is attached to arguments this can affect the way agents reason with them. Properties of context (meta-logical information) may affect properties of arguments (logical information).

Some work has tried to integrate properties of metalogical information [1] in argumentation systems. In [PTS$^+$11] Parsons et al. describe a Trust Argumentation System where properties of a trust network between agents are propagated in an argumentation system where, under certain assumptions, arguments from less trustworthy agents cannot overrule arguments from more trustworthy agents. The approach in this work makes it possible to define formal properties on the acceptability of arguments based on properties of the multi-agent system at hand (a trust network in the specific case). Nonetheless the system in [PTS$^+$11] shows several limitations: it does not support defeasible argumentation; it is based mainly on trust and it is not explicitely extensible to other types of metalogical information or properties on these; it focuses on static properties of the system rather on dynamic ones which may be important in dialogical settings (what changes do new arguments cause in the system? what happens if properties of trust change?).

Other related work comes from the subfield of Belief Revision and Argumentation [FKIS09]. In [Pag04, PC05] Paglieri and Castelfranchi discuss a cognitive framework of belief dynamics called *Data Based Revision*. Here arguments are considered as "persuasions to believe" and used to initiate successful revision. They distinguish between two basic informational categories: *data* an *beliefs*. These represent pieces of information that are simply gathered and stored by the agent (data) and pieces of information that the agent considers to some extent truthful representations of the states of the world (beliefs). The latter are a subset of the former, in fact in this framework belief change is conceived as a two-step process. Paglieri and Castelfranchi's work offers a *parametric* approach to belief change since agents may choose which data to rely more on according to functions on the metalogical information attached to them (e.g. credibility, likeability...). By means of different parameters, agents can develop more or less complex attitudes that determine which data (arguments) they come to believe. This helps to capture individual variation in epistemic dynamics, i.e. specifying different strategies of belief change for different agents and/or for different contexts and tasks. Some limitations of this approach, however, are the following: arguments are considered as "persuasions to believe" and thus they are used only to initiate a belief revision process. As a consequence arguments and their structural properties are used only to evaluate metalogical information attached to them, not for drawing conclusions; the impact of context is then limited to this first step of reasoning.

[1]The term "metalogical information" is inspired by [Bre01] where "meta-knowledge" and "meta-information" are used to refer to non-logical aspects of information, e.g. its reliability

In [vdWDM+12] a formal model for dialogical agents that are able to discuss their preferences is presented. If an agent expresses a subjective statement, this is assumed to have a higher conclusive force over non-subjective conflicting statements. Also in this work, a metalogical property of arguments affects the way beliefs are drawn, but the approach is specific to subjectivity.

In this work I aim at defining a general formal system where agents can assess metalogical information on external data they receive during communication settings (e.g. dialogues) and these metalogical features explicitely affect the conclusions they draw by defeasible argumentation (given certain metalogical properties it is clear what effects they have on arguments).

In the remainder of this introduction I will describe related work more in detail, in particular I will describe Parsons et al.'s work on argumentation and trust [PTS+11], Paglieri and Castefranchi's DBR framework [Pag04] and introduce general work on Belief Revision and Argumentation. Then I will present some general assumptions and approach choices of this work: I will show examples of metalogical information on arguments, show the intuitions behind the properties whose formalization is the aim of this thesis and then generally introduce the framework for structured defeasible reasoning ASPIC+. Finally, I will state the problem statement of this thesis and its research questions.

1.1 Related Work

In this section I will outline two branches of research. I will describe the work in [PTS+11]; several of the ideas in this paper have inspired the work in the current thesis. Secondly, I will give an overview of Belief Revision and Argumentation; literature in belief revision and its more recent blendings with argumentation systems have inspired part of my approach.

1.1.1 Argumentation-based reasoning in agents with varying degrees of trust

The following is a summary of [PTS+11]. In this work, the authors define an argumentation system that they call AT_1 which has the ability to ensure that arguments grounded in information from untrustworthy agents cannot overrule arguments grounded by more trustworthy agents.

Trust is a very important property in multi agent systems. Autonomous agents should be able to carry out their owners' wishes while interacting with other entities. It is thus important that they reason about whether they should trust these other entities, how much they are trusting them. Argumentation is a model of reasoning that seems well-suited for agent-based systems. This work assumes a model for a network of trust, how much agents trust each other, embedding it in a model of argumentation. Then, it proposes some desired properties for this trust-argumentation system. Since the model is (deliberately) vague on some respects the authors then show how, making some assumptions on some of these systems, some of the properties previously described can be satisfied.

This work assumes a network of trust, i.e. a graph $(Ags, \mathcal{T})$ where $\mathcal{T}$ is a set of edges. In this graph, the set of agents is the set of vertices, and the trust relations define the arcs. Also, a measure of trust is defined such that

$$tr : Ags \times Ags \to \mathbb{R}$$

The trust network is consistent with the trust function:

$$tr(ag_i, ag_j) = 0 \;\Leftrightarrow\; (ag_i, ag_j) \notin \mathcal{T}$$

Given two agents, we can have indirect trust between them if there is a path in the trust network connecting the two. Given a path $(ag_i, ag_{i+1}, ..., ag_j)$ in the trust network, one can have: $tr(ag_i, ag_j) = tr(ag_i) \otimes^{tr} tr(ag_{i+1}) \otimes^{tr} .... \otimes^{tr} tr(ag_{j-1}, ag_j)$ for some operation $\otimes^{tr}$. The authors also define another operation $\oplus$ for combining different trust paths, but its description is beyond the scope of this work.

The authors assume the argumentation system used in [PWA03], an argument-based nonmonotonic logic; arguments are built as classical proofs from consistent premises. Counterarguments are arguments that negate a premise of their target (undercutting), whose defeat is defined in terms of a priority relation on premises between the two; defeasible inference is then defined with the grounded semantics of [Dun95].

The available information to each agent ag_i is given by $\Delta_i = \Sigma_i \cup \{\bigcup_j CS_j\}$ where Σ_i is the knowledge base each agent mantains and CS_j denotes the information that agent ag_j made public.

Agents will have different degrees of beliefs for the formulas in Δ_i. So if an argument is $A = (S, p)$, we can compute the belief degree on that argument for agent Ag_i as $bel_i(A) = bel_i(s_1) \otimes^{bel} bel_i(s_2) \otimes^{bel} ... \otimes^{bel} bel_i(s_n)$ where $S = s_1, s_2, ...s_n$ (the set S of premises is also called *support* of A). The operator $\otimes^{bel} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ aggregates different levels of belief degree into a new unique value. The authors leave this operator generally abstract but discuss some instantiations and their properties (see TAS Proposition 9 at the end of the current subsubsection). Thus one (belief-based) order $\geq_i^{bel}$ over arguments is derived from belief degrees.

The following is the formal description of an argumentation system in the work.

Definition 1. *An argumentation system is a triple $(A(\Delta_i), Undercut, \succ_i^{arg})$ where $A(\Delta_i)$ is the set of all arguments which can be made from Δ_i. $Undercut$ is a binary relation collecting all pairs of arguments A_1 and A_2 such that A_1 undercuts A_2 and $\succ^{arg}$ is a preference over arguments.*

Note in this definition that the order over beliefs is not the same as the preference over arguments in the definition of argumentation system; the relation $\succ_i^{arg}$ will be defined in terms of $\geq_i^{bel}$

The authors try to answer the question: if an agent makes use of the information obtained from an acquaintance, how should a level of trust towards this other entity affect its conclusions? To do this, they extend the argumentation system defined above including trust levels among agents.

Definition 2. *A Trust Argumentation System (TAS) is a tuple* $(Ags, A(\Delta_i), Undercut, \succ_i^{arg}, \mathcal{T})$ *where* $\mathcal{T}$ *is a trust network.*

As mentioned before, there are a few parameters which needs to be instantiated in this model:

- the trust operators $\otimes^{tr}$ and $\oplus^{tr}$ (that calculates respectively *indirect trust* on a given path and *combine different paths* in a trust network);
- how we can use the trust levels $tr(Ag_i, Ag_j)$. This is achieved by means of the function *ttb*, so that we can calculate the degree of beliefs an agent has on formulas from the trust levels towards the source of that formula;
- we also need to specify $\otimes^{bel}$, that is how beliefs degrees on formulas are combined as belief degrees over arguments;
- how to determine the arguments order $\succ^{arg}$ from the order $\geq^{bel}$.

The authors proceed discussing instantiations of this model and define some of their desirable properties. Some of these desiderata proposed by the authors are concerned with properties of paths on the trust network, since these properties are beyond the scope of my work I will not describe them in this venue. I will now describe some properties concerning the degrees of belief. Recall that the degree of belief of the conclusion of an argument can be loosely seen as the degree of belief on the argument.

TAS Property (Property 6)**.** *Given an argument* (S, p) *where* $S = \{s_1, s_2, ..., s_n\}$*, then:*

$$bel_i(p) \leq min_{j=1,...m} bel_i(s_j)$$

The belief of the conclusion of an argument should be believed no more than the minimum of the elements of its support.

TAS Property (Property 7)**.** *Given two arguments* (S, p) *and* (S', p')*,* $|S| \geq |S'|$ iff $bel(p) \leq bel(p')$

If an argument has a larger support it should not have a greater belief degree than an argument with a smaller support (they don't have to have the same conclusion)

TAS Property (Property 8)**.** *Given two arguments* (S, p) *and* (S', p')*, if* $S \subseteq S'$ iff $bel(p) \geq bel(p')$

If the support of an argument support is included in another then the first should not have a lower belief degree than the second.

The authors prove that *Property 7* implies *Property 8* but also that they are not equivalent.

TAS Property (Property 9)**.** *If agent* ag_i *has two arguments* (S, p) *and* $(S,', p')$*, where two supports have corresponding sets of agents* Ags *and* Ags' *then* (S, p) *is stronger than* (S', p') *only if* ag_i *considers* Ags *to be more trustworthy than* Ags'*.*

If this property holds, arguments grounded in information coming from less trustworthy agents will not be able to defeat arguments coming from more trustworthy sources. Note that in the above property trust is compared between sets of agents. Given two agents ag and ag' we say that $ag \geq^{tr} ag' \Leftrightarrow tr(ag) >= tr(ag')$. For two sets of agents Ags and Ags' we say that $Ags \geq^{tr} Ags' \Leftrightarrow min_{ag \in Ags}(tr(ag)) \geq max_{ag' \in Ags'}(tr(ag'))$

If this property is obeyed the following proposition holds.

TAS Proposition (Proposition 1)**.** *In a trust argumentation system* $(Ags, A(\Delta_i), Undercut, \succ_i^{arg}, \mathcal{T})$ *if an argument* (S, p) *with corresponding set of agents* Ag*, is acceptable, then, given Property 9, a new argument* (S', p') *with corresponding set of agents* Ag' *cannot make* (S, p) *not acceptable if* Ag_i *considers* Ag' *to be less trustworthy than* Ag*.*

As *Proposition 1* shows, *Property 9* can prevent new arguments from less trustworthy agents to make unacceptable more trustworthy arguments that otherwise would be acceptable.

Last, the authors show that particular instances of the system may satisfy some of the properties above. For example they prove the following proposition.

TAS Proposition (Proposition 9)**.** *A Trust Argumentation System that uses minimum for* $\otimes^{bel}$*, maximum for* $\oplus^{tr}$ *and adopts* $L2$ *and* $O1$ *satisfies* Property 9.

In the proposition above $L2$ is a definition of ttb such that

$$ttb(tr(ag_i), tr(ag_j)) = tr(ag_i, ag_j) \cdot min_j\{bel_i(s_j) | s_j \in \Sigma_i\}$$

whereas $O1$ refers to the following simple relation between $\succ_i^{arg}$ and $\succ_i^{bel}$:

$$(S, p) \succ_i^{arg} (S', p') \Leftrightarrow (S, p) \succ_i^{bel} (S', p')$$

1.1.2 Belief Revision

Belief revision is the study of how a knowledge base is updated when removing existing information, accommodating new information or a combination of both. One of the most influential theories of belief revision is the eponymous AGM theory of Alchourron et al. [AGM85], in which they propose the *AGM Postulates* for which they consider three types of changes that can be made to a knowledge base: expansion (where a new formula is added to a belief system together with its logical consequences), revision (where a formula is added to a belief system with which it is inconsistent and that is changed to restore consistency) and contraction (where a formula is retracted from a belief system without adding any new one).

The original AGM framework has been extended in many ways introducing new types of operators in addition to the three standard ones. Two of these operators relevant to this thesis are *selective revision* and *screened revision.* Selective revision [FH99] is a type of non-prioritized revision, that is a process in which new information is received and weighed against old information, with no special priority assigned to the new information due to its novelty (differently from the AGM approach); part or whole of the input information may then be accepted. Screened revision [Mak97] is a variant of non-prioritized revision where there is a set of potential core beliefs that are immuned to revision.

1.1.3 Belief Revision and Argumentation

In [FKIS09, FGKIS11] Falappa et al. develop a conceptual view on the relation between argumentation and belief revision. This conceptual view is first defined by four steps of reasoning:

- *Receiving new information*: new information can be in many forms, e.g. a propositional fact or an argument.
- *Evaluating new information*: Upon receipt of new information, an agent will evaluate it in order to be convinced it is true. If the information is an observation, truth is usually assumed. However, if it is communicated by another agent, some justification will be required.
- *Changing beliefs*: If the agent chooses to accept the information, it will apply belief revision techniques to incorporate it into its knowledge base.
- *Inference*: From its new epistemic state, the agent derives plausible beliefs that guide its behaviour.

In this thesis I will focus on the first three steps: modelling the way agents receive information, their evaluation and imposing properties on the way the change their beliefs given the metalogical content of the (possibly new) information. The final step of inference will be left to the argumentation machinery.

1.2 Research assumptions

In the following paragraphs I describe some of the assumptions for work in the present thesis.

In the introduction I have said how the information agents receive may have a certain context, representing some non-logical of its content. The way we aim at using these contexts is to affect the conclusions taken by argumentative agents; one way to do this is by letting contexts affect logical properties of arguments. In the next paragraphs I give some more specific examples of the contexts to give an intuition of the type of applications they can provide.

Examples of context

One important example of context in MAS is *trust* [PTS^{+}11, PSM12]. Trust is an approach for measuring and managing the uncerttainty about autonomous entities and the information they deal with. Trust is a type of metalogical information that is specifically related to the source of the information.

Paglieri and Castelfranchi [Pag04] propose contexts that are meant to describe the "overall epistemic processing of a cognitive agent":

- Relevance: a measure of pragmatic utility of the datum;
- Credibility: a measure of the number of all supporting data;
- Importance: a measure of epistemic connectivity of the datum;
- Likability: a measure of the motivational appeal of the datum;

Another specific example of context is *priority*, the *freshness* of a datum. This contextual information is fundamental for a standard assumption in the AGM approach to Belief Revision [Gär88], the axioms of *success*: in AGM new information is prioritized over old one. One way to represent *freshness* as the numerical context of a datum received at time t is representing it as $\frac{1}{t}$.

Other type of contexts may represent values that are not directly available from the environment but may still be significant, e.g. measures that allow to identify good arguments in persuasion settings characterized by uncertainty and inconsistency as described in [Hun13].

Properties of context and properties of arguments

The examples above show cases where contextual information in agent settings may have a certain relevance. If information is represented by means of arguments this context may refer to the argument themselves. If a certain relation on different contexts (and their corresponding pieces of information) holds, what is the relation between the corresponding arguments? If two pieces of information come from differently trustworthy agents, what is the relation between the two resulting arguments, for example in terms of their logical properties?

This scenario is schematized in Figure 1.

As discussed in the previous subsection, the work in [PTS^{+}11] gives one possible answer to the question above: the relation between two pieces of information I_1, I_2 is such that I_1 comes from a more trustworthy source than I_2, then the relationship between the two respective arguments will be "Argument $A(I_1)$ cannot make $A(I_2)$ unacceptable" (see Proposition 9).

One of the aims of this thesis is to provide mappings from properties or relations of contexts to properties or relations of arguments. In the next sections I will define an operator for a process that I will call *contextualized revision* to provide such mappings.

There are some meta-properties we want from contextualized revision. First, we want it to be *general*, that is making almost no assumptions on the underlying metalogical information. Since the metalogical information above can be captured by numbers, I will assume context attached to arguments as a function from arguments to numbers. Thus the properties and relations considered on arguments will have a *numerical nature*. Moreover, a quantitative-oriented approach is also more suitable for experimental settings [KMPV12].

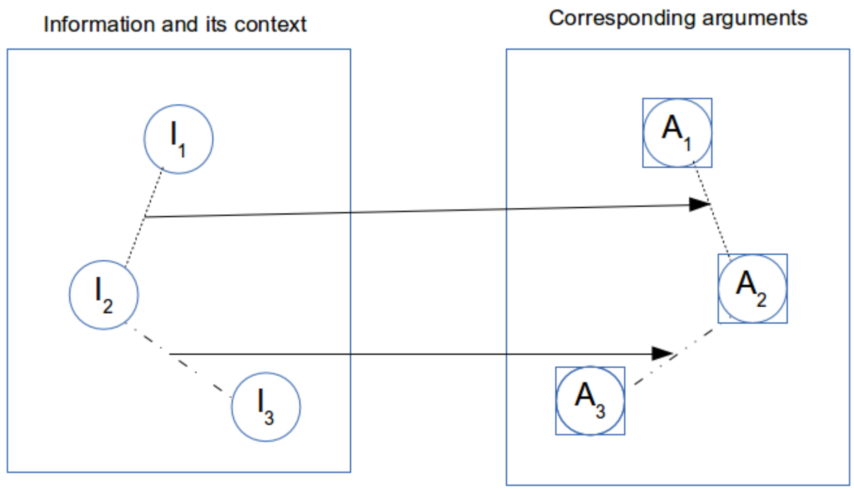

Figure 1: **Context and arguments**. The dotted and dashed lines on the left represent relations between context of pieces of information. The corresponding lines on the right are resulting relations between arguments.

This form of revision should also be able to deal with the *dynamics* of arguments aquisition. We aim at being able to formalize the properties of the set of arguments constructable by an agent after a new argument with attached a certain context is received by it (given a starting situation, e.g. a previous set of arguments and their properties). These dynamical features may constitute a tool for designing strategical dialogical agents [Dij12, Sna12].

There is another sense in which such a framework can be considered dynamic: it must be able to express not only changes in the arguments but also changes in the contexts attached to them. These changes of contexts may be a consequence of the agent's reasoning process itself, thus we want that changes on contexts may be expressed as arguments.

ASPIC+

This work will assume a specific argumentation system: ASPIC+ [Pra10]. This system is chosen for several reasons. It is a rich system of structured argumentation with certain proven properties, which provides the ability to explore the effect of dynamics on the structure of arguments, and the properties of the system as a whole. In many MAS settings, communication exchange is by means of dialogues where agents use formulas to communicate with each other; hence arguments are structured in these contexts. It is rich in the sense that it allows for several types of premises (assumptions and axioms) and rules (defeasible and strict, based on Pollock's theory of defeasible reasoning [Pol87]). Thus ASPIC+ can be a more general case study than other systems (such as the one used in [PTS+11]). Moreover, the fact that ASPIC+ uses two types of inference rules enables us to investigate rule-based dynamics, i.e. how the use of a type of rule affects the revision process. Finally, ASPIC+ is the argumentation system used in Argument Revision [Sna12], which I will use as an example to further applications of formal properties on metalogical information of arguments.

1.3 Research aim

The problem I will tackle is the following:

Problem Statement. *Is it possible to define a geneal theoretical framework, usable in MAS settings, to formally describe reasoning dynamics in argumentative agents by providing formal connections between metalogical properties of the information and logical properties of the arguments expressing that information?*

The method used to answer the problem statement is as follows. First, describe an abstract formal framework built on top of ASPIC+ that describes connections between metalogical information and arguments. Second, formulate a way to define context on arguments in general MAS settings. Third, describe a concrete way to formalize the framework using properties of ASPIC+. Fourth, validate the defined properties and the framework in realistic dialogical settings between agents.

Research Questions

The following research questions stem from the problem statement above.

The first step is to define a formal language to express connections between the context of arguments and their logical properties.

RQ 1: How can we define an abstract formal system for properties of dynamics of arguments with contextual information?

We need to define a formal model for describing the state of the agent that employs arguments with context as well as the dynamic aspects of the system, i.e. how the state can be changed. From this the two following subquestions:

RQ 1.a: What is the formal definition of a system able to express arguments, contextual properties and constraints on these?

RQ 1.b: What is the formal definition of an abstract operator for describing the acquisition of arguments with a certain context?

Given a way to formally describe the connections between arguments and their context, we need to choose which properties we want on these connections that would make the system useful in general agent settings.

RQ 1.c: What general properties should we require for the arguments in the system?

Then these properties have to be formalized in the system.

RQ 1.d: How can we formally describe these properties in the system?

Once we have a system for describing arguments, their context and the connections between the latter and logical properties of the former, we can proceed to another aspect of the system: we have to describe how to go from metalogical aspects of information to a numerical context value on arguments. Thus we need to investigate how to define context on arguments to make the abstract system of the previous research question useful in general settings.

RQ 2: How can we define context on arguments?

RQ 2.a: On which elements of arguments can we assume context is defined?

RQ 2.b: What functions to use to aggregate context of arguments from different elements that are general with respect to the nature of the metalogical information?

Once described a system by which plausibly describe context on arguments, we need to formalize an instance of the model using the way ASPIC+ arguments are constructed.

RQ 3: How can we define a way to modify an agent's knowledge base and an argumentation system so that the properties described in the abstract system of RQ 1 are respected?

Finally, we want to show explicit connection between the formal framework obtained by the three previous questions and settings of agent interaction. To answer this question I define the three subgoals of describing how to use the system in dialogues, how to make it support changes in contexts deriving from the argumentation system itself and how to usefully apply the notions built to answer the previous questions to improve other research in MAS.

RQ 4: Can this system be used in realistic settings of agent interaction?

RQ 4.a: Can this system be concretely used in dialogues?

RQ 4.b: Can the system model changes in the context values due to revision itself?

RQ 4.c: Are the properties defined in RQ 1 flexible enough to formalize and implement useful properties in dialogues or other agent settings?

1.4 Thesis outline

This thesis is structured as follows: Section 2 presents the formal background for logics of defeasible argumentation. Section 3 presents the abstract model of Contextualized Argumentation Theories and Contextualized Revision. Section 3 will answer RQs 1 and 2. Section 4 introduces Contextualized Knowledge Bases as a simple approach to updating knowledge bases by which to obtain properties for Contextualized Revision. Section 4 will answer RQ 3. Section 5 introduces meta-argumentation and uses it to show applications of Contextualized Revisions in dialogue with a dynamical context. Section 5 will answer RQs 4.a and 4.b. Section 6 introduces Argument Revision and shows how to extend it with properties of Contextualized Argumentation Theories. Section 6 will answer RQ 4.c. Finally, Section 7 concludes the thesis and identifies potential areas for future work.

2 Background

In this section I will describe a fundamental component of the agent's reasoning architecture, the logic for defeasible argumentation ASPIC+. I will then proceed describing the considered dialogue system and some aspects of the architectures of the agents.

2.1 Logics for defeasible argumentation

Defeasible logics express a form of reasoning where conclusions can be withdrawn when new information becomes available. Logics for defeasible argumentation (or argumentation systems for short) conceptualize and formalize this type of reasoning as the construction and comparison of arguments for and against a certain conclusion. In these systems conflict between arguments allows new information to "take over" old, thus bringing about a process of inference which is non-monotonic.

In argumentation systems, the fundamental units of knowledge, formulas, are represented by an underlying logical language.

Arguments can be more or less simple: an elementary argument has a conclusion, premises and an inference connecting the conclusion, while a complex argument is a tree of arguments.

Argumentation systems use a binary defeat relation to determine which argument of two attacking arguments wins.

Once defined the defeat relation within a set of arguments, the outcome of an argumentation system is a classification of each of the arguments in one of the three statuses: winning (justified), losing (overruled) and ties (defensible). This classification can be seen as a form of inference in argumentation systems.

I will now formalize the framework for argumentation system ASPIC+.

Definition 3 (Argumentation system). *An* argumentation system *is a tuple* $AS = (\mathcal{L}, Not, \mathcal{R}, \leq_d)$ *where*

- *$\mathcal{L}$ is a logical language,*
- *Not is a contrariness function from $\mathcal{L}$ to $2^{\mathcal{L}}$,*
- *$\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ is a set of strict ($\mathcal{R}_s$) and defeasible ($\mathcal{R}_d$) inference rules such that $\mathcal{R}_s \cap \mathcal{R}_d = \emptyset$,*
- *$\leq_d$ is a partial preorder on $\mathcal{R}_d$.*

Definition 4 (Contrary)**.** *Let $\mathcal{L}$, a set, be a logical language and Not a contrariness function from $\mathcal{L} \to 2^{\mathcal{L}}$. If $\phi \in Not(\psi)$ then if $\psi \notin Not(\phi)$ then ϕ is called a contrary of ψ , otherwise ϕ and ψ are called contradictory. The latter case is denoted by $\phi = -psi$ (i.e., $\phi \in \psi$ and $\psi \in \phi$).*

Definition 5 (Strict and defeasible rules)**.** *Let $\phi_1, ..., \phi_n, \phi$ be elements of $\mathcal{L}$*

- *strict rules are of the form $\phi_1, ..., \phi_n \to \phi$*
- *defeasible rules are of the form $\phi_1, ..., \phi_n \rightsquigarrow \phi$*

ASPIC+ constructs arguments from a pre-ordered set of formulas: a knowledge base.

Definition 6 (Knowledge Base)**.** *A* knowledge base *in an argumentation system $(\mathcal{L}, Not, \mathcal{R}, \leq_d)$ is a pair $(\mathcal{K}, \leq_K)$ where $\mathcal{K} \subseteq \mathcal{L}$ and $\leq_K$ is a partial preorder on $\mathcal{K} \setminus \mathcal{K}_n$. Here $\mathcal{K} = \mathcal{K}_n \cup \mathcal{K}_p$ where these subsets are disjoint and*

- *$\mathcal{K}_n$ is a set of necessary axioms.*
- *$\mathcal{K}_p$ is a set of ordinary premises.*

Intuitively arguments cannot be attacked on their axiom premises, but they can be attacked on their ordinary premises. Defeat depends however on the comparison of the arguments involved.

As said above, in ASPIC+ arguments are represented as trees obtained by chaining inference rules. Here I give a formalization of the constructable arguments from a knowledge base. The following functions on arguments are defined: $Prem$ (premises), $Conc$ (Conclusions), Sub (sub-arguments), $DefRules$ (defeasible rules), $TopRule$ (last inference rule).

Definition 7 (Argument)**.** *An argument A on the base of a knowledge base $(\mathcal{K}, \leq_K)$ in an argumentation system $(\mathcal{L}, Not, \mathcal{R}, \leq_d)$ is:*

1. *ϕ if $\phi \in \mathcal{K}$ with:*
   $Prem(A) = \{\phi\}$
   $Conc(A) = \phi$
   $Sub(A) = \{\phi\}$
   $DefRules(A) = \emptyset$
   $TopRule(A) = undefined$

2. $A_1, ..., A_n \rightarrow \phi$ *such that there exists a strict rule* $Conc(A_1), ...Conc(A_n) \rightarrow \phi$ *in* $\mathcal{R}_s$ *with:*
   $Prem(A) = Prem(A_1) \cup ... \cup Prem(A_n)$
   $Conc(A) = \phi$
   $Sub(A) = Sub(A_1) \cup ... \cup Sub(A_n) \cup \{A\}$
   $DefRules(A) = DefRules(A_1) \cup ... \cup DefRules(A_n)$
   $TopRule(A) = Conc(A_1), ...Conc(A_n) \rightarrow \phi$

3. $A_1, ..., A_n \rightsquigarrow \phi$ *such that there exists a defeasible rule* $Conc(A_1), ...Conc(A_n) \rightsquigarrow \phi$ *in* $\mathcal{R}_d$ *with:*
   $Prem(A) = Prem(A_1) \cup ... \cup Prem(A_n)$
   $Conc(A) = \phi$
   $Sub(A) = Sub(A_1) \cup ... \cup Sub(A_n) \cup \{A\}$
   $DefRules(A) = DefRules(A_1) \cup ... \cup DefRules(A_n)$
   $TopRule(A) = Conc(A_1), ...Conc(A_n) \rightsquigarrow \phi$

Arguments can be classified according to their premises and to their inference rules:

- *strict* if $DefRules(A) = \emptyset$
- *defeasible* if $DefRules(A) \neq \emptyset$
- *firm* if $Prem(A) \subseteq \mathcal{K}_n$
- *plausible* if $Prem(A) \nsubseteq \mathcal{K}_n$

ASPIC+ assumes any pre-order that satisfies the two following basic assumptions:

Definition 8 (Admissible argument ordering)**.** *Let A be a set of arguments. Then a partial pre-order $\preceq$ on A is an admissible argument ordering iff*

if A is firm and strict and B is defeasible or plausible, then $B \prec A$

if $A = A_1, ...A_n \rightarrow \phi$ then for all $1 \leq i \leq n$, $A \preceq A_i$ and for some $1 \leq i \leq n$, $A_i \preceq A$

(Here $B \prec A$ means $B \preceq A$ and $A \npreceq B$.)

Given a knowledge base, an argumentation system and an admissible ordering we can define an *argumentation theory.*

Definition 9 (Argumentation Theory)**.** *An Argumentation Theory is a triple $AT = (\mathcal{AS}, KB, \preceq)$ where $\mathcal{AS}$ is an argumentation system, KB is a knowledge base in $\mathcal{AS}$ and $\preceq$ is an admissible ordering of the set of all arguments that can be constructed from KB in $\mathcal{AS}$ (below called the set of arguments on the basis of AT or $Args(AT)$)*

Next, I shall provide the definitions of two argument orderings, called the *weakest-link* and *last-link* orderings [Pra10]. These two definitions require lifting a partial pre-order $\leq_e$ of elements of two sets to an ordering on the latter. The following definition induces a strict partial order $\prec_s$ of the two sets:

$$S_1 \prec_e S_2 \text{ iff } \exists e_1 \in S_1 \text{ s.t. } \forall e_2 \in S_2 \text{ it holds that } e_1 \leq_e e_2$$

The *weakest-link* principle prefers an argument if this is preferred on both its premises and its defeasible rules.

Definition 10 (Weakest link principle)**.** *Let A and B be two arguments. Then $A \prec B$ iff either*

1. *A is firm and strict while B is defeasible or plausible; or*
2. - *$Prem(A) \prec_s Prem(B)$; and*
   - *If $DefRules(B) \neq \emptyset$ then $DefRules(A) \prec_s DefRules(B)$*

The *last link ordering* compares arguments on the base of their last defeasible rules or, in case both arguments are strict, on the base of their premises. The notion of *last defeasible rules* is defined as follows

Definition 11 (Last defeasible rules)**.** *Let A be an argument*

- *$LastDefRules(A) = \emptyset$ iff $DefRules(A) = \emptyset$*
- *if $A = A_1, ..., A_n \leadsto \phi$ then $LastDefRules(A) = TopRule(A)$, otherwise $LastDefRules(A) = LastDefRules(A_1) \cup ... \cup LastDefRules(A_n)$*

Definition 12 (Last link principle)**.** *Let A and B be two arguments. Then $A \prec B$ iff one of the following holds*

1. *A is firm and strict while B is defeasible or plausible;*
2. *$LastDefRules(A) \prec_s LastDefRules(B)$;*
3. *$LastDefRules(A) = LastDefRules(B) = \emptyset$ and $Prem(A) \prec_s Prem(B)$.*

Attack and defeat

ASPIC+ uses attack and defeat in order to do inference. Attack is a purely syntactic relation between arguments and does not involve any notion of preference, which on the other hand determines defeat.

An argument can be attacked in three ways: on a (non-axiom) premise (undermine), on a defeasible inference rule (undercut) or on a conclusion (rebuttal).

Definition 13 (Rebuttal)**.** *Argument A rebuts argument B (on B') iff $Conc(A) \in Not(\phi)$ for some $B' \in Sub(B)$ of the form $B''_1, ..., B''_n \leadsto \phi$. In such a case A contrary-rebuts B iff $Conc(A)$ is a contrary of ϕ.*

Definition 14 (Undermining)**.** *Argument A undermines argument B (on ϕ) iff $Conc(A) \in Not(\phi)$ for some $\phi \in Prem(B) \setminus \mathcal{K}_n$ of the form $B_1'', ..., B_n'' \rightsquigarrow \phi$. In such a case A contrary-undermines B iff $Conc(A)$ is a contrary of ϕ.*

Definition 15 (Undercut)**.** *Argument A rebuts argument B (on B') iff $Conc(A) \in \overline{B'}$ for some $B' \in Sub(B)$ of the form $B_1'', ..., B_n'' \rightsquigarrow \phi$.*

Definition 16 (Attack)**.** *Argument A attacks B iff A rebuts, undermines or undercuts B.*

We can use ordering on arguments to define which attacks result in defeat.

Definition 17. *An argument A defeats an argument B if A if one of the following cases occurs:*

- *A undercuts B*
- *A succesfully rebuts B, i.e. A rebuts B on B' and either A contrary-rebuts B' or $A \not\prec B'$*
- *A succesfully undermines B, i.e. A undermines B on ϕ and either A contrary-undermines B or $A \not\prec \phi$*

In the remainder of this work I will not assume any specific logic for ASPIC+. Examples will be presented using propositional logic but the underlying concepts are extensible to other cases, unless specified otherwise.

2.2 Argumentation Frameworks

It is possible to transform every ASPIC+ argumentation theory in an *abstract argumentation framework* [Dun95], a set of arguments (whose structure we abstract from) and an attack relation among them. This allows us to define several notions of semantics for arguments.

Definition 18. *A Dung argumentation framework is a tuple $AF = (Args, \mathcal{R})$, where $Args$ is a set of arguments, and $\mathcal{R} \subseteq Args \times Args$.*

Given an argumentation theory $\mathcal{AT}$ the corresponding Dung argumentation framework is given by the $Args(\mathcal{AT})$ and the defeat relation amongs the arguments.

Dung defines the acceptability of arguments, and the characteristic function and admissible extensions of a framework:

Definition 19. *Let $AF = (Args, R), S \subseteq Args$ and let $A, B, C, ...$ denote arguments in $Args$. Then:*

1. *S is conflict free iff $\forall A, B \in S$ it is not the case that $(A, B) \in R$*
2. *A is acceptable with respect to S iff $\forall B \in Args : if (B, A) \in R$ then there is a $C \in S$ such that $(C, B) \in R$*

3. *The characteristic function of AF, denoted F_{AF}, is defined as follows:*
   - $F_{AF} : 2^{Args} \rightarrow 2^{Args}$
   - $F_{AF}(S) = \{A | A \textit{ is acceptable w.r.t. } S\}$
4. *If S is conflict free, then S is an admissible extension of AF iff each argument in S is acceptable with respect to S (i.e., $S \subseteq F_{AF}(S)$)*

For $s \in \{complete, preferred, stable, grounded\}$, an argument is said to be *sceptically justified* under the s-semantics if it belongs to all extensions. An argument is said to be *only credulously justified* under the s-semantics if it belongs to at least one, but not all, extensions. Sceptical and credulous justification coincide for the grounded semantics given that a framework only ever has a single grounded extension.

Intuitively, an argument A is acceptable with respect to S if for any argument B that attacks A, there is a C in S that attacks B, in which case C is said to defend or 'reinstate' A. An admissible extension S can then be interpreted as a coherent defendable position. From hereon, i will write F instead of F_{AF}.

Admissibility is augmented by preferred, complete, stable and grounded semantics:

Definition 20. *Let $AF = (Args, R)$, S a conflict free subset of $Args$, and F the characteristic function of AF . Then:*

- *S is a preferred extension iff S is a set inclusion maximal admissible extension*
- *S is a complete extension iff each argument which is acceptable w.r.t. S is in S ($S = F(S)$)*
- *S is a stable extension iff $\forall B \in S, \exists A \in S$ such that $(A, B) \in R$*
- *S is the grounded extension iff S is the least fixed point of F (the smallest complete extension).*

In ASPIC+, given a set of structured arguments, a Dung-style abstract framework can be derived then evaluated using the standard semantics. This allows the acceptability of structured arguments to be ascertained.

In subsequent sections, given an ASPIC+ argumentation theory $\mathcal{AT}$, the notation $A \in E(\mathcal{AT})$ means A is an acceptable argument in the abstract framework derived from the argumentation theory $\mathcal{AT}$, under some unspecified, unique-extension semantics, subsumed by complete semantics [2] (a semantics is said to be a unique-extension semantics if it prescribes exactly one extension for any argumentation framework).

[2]We use this constraint to be in line with the original work Section 6 is based on, [Sna12]

3 Contextualized Revision

In this section I describe formally the problem of contextualized revision (revision with context) for allowing agents to reason with a context on arguments, i.e. meta-logical information. Although all desired properties and definitions in this section are general and independent of the nature of the context, I will use trust as a study case and as a concrete example in most of the section.

3.1 Introduction to contextualized revision

With the expression *contextualized revision* or *revision with context* I informally mean a reasoning process of argumentative agents where arguments (or their structural elements) are provided with a function, called *context*. This function is considered to be external to the argumentation logic, capturing meta-logical information on it. An discussed in the introduction of this thesis several kinds of meta-logical information may be available in MAS settings and may play a role in the way agents reason. This section will describe part of this role and how this it may be formalized in argumentation in forms of desiderata on the process of contextualized revision. How to capture this meta-logical in logics of argumentation will be discussed later on in this work when presenting meta-argumentation (see Sections 5 and 6).

There are differences and similarities between classical revision and Contextualized Revision. Classical revision aims at consistency in a flat set of beliefs. The aim of contextualized revision is to use this properties of a (numerical) function to impose constraints on the relation between arguments. Several desired properties on the contextualized revision operator expressing these constraints will be presented.

This function expresses properties of arguments leading the reasoning process and in some of the scenarios I will examine in this work it can be seen as a form of entrenchement. Although I use the term "revision" to refer to this process of reasoning dynamics, there are several differences with classical *Belief Revision*. Whereas classical Belief Revision focused on consistency, Success and Failure of beliefs (revision, expansion and contraction), I will study some loosely related concepts in the perspective of structured argumentation, focusing on features of arguments in ASPIC+. Concepts closer to classical revision will be examined in later sections when I discuss extensions of Argument Revision [Sna12] with contextualized revision.

I shall now introduce the formal setting. In this thesis I assume a unique logical language $\mathcal{L}$, left implicit if there is no danger of ambiguity.

Definition 21 (Contextualized Argumentation Theory)**.** *A Contextualized Argumentation Theory (CAT) is a pair* (Ctx, AT) *where* AT *is an ASPIC+ Argumentation Theory and* $Ctx : S_{Ctx} \to \mathbb{R}$ *where* S_{Ctx} *is a set of arguments.*

The function Ctx, also called *context*, is a real function on a set of the arguments.

Definition 22 (Contextualized revision (revision with context))**.** *A contextualized revision function takes as input a CAT (Ctx, AT) and an argument A, the contextualized revision function returns a new CAT:*

$$(Ctx, AT) \otimes A = (Ctx, AT\prime)$$

Given this definition we can see why we made no assumptions on the domain of the context function with respect to AT. The domain of the context function is not a subset of AT because it may be defined on A, the argument we revise with, which may as well not be present in the original argumentation theory — notice that the revision process leaves the context untouched, so the context function is not extended (cases where this assumption is dropped will be discussed later on in this thesis). The domain of the context function is neither a superset of the original AT as not every argument in it has to have the context defined on it. If we assumed the existence of a universe $\mathbb{U}$ of all the possible arguments we can construct (as done in other works on Belief Revision and Argumentation such as [RMF$^+$08]) we might define the context function on any subset of $\mathbb{U}$. The existence of such a universe is not assumed as this would be limiting in dialogical settings: as pointed out in [Sna12], this would require a shared knowledge between dialogue participants of all the possible arguments.

But is it plausible to assume the existence of such a context function on an arbitrary set arguments? The context function is formally defined over arguments, however some of the properties of argumentative settings the context aims at modelling do not refer to the arguments themselves (as ASPIC+ entities) but to other features of the arguments. Context functions may also be generally dependent on individual agents. For example, my main example of context function in this thesis, trust, is commonly attached to the source of the argument. Later in this section I will describe context tranformers that will allow us to map contextual properties of the setting tp contexts on arguments in a plausible way starting from arguments's structure.

In the remainder of this section I will keep the contextualized revision operator abstract and I will discuss some of the desired properties for it. These properties intuitively describe several attitudes an agent can have towards the information on arguments provided by the context function. In the remainder of this section, I will assume a specific context, trust. The properties above should be considered valid for a rather general class of context. Nevertheless the choice of presenting the desiderata below by the trust function is motivated by two main reasons. First, connection with literature on trust and argumentation. Trust is a relevant topic in MAS [HH11, SS05] and recently several works on trust and argumentation appeared [PTS$^+$11, PSM12, TSP12]. As discussed in Section 1, some of this literature has inspired part of this work's approaches. A second reason trust has been chosen is for the sake of a more concrete discussion. Some of the general desiderata below may be more effective when context represents a measure of reliability of arguments; trust is a good example of such a property. Although other examples will be presented, a thorough discussion of which context functions aptly match the desired properties in this section is out of the scope of this thesis.

3.2 Reasoning by contextual revision

So far we described a general setting where new arguments can be introduced in an Argumentation Theory by contextualized revision, here I present a few proposals on the type of reasoning properties that we may want when dealing with contextualized revision.

I shall now introduce the context function that will be used in this section.

Definition 23 (Trust function on arguments)**.** *A trust function is a function of the type* $\tau_A : S \to [0, 1]$ *where* S *is a set of ASPIC+ arguments and* $\tau_A(A) = 1$ *if* $A \in S$ *is firm and strict.*

The assumption on strict and firm arguments will be clarified later and has to do with the special way this type of arguments is treated by admissible orderings in ASPIC+.

3.2.1 An example of contextual revision with trust

I will introduce some motivation behind the desiderata below by means of an example.

Albert (α) is the agent of whom we aim at describing the revision process. Albert interacts with the two agents BraggingBuddy (β), an old friend of his that sometimes utters more boasting than truths, and SageSagan (σ), who is obsessed with a rational understanding of the world and is thoughtful in anything he says. Albert receives argument A_β and A_σ respectively from the other two agents. He trusts both agents but he trusts information from σ more than information from β; thus we have $\tau(A_\sigma) > \tau(A_\beta)$.

How should trust affect α's conclusions? Assuming he has no way of determining the reliability of these two arguments other than trust, he may want to be sure that information from β cannot take over information from σ (if the two are conflicting). In terms of argumentation in ASPIC+ this means that A_β should not defeat A_σ. Desideratum 3 expresses this scenario. At the same time, if the information are conflicting we want σ's information to be able to take over β's. Desideratum 2 expresses this property.

A third property, not directly related to trust, has to do with a general desirable feature of dynamics of argument acquisition. The abstract operator of the previous subsection aims at modelling the receipt of an argument and the consequent modifications on an argumentation theory. It may be that this acquired argument is not trustworthy enough, as in the previous example, or that some other property of contextualized revision limits the strength or the use of this argument for other reasons. In either case, one desirable property of a revision operator is that in the future (i.e. after further revisions), if an agent aims at using that argument, for example because the context function changed or the argumentation theory underwent other modifications, it can still utilize it even if it did not in the past; in other words the revised argumentation theory should keep the argument. This property will be called *Total memory* in the following paragraphs.

3.2.2 Discussion: gathering revision properties

From the example above, we may gather some initial intuitions for initial properties. Here I informally introduce what will be formalized in the next section as desired properties: three of the constraints that we may want from revision with context.

- Total memory: arguments should not be removed from the AT, even if they have to overruled by others;
- Enabling attack (also called consistency with context ordering): if an argument is more trustworthy than another, the preference in the AT should enable it to defeat it;
- Screened defeat: if an argument is strictly more trustworthy than another it should defeat it.

Notice that the last two properties refer to pairs of conflicting arguments. Also notice that the requirements of the third property, screened defeat, will be refined excluding cases of preference-independent attacks (see Property on CAT 2).

3.3 Desiderata on Contextualized revision

This section formalizes the notions of the previous section introducing formal desired properties on contextualized revision by means of desiderata on Argumentation Theories in ASPIC+.

These desiderata express properties on contextualized argumentation theories. (If there is no danger of ambiguity, the following definitions and properties will leave the argumentation theory implicit).

I shall now present some desired properties for contextualized revision.

Desideratum 1 (Deductive Monotonic Contextualized Revision)**.** *A contextualized revision operator* $\otimes$ *is deductively monotonic if for all argumentation theories* AT *and arguments* A*, we have that* $\forall A' \in Args(AT) Prem(A') \subseteq \mathcal{K}(AT')$ *where* $(Ctx, AT) \otimes A = (Ctx, AT\prime)$

If this property is satisfied, a contextualized revision operator does not carry out destructive changes in the argumentation theory.

Desideratum 2 ensures consistency between the ordering induced from trust and the one on arguments. The next one, Desideratum 3, guarantees that under some circumstances arguments cannot defeat each other if trust is used to assign them preferences.

The desiderata express constraints on Contextualized Argumentation Theories. These properties are then extended to contextualized revision.

Property on CAT 1 (Consistency with context ordering)**.** *A CAT* (τ_A, AT) *satisfies the property of consistency with context ordering if given two arguments* $A, B \in Args(AT)$ *such that* $\tau_A(A) \leq \tau_A(B)$*,* $A \preceq B$ *(where* $\preceq$ *is the ordering relation in* AT*).*

Desideratum 2 (Consistency with Context Ordering)**.** *A contextualized revision operator* $\otimes$ *satisfies the Desideratum of Consistency with Context Ordering if if* CAT *satisfies the property of the same name then* $CAT \otimes A$ *satisfies it too, for each* CAT *and* A*.*

The next desired property connects the trust levels of two arguments with constraints on their defeat relationship. The next property imposes a condition on pairs of argument:the lack of attacks which are preference-independent.

Definition 24 (Preference-independent attack)**.** *An argument* A *carries out a preference-independent attack on an argument* B *if* A *contrary-attacks* B *or* A *undercuts* B*.*

Defeat in preference-independent attacks is automatic in ASPIC+, thus trust cannot affect them since it imposes only a preference between arguments.

Property on CAT 2 (Screened defeat)**.** *A CAT* (τ_A, AT) *satisfies the property of* Screened Defeat *if given two arguments* $A, B \in Args(AT)$ *such that:*

- $\tau_A(A) < \tau_A(B)$*;*
- *A does not carry out a preference-independent attack on B*

then A does not defeat B in AT.

Desideratum 3 (Screened defeat)**.** *A contextualized revision operator* $\otimes$ *satisfies the Desideratum of Screened Defeat if if* CAT *satisfies the property of Screened Defeat then* $CAT \otimes A$ *satisfies it too, for each* CAT *and* A *.*

A Contextualized Argumentation Theory for which the properties above do not hold is of little use (at least for the settings considered in this thesis), thus the following definition:

Definition 25 (Effective Contextualized Argumentation Theory)**.** *A Contextualized Argumentation Theory is* effective *if it satisfies the properties of Consistency with Context Ordering and Screened Defeat.*

How to build effective Contextualized Argumentation Theories by means of an operator of contextualized revision will be the subject of the next section.

Note that the desiderata above assume a definition of the τ_A function. In the next subsection I will propose how to derive this function from trust on argument premises.

3.4 From trust on formulae to trust on arguments

Arguments in ASPIC+ are structured pieces of information composed by several elements: premises and rules that in turn form other subarguments. So far, I assumed that each argument has a specific level of trust attached to it. However, when an agent considers the trust level of a certain argument it may be that not all the argument's sub-arguments come from the same source.

In this section I propose how the trust level τ_A of a single argument can be obtained from the trust of its composing elements: premises and rules (this is also close to the approach in [PTS$^+$11] where a trust value over an agent is translated to a belief value on premises and only then it becomes an ordering over arguments). Formulae, that will then become premises and rules in arguments, are the basic unit of exchange in several dialogue protocols [PWA03, Pra05], for this reason trust is assumed as "given" on these elements, rather than on whole arguments. Thus I assume that trust on arguments, τ_A, is defined for an argument A only if for all $\phi \in Premises(A) \cup DefRules(A) \cup StrictRules(A)$ a trust function on formulae, τ, is defined [3]. The set $Premises(A) \cup DefRules(A) \cup StrictRules(A)$ will be from now on called $Formulae(A)$.

ASPIC+ has a type of premise which is treated specially : necessary axioms (see definition of knowledge base and admissible orderings in the background section on ASPIC+). Given a KB, trust level on any formula that belongs to K_n is assumed to be 1.

This function above is the one on which trust on arguments can be defined. Desiderata 2 and 3, however, require a trust function defined on arguments. I shall now discuss a few ways to get from trust on premises to trust on arguments. This passage from τ to τ_A is partly inspired from operators for combining belief degrees discussed in [PTS$^+$11].

Definition 26 (Trust on arguments as minimum)**.** $\tau_A : S \to [0, 1]$
$\tau_A(A) = min_{\phi \in Formulae(A)}(\tau(\phi))$ *where S is a set of arguments.*

For the next definition of trust over arguments I require that the premises of the arguments all have the same trust level. I shall call these arguments *homogeneously trustworthy.*

Definition 27 (Homogeneously trustworthy arguments)**.**

$$H_\tau^S = \{A \in S | \exists t \in [0, 1] s.t. \forall \phi \in Formulae(A) \tau(\phi) = t\}$$

The set of arguments S in H^S will be considered implicit when there is no risk for ambiguity.

Definition 28 (Homogeneous trust on arguments)**.** $\tau_A : H\tau \to [0, 1]$
$\tau_A(A) = t$ *(where t is the trust level of each of the formulae in $Formulae(A)$)*

The functions in definition 26 and 28 will also be referred to as respectively τ_{min} and τ_{hom}.

Notice that τ_{hom} is a restriction of τ_{min} over homogeneously trustworthy arguments.

The definitions above represent measures of trust for arguments starting from trust on their premises. Definitions 26 reflect a cautious attitude towards arguments which contain heterogeneous trust levels: the agent considers the least trustworthy subargument in order to evaluate the whole argument. Definition

[3] In general I assume a generic logical language $\mathcal{L}^U$ and universe of rules $\mathcal{R}^U$ such that $Dom(\tau) = \mathcal{L}^U \cup \mathcal{R}^U$

28 can be used in contexts where the arguments the agent reason with come all from the same source. More sophisticated options may be explored with an eye on literature on trust and argumentation [PTS+11] or trust in MAS [HH11].

So far I have discussed desired properties on contextualized revision and showed how a context function on arguments can be built from a context function from the formulae in them. In the next section I will present a way to obtain a context revision operator by exploiting the context function to determine arguments' preferences.

Generalizing with context transformers

In the previous paragraphs I described a different type of context function than the one used in the definition of $CATs$. These functions are defined on the formulae that constitute arguments and their values are used to produce a context on the arguments themselves (e.g. τ_A).

In particular I proposed two different definitions of τ_A: τ_{min} and τ_{hom}. These are not to be considered the only possible definitions of a context function on arguments from a context on formulae, but are simply the ones we are using for further investigation of the theoretical framework at hand, given their generality. At the same time, the rationale behind τ_{min} and τ_{hom} is still applicable in a number of domains, so it is worth abstracting from the way we defined them in the previous sections.

The following definitions are not completely formalized because they would be very similar to definitions 26 and 28, however it is straightforward to generalize the results in the previous section from the following descriptions.

I call (context) *transformer* (usually denoted with $\mathcal{T}$) any operator that takes a context on formulae $ctx : \mathcal{L}^U \cup \mathcal{R}^U \to \mathbb{R}$ and returns a context on arguments $Ctx_A : S \to \mathbb{R}$ (where S is a set of arguments).

Definition 29 (Min-Transformer)**.** *Given a context function on formulae ctx, the application of the tranformer $\mathcal{T}_{min}$ on ctx, $Ctx_A = \mathcal{T}_{min}(ctx)$, is a context on arguments behaving analogously to τ_{min}*

Definition 30 (Hom-Transformer)**.** *Given a context function on formulae ctx, the application of the tranformer $\mathcal{T}_{hom}$ on ctx, $Ctx_A = \mathcal{T}_{hom}(ctx)$, is a context on arguments behaving analogously to τ_{hom}*

The operators above are not only a further layer of abstractions for the definitions in the previous section, in fact they will turn useful later on in this work (namely with *Contextualized Knowledge Bases*) allowing a higher degree of modularity.

4 Contextualized Revision by update of preferences

In this section I propose a possible way to build the abstract contextualized revision operator described in the previous section. For this purpose I will describe the basic notions of Contextualized Knowledge Bases and their *evaluation*. By using the properties of ASPIC+ argumentation theories, I will prove how these notions can achieve effective contextualized revision under certain assumptions.

4.1 Contextualized Knowledge Base

The notion of Contextualized Knowledge Base allows to concretely instantiate contextualized revision operators.

Definition 31 (Contextualized Knowledge Base). *Given a logical language $\mathcal{L}$, a Contextualized Knowledge Base (CKB) on $\mathcal{L}$ is a triple $(\mathcal{K}, \mathcal{R}, pref)$ where*

- $\mathcal{K} \subseteq \mathcal{L}$*;*
- $\mathcal{R}$ *is a set of rules (in the sense of ASPIC+) over* $\mathcal{L}$*;*
- $\mathcal{R}_d \subseteq \mathcal{R}$ *is the set of all the defeasible rules in* $\mathcal{R}$*;*
- $pref$ *is a function from* $\mathcal{K} \cup \mathcal{R}$ *to* $\mathbb{R}$*.*

The set $\mathcal{K}$ in a CKB is close to a knowledge base in ASPIC+. Definition 32 of the Argumentation Theory derived from a CKB will show how a CKB is closer to $\mathcal{K}_p$ as necessary axioms are not being used in CKBs.

The set $\mathcal{R}$ of rules in a CKB may contain strict as well as defeasible rules. To simplify the formalism I assume that strict rules (that is any rule in $\mathcal{R} \setminus \mathcal{R}_d$) are in the domain of $pref$, however since strict rules will not be used when building an argumentation theory from $pref$ we can assume $pref(r_s)$ to be an undefined value for any strict rule r_s.

In the following I will assume the existence of a contrariness function $Not : \mathcal{L} \rightarrow 2^{\mathcal{L}}$ (see definition 3).

Definition 32 (Argumentation Theory from CKB). *Given a Contextualized Knowledge Base* $CKB = (\mathcal{K}, \mathcal{R}, pref)$ *an* Argumentation Theory from CKB *is* $AT_{ord}(CKB) = (AS, KB, \preceq_A)$ *where:*

- $AS = (\mathcal{L}, Not, \mathcal{R}, \leq_d)$
- $\leq_d$ *is such that* $\forall r_1, r_2 \in \mathcal{R}_d\ r_1 \leq_d r_2 \Leftrightarrow pref(r_1) \leq pref(r_2)$
- $KB = (\mathcal{K}, \leq_K)$
- $\leq_K$ *is such that* $\forall \phi_1, \phi_2 \in \mathcal{K}\ \phi_1 \leq_K \phi_2 \Leftrightarrow pref(\phi_1) \leq pref(\phi_2)$
- *the admissible ordering over arguments* $\preceq_A$ *is derived by applying the weakest-link or last-link principles to* $\leq_d$ *and* $\leq_K$*, denoted respectively by* $AT_{ll}(CKB)$ *or* $AT_{wl}(CKB)$ *(If the assumed admissible ordering is not relevant I will use the notation* $AT(CKB)$*)*

Note that all the formulas in the resulting Knowledge Base in an Argumentation Theory are ordinary premises.

Now that we defined CKBs we want to see how to build them. There are two main elements: the $formulae$, by which to build the arguments (given by $\mathcal{K}$ and $\mathcal{R}$) and the $preferences$ by which comparing these arguments according to some prior The following notion is the basic building block for achieving contextualized revision with contextualized knowledge bases.

Definition 33 (Evaluation of an argument in a CKB)**.** *Given a Contextualized Knowledge Base CKB, an argument A and a context function $ctx : Formulae(A) \to \mathbb{R}$, the evaluation of argument A in CKB by context ctx is a function eval that returns a new CKB:*

$$CKB\prime = eval(CKB, A, ctx)$$

where $CKB\prime = (\mathcal{K}\prime, \mathcal{R}\prime, pref\prime)$ and

- $\mathcal{K}\prime = \mathcal{K} \cup Prem(A)$
- $\mathcal{R}\prime = \mathcal{R} \cup Rules(A)$
- $pref\prime(\phi_A) = ctx(\phi_A)\ \forall \phi_A \in Formulae(A)$
- $pref\prime(\phi) = pref(\phi)\ \forall \phi \in \mathcal{K}(CKB) \cup \mathcal{R}(CKB)$

As the definition above shows, when evaluating an argument the new preference function in $CKB\prime'$ is an extension of that in CKB.

If an agents evaluates its Contextualized Knowledge Base with a new argument two things happen: the formulae in the argument are stored in the Contextualized Knowledge Base and the preferences corresponding to each of the formulae are updating by means of the context values. Notice that when evaluating a Contextualized Knowledge Base with a new argument we update it with both premises and rules. The reason for this is to enable agents that acquire new arguments to be able to make use of the inference rules in those arguments in the future.

The simple way Contextualized Knowledge Bases deal with preferences on premises and rules comes with a cost and will account for most of the technical challenges in the next subsection for proving that this formalism may support effective Contextualized Argumentation Theories.

The following is an example of the use of evaluation in CKBs to have contextualized revision. We assume a sequence of incoming arguments and to start from an empty CKB evaluating each argument at the time. Formally, We have in input a set of incoming arguments $\mathcal{I} = \{\mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_n\}$, a context function on arguments ctx and a context transformer.

Example 1. *Our agent has received the following sequence of arguments $\mathcal{I} = \{\mathcal{A}_1, \mathcal{A}_2\}$ where each formulas is contextualized by means of a trust function τ.*

For convenience of notation, the functions τ and $pref$ are represented as dictionaries, i.e. sets of pairs $(key, value)$.

The two arguments are:

- $\mathcal{A}_1 = (\{p, r_1 : p \rightsquigarrow q\}, q)$
- $\mathcal{A}_2 = (\{\neg q\}, \neg q)$

The context is given by $\tau = \{(p, .2), (r_1, .4), (\neg q, .3), \}$

Since it has just started the revision process its CKB *is empty, thus we have* $CKB = (\mathcal{K}_0, \mathcal{R}_0, pref)$, *where* $\mathcal{K}_0 = \mathcal{R}_0 = pref = \emptyset$. *We will build the final Contextualized Knowledge Base by iterating* $CKB_i := eval(CKB_{i-1}, \mathcal{A}_i, ctx)$ *for all* $\mathcal{A}_i \in \mathcal{I}$.

i	$\mathcal{K}(CKB_i)$	$\mathcal{R}(CKB_i)$	$pref(CKB_i)$
0	$\{\}$	$\{\}$	$\{\}$
1	$\{p\}$	$\{r_1\}$	$\{(p, .2), (r_1, .4)\}$
2	$\{p, \neg q\}$	$\{r_1\}$	$\{(p, .2), (r_1, .4), (\neg q, .3)\}$

One thing to notice is that CKBs are unaware of transformers when they are built. Thus, as we will see in the next section, given a specific CKB, they respect the properties of the context function on formulas no matter how strict it is the criterion on arguments (as said before, transformers can be used as criterions the agents have on how to use the meta-logical information and their beliefs).

Definition 34 (Preference-based contextualized revision)**.** *Given a context function ctx (on formulas), a context tranformer* $\mathcal{T}$*, a Contextualized Knowledge Base* CKB *and an argument* A*, preference-based contextualized revision is defined as*

$$(Ctx, AT(CKB)) \otimes_{ckb} A = (Ctx, AT(CKB\prime))$$

where $CKB\prime = eval(CKB, A, ctx)$ *and* $Ctx = \mathcal{T}(ctx)$

The definition above is quite general, assuming an arbitrary transformer and making no assumption on the connection between the context on formulas and the one on arguments. However, as we will see in the next section, only specific instances allow us to satisfy desiderata on $CATs$.

Where are ASPIC+ axioms in CKBs?

The reader may notice that in the previous definition of Contextualized Knowledge Bases all the premises we are used are ordinary assumptions in ASPIC+ Knowledge Bases. The reason for this is the following: axioms are premises on which arguments cannot be attacked, thus protecting arguments with axiom premises from undermining on those premises. As we will see in the next paragraphs, a Contextualized Knowledge Base enables us to have a contextualized revision operator where Screened Defeat holds. The latter property protects arguments from defeat from attacking arguments that are less trustworthy (or that in general have a lower context). In this sense, to some extent, high trustworth premises take the place of axioms and I chose not to include the latter to simplify the definitions on Contextualized Knowledge Bases.

4.2 Desiderata for preference-based contextualized revision

Proposition 1 (Preference-based contextualized revision is deductively monotonic)**.** *Sketch of the proof: for Definition 33 in an Argumentation Theory result of evaluation the set of the premises is monotonic with respect to inclusion.*

In the following propositions $StrictArgs(AT)$ and $DefArgs(AT)$ are respectively the sets of strict and defeasible arguments in an Argumentation Theory AT.

Proposition 2. *Desiderata 2 and 3 hold for any preference-based contextualized revision with context τ_{min}^{strict} (the restriction of τ_{min} to strict arguments) in $AT_{ll}(CKB)$ and $AT_{wl}(CKB)$.*

Proof. The proof is the same for both derived argumentation theories AT_{ll} and AT_{wl}. It is sufficient to show that, given any CKB, the properties on CATs corresponding to desiderata 2 and 3 hold for $(\tau_{min}^{strict}, AT(CKB))$. First note that all arguments in $AT(CKB))$ are plausible (see definition 32). For desideratum 2, we want to show that given any two arguments $A, B \in Args(AT(CKB))$ s.t. $\tau_A(A) \preceq \tau_A(B)$ we have $A \preceq B$. Both arguments are strict, hence assuming either ordering on them (weakest-link or last-link) we have $A \preceq B$ iff $Prem(A) \preceq_s Prem(B)$ (see Definitions 10 and 12). For definition of τ_{min} we have $\tau(p_A^{min}) \leq \tau(p_B^{min})$ where p_A^{min} and p_B^{min} are the least trustworthy premises respectively in A and B. Therefore $\forall p_B \in Prem(B)\ \tau(p_A^{min}) \leq \tau(p_B^{min}) \leq \tau(p_B)$. By definition of $\preceq_s$ this implies $Prem(A) \preceq_s Prem(B)$. Hence $A \preceq B$. For Desideratum 3, we have to show that if $\tau_A(A) < \tau_A(B)$ and A does not contrary attack B or viceversa, A does not defeat B. In order for A to defeat B we should have that A attacks B and $B \prec A$. The latter is not possible since, as shown before, $Prem(A) \prec_s Prem(B)$. □

The following proposition states that strict homogeneously trustworthy arguments can satisfy Desiderata 2 and 3.

Proposition 3. *Desiderata 2 and 3 hold for any preference-based contextualized revision with context τ_{hom}^{strict} (the restriction of τ_{hom} to strict arguments) in $AT_{ll}(CKB)$ and $AT_{wl}(CKB)$.*

Proof. The proof is identical to Proposition 2: it suffices to observe that τ_{hom} is a restriction of τ_{min} to homogeneously trustworthy arguments. □

The proposition above shows that a strict argument inserted by evaluation and coming entirely from a source S can not be defeated by incoming arguments from sources less trustworthy than S (except for contrary attacks). The assumption of homogeneous trust is used to ensure that arguments come from a single source.

These propositions can partly be extended to defeasible arguments, viz. arguments that use defeasible rules.

I will show that using τ_{hom} can be extended to defeasible arguments in a straightforward way, while the same is not true for τ_{min} and the reason for this will be explained later on.

Proposition 4. *Desiderata 2 and 3 hold for any preference-based contextualized revision with context τ_{hom} in $AT_{ll}(CKB)$ and $AT_{wl}(CKB)$.*

Proof. As in Proposition 2, it is sufficient to prove that given any CKB, the properties on CATs corresponding to desiderata 2 and 3 hold for $(\tau_{hom}, AT(CKB))$. Let us consider two arguments $A, B \in Args(AT(CKB))$ s.t. $\tau(A) < \tau(B)$. First observe that if $\tau(A) < \tau(B)$ then $Prem(A) \prec_S Prem(B)$ and $DefRules(A) \prec_S DefRules(B)$. This because given any formula (premise or rule) ϕ_A and ϕ_B respectively such that $\phi_A \in Formulae(A)$ and $\phi_B \in Formulae(B)$, we have $\tau_{hom}(\phi_A) < \tau_{hom}(\phi_B)$. If $DefRules(A) \neq \emptyset$, given any rule $r_A \in DefRules(A)$ we have that $\tau_{hom}(r_A) = \tau_{hom}(\phi_A) < \tau_{hom}(\phi_B) = \tau_{hom}(r_B)$ where r_B is an arbitrary defeasible rule of B. Hence $DefRules(A) \prec_S DefRules(B)$. Reasoning in an analogous way, we can show that $Prem(A) \prec_S Prem(B)$ and assuming $\tau(A) \leq \tau(B)$ we can have $Prem(A) \preceq_S Prem(B)$ and $DefRules(A) \preceq_S DefRules(B)$.

We can now observe that, both with Last Link or Weakest Link (see Definitions 10 and 12), it holds that $\tau(A) \leq \tau(B) \Rightarrow A \preceq B$, hence Property 2 is satisfied. To show that Property 2 holds, we can observe that under its assumptions $A \prec B$. Now we only need show that we have $B \not\prec A$. The latter holds since $Prem(A) \prec_S Prem(B)$ and $DefRules(A) \prec_S DefRules(B)$ and neither A nor B is strict and firm. Hence Property 3 holds.

$\square$

We saw that we could show properties of CKBs for τ_{min} with strict arguments, whereas the properties above could not be easily shown to be holding for general arguments. This technical difficulty is due to a feature of the admissible orderings we are considering (weakest-link and last-link) and to the way Contextualized Knowledge Bases work. This will be exemplified by the following.

Example 2. *Let us consider the following Knowledge Base, rule set and trust function.*

- $\mathcal{K} = \{p, s\}$
- $\mathcal{R} = \{r_1 : p \rightsquigarrow q, r_2 : s \rightsquigarrow \neg q\}$
- $\tau(p) = 0.7; \tau(s) = 0.8; \tau(r_1) = 0.9; \tau(r_2) = 0.6$
- $A = (\{p, r_1\}, q)$*;* $B = (\{s, r_2\}, \neg q)$

We are considering $\tau_A = \mathcal{T}_{min}(\tau)$ as a context on arguments. Thus we have $\tau_A(A) < \tau_A(B)$. However if we build a CKB by evaluating the two arguments, we have $Prem(A) \not\prec_S Prem(B)$ and $DefRules(A) \not\prec_S DefRules(B)$. Thus neither with last-link nor weakest-link we can have $A \prec B$.

5 Using meta-argumentation for dynamic contexts and applications to dialogues

In the previous sections I showed how we could use context values, conceived as properties of the system at hand, to influence properties of arguments. In some settings these properties and consequent context may come directly from the arguments themselves. For example agents may argue about the trustworthiness of a third source of information [PSM12] or state whether some of their statements are subjective [vdWD12].

In this section I will use the CKBs presented in the last section to define a meta-argumentation approach by which agents may change the preferences they use during contextualized revision. This approach will be presented in the setting of agent dialogues and by specific examples on trust.

The meta-argumentation approach presented here extends the possibilities of Contextualized Revision to contexts so that agents can use arguments to express context and reason on these. The meta-argumentation system proposed is based on the work of [vdW11].

These preliminaries on meta-argumentation will also be used in the next section on Argument Revision.

5.1 Preliminaries on meta-argumentation

A general idea of meta-argumentation as an approach is that one can express preferences on arguments as arguments themselves. A predicate $\prec$ is allowed at a higher-level to express order between arguments. Thus, the meta-argument $A = A_1 \prec A_2$ expresses that argument A_1 has less conclusive force than A_2.

In this section we introduce the formal preliminaries of meta-argumentation. The following notions are based on the work in [vdW11]. In this section we will use meta-argumentation to describe context by means of justified conclusions at the meta-level. Thus we will use meta-argumentation to determine context levels on the basis of properties of arguments and their elements. We will not use the full power of meta-argumentation by deriving a defeat relation at the object level (see definition of argumentation system below) from the justified conclusions at the meta-level. A more powerful meta-argumentation framework will be described in Section 6.

The following is the notion of meta-language, which refers to elements of a particular argumentation system, with respect to which it is at the meta-level.

Definition 35 (Meta-language)**.** *Let $\mathcal{AS} = (\mathcal{L}, \mathcal{R}_s \cup \mathcal{R}_d, Not)$ be an argumentation system and let $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$. A Meta-Language w.r.t. $\mathcal{AS}$ is a first-order language $\mathcal{L}'$ such that $\mathcal{L}' \neq \mathcal{L}$ and with at least:*

- *constants:*
  - *for each formula $\phi \in \mathcal{L}$, there is a constant $\underline{\phi}$,*
  - *for each inference rule $r \in \mathcal{R}$, there is a constant $\underline{r}$,*

– *for each argument A in $Args(\mathcal{AS})$ there is a constant $\underline{A}$,*

– *for each subset X of $\mathcal{L}$, $\mathcal{R}$ or $Args(\mathcal{AS})$ there is a constant $\underline{X}$*

- *functions:*

  – *for each function on arguments defined in Definition 7 there is a function symbol, and*

  – *there is an unary function symbol for Not,*

- *predicates:*

  – *unary predicates: np, op*

  – *binary predicates: $in, \preceq_{Args}, \preceq_{\mathcal{R}}$ and $\preceq_{\mathcal{L}}$*

The definition above uses a variation of the definition of argumentation system used in the previous sections (see Definition 3) to align with the work in [vdW11]. In fact here we do not require that the tuple representing the argumentation system contains an the ordering on rules. Given an argumentation system $\mathcal{AS} = (\mathcal{L}, \mathcal{R}, Not)$, we will assume that there exists an ordering $\preceq_{\mathcal{R}}$ over rules. Similarly I will assume that there exists an ordering $\preceq_{\mathcal{L}}$ among the formulae in the logical language $\mathcal{L}$, thus knowledge bases will not be defined as pairs containing a set of fomulae and an ordering but as simple sets. These assumptions will be used in the remainder of this thesis. One of the reasons for such changes in definitions is that the argumentation theory in definition 41 will have an ordering on arguments based determined by justified formulae at the meta-level and not by any fixed ordering in its knowledge base and argumentation system.

The unary predicates "np" and "op" denote that the given formula is a necessary axiom [4] and an ordinary premise respectively in the object-level knowledge base.

The binary predicate $\preceq_{Args}$ is also written as $\underline{x} \preceq_{Args} \underline{y}$ and denotes that y is as strong as or stronger than x. The predicates $\mathcal{R}$ and $\mathcal{L}$ are used similarly and denote relative strength of sets of object-level defeasible rules and sets of object-level formulae.

For ease of notation, if x is a constant, then we will write simply x if there is no ambiguity. For example, instead of $in(\underline{\phi}, \underline{\mathcal{L}})$, we will write $in(\phi, \mathcal{L})$. Further abbreviations are represented in Table 2 ([vdW11]) (notice that the author uses $\mathcal{SR}$ and $\mathcal{DR}$ to denote respectively strict and defeasible inference rules in $\mathcal{R}$).

The following definition describes contrariness for meta-languages, limiting it to symmetrical conflict.

Definition 36. *Let $\mathcal{L}$ be a meta-language for argumentation system $\mathcal{AS}$. A contrariness function for meta-language $\mathcal{L}$ is a contrariness function $Not : \mathcal{L} \rightarrow 2^{\mathcal{L}}$ with at least*

[4] The name "np" is kept for continuity with the work in [vdW11] that uses "necessary premise" instead of "necessary axiom"

Abbreviation	Of
$p \prec_X q$	$p \preceq_X q \wedge \neg(q \preceq_X p)$ (X either Args, $\mathcal{R}$ or $\mathcal{L}$)
$\exists_{x \in X}[\phi]$	$\exists_x[\mathsf{in}(x, X) \wedge \phi]$
$\forall_{x \in X}[\phi]$	$\forall_x[\mathsf{in}(x, X) \supset \phi]$
$\exists_{x_1,\ldots,x_n \in X}[\phi]$	$\exists_{x_1,\ldots,x_n}[\mathsf{in}(x_1, X) \wedge \ldots \wedge \mathsf{in}(x_n, X) \wedge \phi]$
$\forall_{x_1,\ldots,x_n \in X}[\phi]$	$\forall_{x_1,\ldots,x_n}[\mathsf{in}(x_1, X) \wedge \ldots \wedge \mathsf{in}(x_n, X) \supset \phi]$
$X \subseteq X'$	$\forall_{x \in X}[\mathsf{in}(x, X')]$
$X \subset X'$	$X \subseteq X' \wedge \neg(X' \subseteq X)$
$\mathsf{strict}(A)$	$\neg\exists_{r \in \mathsf{rules}(A)}[\mathsf{in}(r, \mathcal{DR})]$
$\mathsf{firm}(A)$	$\forall_{\phi \in \mathsf{premises}(A)}[\mathsf{np}(\phi)]$

Figure 2: Abbreviations in Meta-languages

- *$x \prec y \in Not(y \prec x)$, and*
- *there are no formulae ϕ and ψ s.t. $\phi \in Not(\psi)$ and $\psi \notin Not(\phi)$.*

We may now define meta-argumentation systems:

Definition 37 (Meta-Argumentation System)**.** *Let $\mathcal{AS}$ be an argumentation system. A Meta-Argumentation System on the basis of argumentation system $\mathcal{AS} = (\mathcal{L}, \mathcal{R}, Not)$ is an argumentation system $(\mathcal{L}, \mathcal{R}_s \cup \mathcal{R}_d, Not)$ where*

- *$\mathcal{L}$ is a meta-language for $\mathcal{AS}$,*
- *$\mathcal{R}_s$ is the set of all valid first-order inference rules,*
- *$\mathcal{R}_d$ is a set of defeasible inference rules, and*
- *Not is a contrariness function for $\mathcal{L}$.*

Meta-argumentation systems are a class of argumentation systems, thus it is straigthforward to build a meta-argumentation theory for a meta-argumentation system. Recall that an argumentation theory is a tuple consisting of an argumentation system $\mathcal{AS}$, a knowledge base and argument ordering over the arguments in $\mathcal{AS}$. In the knowledge base we will put several axioms to ensure desirable properties (Figure 3 from [vdW11]):

Definition 38 (Meta-Knowledge Base)**.** *Let $\mathcal{AS}$ be a meta-argumentation system on the basis of an argumentation system $\mathcal{AS}$ and $\mathcal{K} = \mathcal{K}_n \cup \mathcal{K}_p$ is a knowledge base in $\mathcal{AS}$. A meta-knowledge base in $\mathcal{AS}$ on the basis of $\mathcal{K}$ is a tuple $\mathcal{K}' = \mathcal{K}'_n \cup \mathcal{K}'_p$ with*

- *$\mathcal{K}'_n$ containing the axioms in the table in Figure 3,*
- *$x \in X$ iff $in(\underline{x}, \underline{X}) \in \mathcal{K}'_p$,*

Name	Axiom[a]
$\mathtt{trns}'_{\mathsf{Args}}$	$x \preceq_{\mathsf{Args}} y \wedge y \preceq_{\mathsf{Args}} z \supset x \preceq_{\mathsf{Args}} z$
$\mathtt{trns}'_{\mathcal{R}}$	$x \preceq_{\mathcal{R}} y \wedge y \preceq_{\mathcal{R}} z \supset x \preceq_{\mathcal{R}} z$
$\mathtt{trns}'_{\mathcal{L}}$	$x \preceq_{\mathcal{L}} y \wedge y \preceq_{\mathcal{L}} z \supset x \preceq_{\mathcal{L}} z$
$\mathtt{rflx}'_{\mathsf{Args}}$	$\forall_{x \in \mathsf{Args}(\mathcal{AS})}[x \preceq_{\mathsf{Args}} x]$
$\mathtt{rflx}'_{\mathcal{R}}$	$\forall_{x \subseteq \mathcal{DR}}[x \preceq_{\mathcal{R}} x]$
$\mathtt{rflx}'_{\mathcal{L}}$	$\forall_{x \subseteq \mathcal{L}}[x \preceq_{\mathcal{L}} x]$
$\mathtt{adm}'_1$	$\mathsf{strict}(x) \wedge \mathsf{firm}(x) \wedge \neg(\mathsf{strict}(y) \wedge \mathsf{firm}(y)) \supset y \prec_{\mathsf{Args}} x$
$\mathtt{adm}'_2$	$\mathsf{in}(\mathsf{lastRule}(x), \mathcal{SR}) \supset \forall_{y \in \mathsf{dirSub}(x)}[x \preceq_{\mathsf{Args}} y] \wedge \exists_{y \in \mathsf{dirSub}(x)}[y \preceq_{\mathsf{Args}} x]$
$\mathtt{lstLnk}'_1$	$\mathsf{lastDef}(x) \neq \mathsf{lastDef}(y) \wedge \mathsf{lastDef}(x) \prec_{\mathcal{R}} \mathsf{lastDef}(y) \supset x \prec_{\mathsf{Args}} y$
$\mathtt{lstLnk}'_2$	$\mathsf{lastDef}(x) = \emptyset \wedge \mathsf{lastDef}(y) = \emptyset \wedge \mathsf{premises}(x) \prec_{\mathcal{L}} \mathsf{premises}(y) \supset x \prec_{\mathsf{Args}} y$
$\mathtt{set}'_{\mathcal{L}}$	$\exists_{x \in X} \forall_{y \in Y}[\{x\} \prec_{\mathcal{L}} \{y\}] \supset X \prec_{\mathcal{L}} Y$
$\mathtt{set}'_{\mathcal{R}}$	$\exists_{x \in X} \forall_{y \in Y}[\{x\} \prec_{\mathcal{R}} \{y\}] \supset X \prec_{\mathcal{R}} Y$
$\mathtt{prem}'_1$	$\mathsf{op}(x) \wedge \mathsf{np}(y) \supset \{x\} \prec_{\mathcal{L}} \{y\}$
$\mathtt{prem}'_2$	$\mathsf{as}(x) \wedge \mathsf{op}(y) \supset \{x\} \prec_{\mathcal{L}} \{y\}$

[a] All formulae with free variables are implicitly universally quantified.

Figure 3: Axioms for Meta-Argumentation Theories

- $\phi \in \mathcal{K}_p$ *iff* $op(\underline{\phi}) \in \mathcal{K}'_p$,
- $\phi \in \mathcal{K}_n$ *iff* $np(\underline{\phi}) \in \mathcal{K}'_p$,

We can now define meta-argumentation theories:

Definition 39 (Meta-Argumentation Theory)**.** *Let $\mathcal{AS}$ be an argumentation system and $\mathcal{K}$ a knowledge base in $\mathcal{AS}$. A meta-argumentation theory w.r.t. $\mathcal{AS}$ and $\mathcal{K}$ is an argumentation theory $\langle \mathcal{AS}', \mathcal{K}', \leq \rangle$ with:*

- *$\mathcal{AS}'$ a meta-argumentation system w.r.t. $\mathcal{AS}$,*
- *$\mathcal{K}'$ a meta-knowledge base in $\mathcal{AS}$ w.r.t. $\mathcal{K}$, and*
- *$\leq$ an argument ordering over $Args(\mathcal{AS})$ that satisfies the last-link or weakest link principle.*

Note that meta-argumentation theories are a class of argumentation theories (see Definition 9). Thus it is straightforward to construct its corresponding argumentation framework as described in subsection 2.2. We aim at using the status of formulae w.r.t. $\preceq$ in the meta argumentation framework to determine the argument ordering of the object level argumentation theory. The conclusions of the AF of a meta-level argumentation theory will be used to construct the argument ordering of object-level arguments according to grounded semantics, which allows exactly one argument ordering over object-level arguments. This justifies the following definition:

Definition 40 (Grounded Argument Ordering)**.** *Let $\mathcal{AF}$ be the argumentation framework of a meta-argumentation theory $\mathcal{AT}$ w.r.t. $\mathcal{AS}$ and $\mathcal{K}$. The grounded argument ordering on the basis of $\mathcal{AT}$ is a binary relation $\leq$ over $Args(\mathcal{AS})$ such that $A \leq B$ iff $\underline{A} \preceq \underline{B}$ is a justified conclusion of $\mathcal{AF}$.*

Given a meta-argumentation theory, we can build an object-level argumentation theory with the grounded argument ordering of the meta-argumentation theory.

Definition 41 (Argumentation Theory Based On Meta-Argumentation)**.** *Let $\mathcal{AT}$ be an argumentation theory w.r.t. $\mathcal{AS}$ and $\mathcal{K}$. The object level argumentation theory based on $\mathcal{AT}$ is an argumentation theory $\langle \mathcal{AS}, \mathcal{K}, \leq \rangle$ where $\leq$ is the grounded argument ordering on the basis of $\mathcal{AT}$.*

In [vdW11] it is shown that the meta-level argumentation framework satisfies ASPIC+ rationality postulates and that the object-level argument ordering it describes satisfies the last-link principle.

In this section we will make use of meta-argumentation to describe properties on arguments and formulae at the object level such as context levels and we will make use of argumentation theories based on meta-argumentation only in the Section 6 when introducing argument revision. Thus Section 5 will not make use of definitions 41 and 40.

5.2 Updating trust contexts with meta-arguments

The notions in the following paragraphs will enable dialogical agents to argue modifying a context function by means of meta-arguments. Dialogues are systems for inter-agent communication [MP09]. They are described by means of formal protocols defining the agents' communication rules and involve one of more locution acts, i.e. the type of utterances agents can carry out. We will consider a very simple dialogue protocol where the only locution act is $claim(A)$ used for communicating arguments. We assume a binary variant of the locution act, $claim(\alpha, A)$, where α is the agent who uttered A. We will also only cover the case of an agent observing dialogue moves and carrying out contextualized revision on them.

As discussed in [vdWD12], meta-argumentation can be useful to model observation of dialogue moves by agents. In our case agents may have to observe moves for two main reasons: for keeping track of the arguments claimed by other agents and reasoning about the trust context function. The former reason is part of the purpose in [vdWD12] whereas the second one is specific to the setting of Contextualized Argumentation Theory.

Technically, we aim at using a meta-argumentation theory to build a Contextualized Argumentation Theory CAT whose arguments and context function are determined respectively from specific formulae in the meta-knowledge base (i.e. the observed utterances) and from specific justified formulae. The first part of a Contextualized Argumentation Theory (see Definition 21) is its context function, here assumed to represent trust for sake of example. We will enable meta-arguments to determine the trust context function τ in CAT. The other part of CAT is the argumentation theory $\mathcal{AT}(CAT)$; we aim at making sure that the latter is such that CAT is effective in the sense of Definition 25.

This argumentation theory will be produced by a Contextualized Knowledge Base by using definitions and results from the previous section. However, this time we want the meta-arguments to affect the arguments produced by the Contextualized Knowledge Base (because these same meta-arguments affect a context function). We enable this in two ways: updating $\mathcal{K}(CKB)$ with the formulas in the claimed arguments and imposing preferences that reflect the context. These two constraints are exactly those imposed by evaluation, thus we will use the function *eval* on CKBs in the following definitions. Finally, these notions will be formalized in the setting of dialogical agents by including explicit representation of utterance types (only *claim* in our case).

The following definition is an instance of how to define a system that allows reasoning about trust and that brings about a corresponding Contextualized Argumentation Theory.

Definition 42 (Dialogical Trust-based Agent)**.** *A Dialogical Trust-based Agent (DTA) is a tuple* $(\mathcal{AS}_2, \mathcal{K}_2, \tau_{ctx}, CKB)$ *where:*

1. *Let* $\mathcal{AS} = (\mathcal{L}, \mathcal{R}(CKB), Not)$ *is an argumentation system for some logical language* $\mathcal{L}$ *and contrariness function* Not*,*

2. $\mathcal{AS}_2$ *is a meta-argumentation system on the basis of* $\mathcal{AS}$*,*

   - $\mathcal{L}_2$*, the logical language of* $\mathcal{AS}_2$*, is a meta-language for* $\mathcal{AS}$ *and contains the binary predicates* τ *and* claim*,*

3. $\mathcal{K}_2$ *is a meta-knowledge base in* $\mathcal{AS}$ *on the basis of* $\mathcal{K}(CKB)$*,*

4. $\tau_{ctx}(\phi) = t$ *iff* $\tau(\phi, t)$ *(i.e. "*$\tau(\phi) = t$*") is S-justified in the argumentation framework corresponding to the meta-argumentation theory* $\mathcal{AT}_2$ *for some semantics* S*,*

   - $\mathcal{AT}_2 = \langle \mathcal{AS}_2, \mathcal{K}_2, \leq \rangle$ *where* $\leq$ *is an argument ordering over* $Args(\mathcal{AS})$ *that satisfies the last-link or weakest-link principle.*

From now on, given a CKB, we will also denote the argumentation system in the first bullet of the definition above as $\mathcal{AS}(CKB)$.

Notice that τ is a predicate symbol in the meta knowledge base, whereas τ_{ctx} is a function on formulae such as the one defined in Subsection 3.4.

The most important parts of the definition above are (i) the requirement for the meta-logical language to include predicates to describe uttered arguments[5] and context values and (ii) the definition of the context function on the basis of justified conclusions.

By the definition above we obtain a Contextualized Argumentation Theory $(\tau^A_{ctx}, \mathcal{AT}(CKB))$, where τ^A_{ctx} is the context on arguments for τ_{ctx} (i.e. $\tau^A_{ctx} = \mathcal{T}(\tau_{ctx})$ for some tranformer $\mathcal{T}$) .

[5]a similar approach is taken in [vdWD12]

The binary predicate *claim* refers to the *claim* move in argumentative dialogue protocols [Pra05], that in general is used when an agent communicates an argument to one or more other agents in a dialogue. Dialogue protocols usually involve more locution types, but in this venue I will focus on a very simple dialogue protocol based only on the locution *claim*. The predicate is binary so that we can keep track not only of the uttered argument but also of the agent α, the speaker; in an example later in this chapter the agent-parameter will be useful to show the connections with reasoning on trust contexts.

How does a dialogue affect the system above? The connections between the dialogue and the meta-argumentation system may be formalized by a notion similar to the commitment update rules in [vdWD12] (Definition 21). In our setting with our simplified dialogue system we impose that if, at a certain time, an agent with a meta-knowledge base $\mathcal{K}_2$, observes the move $claim(\alpha, A)$ then it updates its meta-knowledge base by the following:

Definition 43 (Updating Knowledge Base during a dialogue)**.** *Let* $DTA = (\mathcal{AS}_2, \mathcal{K}_2, \tau_{ctx}, CKB)$ *be a Dialogical Trust-based Agent, if it observes move* $claim(\mathcal{A})$ *during a dialogue, then the agent is updated to* $DTA' = (\mathcal{AS}_2, \mathcal{K}'_2, \tau'_{ctx}, CKB')$ *such that:*

- *If* $\mathcal{A}$ *is an object-level argument:*
  - $\mathcal{K}'_2 = (\mathcal{K}'_{2,p} \cup \mathcal{K}'_{2,n}, \leq_K)$,
    - $\mathcal{K}'_{2,n} = \mathcal{K}_{2,n}$ *where* $\mathcal{K}_{2,n}$ *is the set of necessary axioms in* $\mathcal{K}_2$,
    - $\mathcal{K}'_{2,p} \supseteq \mathcal{K}_{2,p} \cup \{claim(\alpha, \underline{\mathcal{A}})\}$ [6] *is the smallest set, w.r.t. inclusion, such that* $\mathcal{K}'_2$ *is a meta-knowledge base in* $\mathcal{AS}(CKB')$ *on the basis of* $\mathcal{K}(CKB')$
  - $CKB' = eval(CKB, \mathcal{A}, \tau_{ctx})$,
  - $\tau'_{ctx} = \tau_{ctx}$
- *If* $\mathcal{A}$ *is a meta-level argument (i.e. not an object-level argument):*
  - $\mathcal{K}'_2 = (\mathcal{K}'_{2,p} \cup \mathcal{K}'_{2,n}, \leq_K)$,
    - $\mathcal{K}'_{2,n} = \mathcal{K}_{2,n}$ *where* $\mathcal{K}_{2,n}$ *is the set of necessary axioms in* $\mathcal{K}_2$,
    - $\mathcal{K}'_{2,p} = \mathcal{K}_{2,p} \cup Formulae(\mathcal{A})$
  - $CKB' = CKB$
  - τ'_{ctx} *is such that* DTA' *is a Dialogical Trust-Based Agent (i.e.* τ'_{ctx} *is defined as in bullet 4 of Definition 42).*

Above, an argument is an object-level argument if its premises are a subset of the logical language at the object-language (rules are not a problem because we include them by means of *eval*, thus extending the argumentation system at the object level). The update rule above says that when an agent observes a claimed argument $\mathcal{A}$ at the object level, it would expand the contextualized

[6] α is the agent who made the observed move

knowledge base in the DTA with $\mathcal{A}$. The context values used for the evaluation are taken from the context function in DTA. Also the meta-knowledge base is expanded including a predicate corresponding to the locution act. If $\mathcal{A}$ is an argument at the meta-level, we extend the meta-knowledge base with it. We consider the possible effect of the argument updating the context function τ_{ctx} when a new extension for the new meta-argumentation theory is determined. This is necessary to model the possibility that the claim's conclusions deals with trust levels.

In Section 3 we assumed that context is a function defined on a subset of all possible arguments. When using *eval* in the rule above it is possible that context is not defined on the particular argument we are evaluating with. We can solve this problem by assuming a default value for formulae/arguments whose value is not determined by the predicate τ at the meta-level.

Notice that $\mathcal{K}'_{2,p}$ is well-defined because extending the sets in an argumentation system or a knowledge base at the object level (which is the case with evaluation of contextualized knowledge bases) can only extend a corresponding meta-knowledge base.

The definition of DTA gives rise to a Contextualized Argumentation Theory (τ_{ctx}, CKB). The definition of updating on the other hand gives rise to a contextualized revision process: every time an agent observes the move $claim(A)$ by the definition above we carry out the revision $(\tau_{ctx}, \mathcal{AT}(CKB)) \otimes_{ckb} A = (\tau'_{ctx}, \mathcal{AT}(CKB'))$. The reader may see how, differently from Section 3, here the context function changes with revision.

Notice that the properties of CKB we saw holding for CKBs in the last section still hold here: the resulting CAT is effective. The meta-argumentation formalism shown here allows us to express the interplay between dialogue moves, changes in contexts and consequent contextualized revision.

Notice that given the property of deductive monotonicity of CKBs we have $claim(\alpha, \underline{A}) \in \mathcal{K}_2 \Rightarrow \forall \phi \in Formulae(\mathcal{A})\ \phi \in \mathcal{AT}(CKB)$.

The following is an example of application of Contextualized Knowledge Bases with meta-argumentation.

5.3 Dialogical Trust-based Agents in action: choosing a movie

Agent α is looking for a movie to watch on a Saturday night and asks for advise, let us consider the following dialogue between agents α, β and γ (we will consider only α's perspective listening to these arguments):

1. β: You should watch the movie "Tetsuo The Ironman" by Shinya Tsukamoto because it is good cyberpunk.

2. β: You should watch the movie "Blade Runner" by Ridley Scott because of its director.

3. γ: When β talks about Japanese movies, it is usually too biased

4. γ: Tetsuo is a Japanese movie.

5. γ: You should not watch Tetsuo because it is too violent.

6. γ: You should not watch Blade Runner either because it is based on a novel by a lousy writer.

We will assume that, by default, α attaches a trust level of 0.8 and 0.7 to formulae respectively from β and γ (see defeasible rules below). Since we will deal with defeasible arguments I assume τ_A^{hom} to construct context levels on arguments in order to use properties of effective Contextualized Argumentation Theories with CKBs.

The following table summarizes the arguments constructible by α (the column CKB^i below represents the Contextualized Knowledge Base of the Trust-Based Dialogue Agent α at time i, the union of the sets $\mathcal{R}(CKB)$ and $\mathcal{K}(CKB)$). Some other aspects of the formalism have been simplified and are discussed below.

i	Locution	$\mathcal{K}_2^i$	τ_A^i	CKB^i
0	-	$\{\}$	$\{\}$	$\{\}$
1	$claim_\beta(\mathcal{A}_1)$	$\mathcal{K}_2^0 \cup \{claim(\beta, \underline{\mathcal{A}_1})\}$	$\{\mathcal{A}_1 : .8\}$	$CKB^0 \cup Formulae(\mathcal{A}_1)$
2	$claim_\beta(\mathcal{A}_2)$	$\mathcal{K}_2^1 \cup \{claim(\beta, \underline{\mathcal{A}_2})\}$	$\{\mathcal{A}_1, \mathcal{A}_2 : .8\}$	$CKB^1 \cup Formulae(\mathcal{A}_2)$
3	$claim_\gamma(\mathcal{A}_3)$	$\mathcal{K}_2^2 \cup Premises(\mathcal{A}_3)$	$\{\mathcal{A}_1, \mathcal{A}_2 : .8\}$	CKB^2
4	$claim_\gamma(\mathcal{A}_4)$	$\mathcal{K}_2^3 \cup \{claim(\gamma, \underline{\mathcal{A}_4})\}$	$\{\mathcal{A}_1 : .5, \mathcal{A}_2 : .8, \mathcal{A}_4 : .7\}$	$CKB^3 \cup Formulae(\mathcal{A}_4)$
5	$claim_\gamma(\mathcal{A}_5)$	$\mathcal{K}_2^4 \cup \{claim(\gamma, \underline{\mathcal{A}_5})\}$	$\{\mathcal{A}_1 : .5, \mathcal{A}_2 : .8, \mathcal{A}_4, \mathcal{A}_5 : .7\}$	$CKB^4 \cup Formulae(\mathcal{A}_5)$
6	$claim_\gamma(\mathcal{A}_6)$	$\mathcal{K}_2^5 \cup \{claim(\gamma, \underline{\mathcal{A}_6})\}$	$\{\mathcal{A}_1 : .5, \mathcal{A}_2 : .8, \mathcal{A}_4, \mathcal{A}_5, \mathcal{A}_6 : .7\}$	$CKB^5 \cup Formulae(\mathcal{A}_6)$

The arguments and formulae in the table above are the following:

- $\mathcal{A}_1 : \{goodCyberpunk(Tetsuo) \rightsquigarrow shouldWatch(Tetsuo)\}$
- $\mathcal{A}_2 : \{director(BladeRunner) \rightsquigarrow shouldWatch(BladeRunner)\}$
- $\mathcal{A}_3 : \{japanese(m), claim(\beta, \underline{A}), Conc(\underline{A}) = shouldWatch(m), \phi \in Formulae(A) \rightsquigarrow \tau(\phi) = .5\}$ (here A is a variable on arguments)
- $\mathcal{A}_4 : \{japanese(Tetsuo)\}$
- $\mathcal{A}_5 : \{violent(Tetsuo) \rightsquigarrow \neg shouldWatch(Tetsuo)\}$
- $\mathcal{A}_6 : \{lousyWriterStory(BladeRunner) \rightsquigarrow \neg shouldWatch(BladeRunner)\}$
- $r_{\tau(\beta)} : \{claim(\beta, A), \phi \in Formulae(A) \rightsquigarrow \tau(\phi) = .8)\}$
- $r_{\tau(\gamma)} : \{claim(\gamma, A), , \phi \in Formulae(A) \rightsquigarrow \tau(\phi) = .7\}$

Notice how the *claim* at move 3 changes the context levels. Also note how this move should be contained in a level 3 meta knowledge base if the application of the update policy had to be formalized to allow observing moves involving meta-arguments.

Notice also how we must apply context changes to formulas rather than on arguments; the latter have to be determined by the evaluation in the Contextualized Knowledge Base otherwise non-admissible orderings in ASPIC+ argumentation theories may arise.

We assume that the $\mathcal{AS}_2$ of the Dialogical Trust-based Agent α contains $r_{\tau(\beta)}, r_{\tau(\gamma)}$ and that preferences are symmetric at this second level. We also assume that the agents has a prior preference on $\mathcal{A}_3$ higher than any argument using $r_{\tau(\gamma)}$ as an inference rule, this is to ensure that the conclusion of $\mathcal{A}_3$ is justified in the argumentation theory at the second level.

I will now highlight some of the conclusions α may draw at certain points in time. Notice first that the $pref$ values in CKB will be equal to the context level at each step i. Before move 3, α may derive the conclusion that it should watch both movies (there are no attacking arguments for β's recommendations). After move 4, α can form an argument whose conclusion sets the context of the formulae in $\mathcal{A}_1$ to .5. Notice how the context changes consequently. After move 5, γ claims $\mathcal{A}_5$ which attacks $\mathcal{A}_1$. Since the trust levels of the formulae of $\mathcal{A}_1$ have been reduced after the previous step, now we have $\tau_A(\mathcal{A}_1) = .5$ whereas $\tau_A(\mathcal{A}_5) = .7$. Hence after move 5 we have $\mathcal{A}_1 \prec \mathcal{A}_5$ and α may conclude that it should not watch Tetsuo. After move 6, γ attacks the other recommendation by β. This time, however, argument A_2 is considered more trustworthy than argument $\mathcal{A}_6$ because of the default context on them, thus it is $\mathcal{A}_6$ that would be defeated and α can still conclude that it should watch Blade Runner.

The example above shows some features of Contextualized Argumentation Theories as well as some limitations of Contextualized Knowledge Bases and the extensions of my formalism to meta-argumentation:

- Consistency with trust orderings and screened defeat may allow to challenge specific arguments in a predictive manner: if we know (as a property of the system) that agent β is not reliable on Japanese movies we are able to observe specific properties of the arguments from it. In this case, after step 3 we know that any arguments from β that advise japanese movies will be defeated by conflicting arguments with (strictly) higher reliability.

- Meta argumentation and contextualized revision allow specific (indirect) changes of preferences in arguments by describing context on their structural elements.

- Contexts may encapsulate specific properties of a MAS and decide how this influences the conclusive force of arguments; the example above shows this for trust but other properties are possible. For example, this approach may be used to define more sophisticated ways *subjective arguments* in [vdWD12] are affected.

- The example shows some limitations of the tranformer $\mathcal{T}_{hom}$: this transformer requires that the formulae in an argument all have the same context value, hence the structure of $r_{\tau(\beta)}$, $r_{\tau(\gamma)}$ and $\mathcal{A}_3$.

- It is not clear how the operators described in Section 4 may enable an agent to adopt prior preferences on arguments, that is preferences on arguments that are not acquired by other agents by contextualized revision. Moreover when context on some formula changes, the current formalism implicitely assumes that corresponding preferences in the object level CKB changes as well, whereas they may require a further revision.

- Another limitation is the need for a higher tower of argumentation systems [7]. In fact, in the example above, move 3 is assumed to update the argumentation system at the meta-level with levels of trust, but recall from Definition 42 that DTAs describe a tower of only size 2. This does not allow us to introduce an explicit observed claim in the knowledge base of the agent.

- Finally, we make no assumption on the preferences at the meta-level, that is on *how* the agents may actually reason on trust (the current framework only enables them *to* reason on trust). This issue will be addressed in the next subsection when we will see how agents may reason on trust being protected by certain types of naive manipulations.

5.4 Agents that are skeptical towards manipulations of trust

Earlier in this section, we saw how to allow agents to discuss about their trust levels and how we can use Contextualized Knowledge Bases to incorporate this information in the argumentation theories in order to guarantee properties of effective contextualized argumentation theories. The approach we followed was to define Dialogical Trust-based Agents, a notion that used a meta-argumentation theory to reason about terms at the object-level and in turn about context. In this subsection we will address some of the limitations of Dialogical Trust-based Agents and allow agents to protect themselves from manipulations of trust under certain circumstances.

The formalization we have given so far for dialogical agents shows a property which is undesirable in many dialogical settings: agents may claim trust levels (theirs or other agents') and the argument for this claim may end up being justified and actually change their trust levels. This happens because we make too few assumptions on the arguments at the meta-level. To exemplify this undesirable effect, let us assume that, at the end of the example in the previous subsection, agent β claims to α: "You can consider me very acknowledgeable on any topic" — thus conflicting with γ's statement at argument $\mathcal{A}_3$. One way to formalize β's claim is: $\mathcal{A}_\beta = \{claim(\beta, \underline{A}), \underline{\phi} \in Formulae(\underline{A}) \rightsquigarrow \tau(\underline{\phi}) = 1\}$. This argument would be added in the meta-knowledge base of a DTA (second case of Definition 43) and if there is no defeating argument, $\mathcal{A}_\beta$ would be justified. As a consequence τ_{ctx} would be modified accordingly (last bullet

[7] Tower of argumentation systems are a generalization of the approach used here with a single pair of meta-level/object-level argumentation systems. See [vdWD12] for further details.

of Definition 43), in this case enabling arguments from β to have full trust. We can see how our previous formalization of DTAs enables makes agents too naive: other agents would easily convince them that they are trustworthy and all properties such as screened defeat would offer no protection at all.

I shall now propose a way to tackle this issue. First, notice that this problem might have occurred if some other agent had claimed a high trust level for γ. Intuitively, we do not want this to occur, *unless* we trust significantly the agent who utters such a trust-manipulation argument. If an agent Ag claims that a trust level of another agent is t (where t is higher than our current beliefs on the trust levels), how much trust do we require from Ag in order to (possibly) believe its proposition? I propose that such an agent should be trusted at least t in order for its claim to be believed, or even considered, since it may always be overruled by other incoming information. This principle, that from now on I will call *manipulation-protection principle*, may be informally stated as:

The trust level τ_{Ag_1} attributed to an agent Ag_1 may be increased at τ'_{Ag_1} out of information by an agent Ag_2 only if $\tau_{Ag_2} \geq \tau'_{Ag_1}$.

Before I discuss how to formalize the intuitions behind such a principle, notice that the behavior exemplified by $\mathcal{A}_\beta$ may also occur if the latter were a strict argument, rather than a defeasible one. In the rest of this subsection we will discuss only the defeasible case, showing a solution for issues arising from it and will assume that strict arguments whose conclusions are of the type $\tau(\underline{\phi}, t)$ are forbidden, thus avoiding the issues arising in the strict case.

In the next paragraphs I will show how the properties (in particular Screened Defeat) of effective CATs may be used at the meta-level to tackle this problem. Before showing the formal solution I will discuss some of the observations that lead to it. Consider a formula ϕ, let $\tau_{cur}(\phi)$ be the current trust level we attach to ϕ; we aim at imposing that whenever an agent Ag_{Src} — to whose arguments we attach the trust level $\tau_{cur}(Ag_{Src})$ — utters an argument with conclusion "$\tau(\phi) = \tau_{new}$", this argument is not justified when $\tau_{cur}(Ag_{Src}) < \tau_{new}$, with $\tau_{new} > \tau_{cur}(\phi)$. One way we can do this is by having an (atomic) argument with conclusion $\psi_{\tau(\phi)} \equiv$ "$\tau(\phi) \leq \tau_{cur}(\phi)$" (read "the context of a formula ϕ cannot be higher than its current trust level") at the meta-level. Let us now observe how strong should such an argument be and consequently how to use it at the meta-level. We aim at having this argument protect us against trust manipulation, we may then think of using $\psi_{\tau(\phi)}$ as an axiom at the meta-level. However, such an approach would not be satisfying, since it would prevent *any* increase of trust levels: in fact we aim at protecting this only whenever the source of the information is not trustworthy enough (namely when it has a lower trust than the increased trust level). Hence we do not want $\psi_{\tau(\phi)}$ to be this strongly preferred. How strong should $\psi_{\tau(\phi)}$ be then? Notice that we only want $\psi_{\tau(\phi)}$ to defeat a contradictory conclusion when this conclusion is not trustworthy enough. But this is exactly what Screened Defeat does! (see subsections 3.2 and 3.3) In particular, we may achieve the manipulation-protection principle above if the property of screened defeat holds and we attach to $\psi_{\tau(\phi)}$ a context (trust) value strictly greater than $\tau_{cur}(\phi)$. Thus only arguments $\mathcal{A}$ such that

$\tau_{cur}(\mathcal{A})$ is greater than $\tau_{cur}(\phi)$ may be defeated by the atomic argument $\psi_{\tau(\phi)}$. This corresponds very closely to our principle above.

In the remainder of this section I will propose an improved definition of DTA that formalizes these observations. For this purpose I shall define the notion of meta-contextualized knowledge base, in order to deal with trust preferences on statements at the meta-level in particular with statements of the form $\psi_{\tau(\phi)}$ as above.

Definition 44 (Meta-Contextualized Knowledge Base)**.** *Let $CKB = (\mathcal{K}, \mathcal{R}, pref)$ be a Contextualized Knowledge Base on a logical language $\mathcal{L}$. Let $\mathcal{L}'$ be a meta-language on the basis of the argumentation system $(\mathcal{L}, \mathcal{R}^U, Not)$ (where $\mathcal{R}^U$ is a set of rules $\mathcal{R}^U$ s.t. $\mathcal{R}^U \subseteq \mathcal{R}$ and Not is some contrariness function on $\mathcal{L}$). Let $CKB' = (\mathcal{K}', \mathcal{R}', pref')$ be a Contextualized Knowledge Base on $\mathcal{L}'$, CKB' is a meta-Contextualized Knowledge Base on the basis of CKB if $\mathcal{K}'$ is a meta-knowledge base in $\mathcal{AS}(CKB') = (\mathcal{L}', \mathcal{R}', Not')$ for some contrariness function Not' on $\mathcal{L}'$,*

The assumption on $\mathcal{R}^U$ above is required because the set of rules may increase by evaluating arguments in the contextualized knowledge base. If all the possible arguments an agent can observe have rules in the set $\mathcal{R}^U$, we have the guarantee that the meta-language $\mathcal{L}'$ has a symbol $\underline{r}$ for any rule r an evaluated argument may have. [8]

The use of Meta-Contextualized Knowledge Bases will be shown with the next definition.

Definition 45 (Anti-manipulation Dialogical Trust-based Agent)**.** *An anti-manipulation DTA is a triple $(CKB_2, \tau_{ctx}, CKB_1)$ where:*

- *CKB_2 is a meta-contextualized knowledge base on the basis of CKB_1*
- *$\mathcal{L}'$, the (meta) language of CKB_2, contains the binary predicates τ and claim,*
- *τ_{ctx} is a context function s.t. it is defined on the members of $Dom(\tau_{ctx}) = Formulae(CKB')$,*
- *$\tau_{ctx}(\phi) = t$ iff $\tau(\phi, t)$ (i.e. "$\tau(\phi) = t$") is S-justified in the argumentation framework corresponding to the meta-argumentation theory $\mathcal{AT}(CKB_2)$ for some semantics S.*

The definition above uses two contextualized knowledge bases, one of which, CKB_2, may be used to reason on the formulae at the object-level and to define the context function τ_{ctx}. Notice that while only CKB_2 can modify τ_{ctx} (last bullet of definition above), this context function may be used for the evaluation of arguments in both CKBs. This will be shown in the update rule in Definition 47 and it is an important difference with Dialogical Trust-based Agents where the strength of the arguments at the meta-level had no specific constraints and was independent of the trust levels.

[8]Notice that a similar assumption on $\mathcal{R}(CKB)$ may be used to improve the definition of Dialogical Trust-based Agent given in the previous subsection.

Definition 46 (Trust protection statement)**.** *Let* $(CKB_2, \tau_{ctx}, CKB_1)$ *be an Anti-manipulation Dialogical Trust-based agent, let* $\phi \in \mathcal{L}$ *(defined as in Definition 44), the trust protection statement for* ϕ *w.r.t.* τ_{ctx} *is a formula* $\psi_{\tau\phi} \in \mathcal{L}'$ *(where* $\mathcal{L}' = \mathcal{L}(CKB_2)$*) s.t.* $\psi_{\tau(\phi)} = \tau(\phi) \leq \tau_{ctx}(\phi)$ *where* $\tau_{ctx}(\phi)$ *is the actual value of the context function and* τ *is the predicate in* $\mathcal{L}'$*.*

Definition 47 (Update rules for anti-manipulation agents)**.** *Let* $ADTA = (CKB_2, \tau_{ctx}, CKB_1)$ *be an anti-manipulation dialogical trust-based agent, then the agent is updated to* $ADTA' = (CKB_2', \tau_{ctx}', CKB_1')$ *where:*

- *If* $\mathcal{A}$ *is an object-level argument:*
  - $CKB_2' = eval(CKB_2, claim(\alpha, \underline{\mathcal{A}}), \tau_{ctx})$,
  - $CKB_1' = eval(CKB_1, \mathcal{A}, \tau_{ctx})$,
  - $\tau_{ctx}' = \tau_{ctx}$
- *If* $\mathcal{A}$ *is a meta-level argument (i.e. not an object-level argument):*
  - $CKB_2' = eval(CKB_2^{\psi}, \mathcal{A}, \tau_{ctx})$ *where:*
    * *Let* $Formulae(\mathcal{A}) = \{\phi_1, ...\phi_n\}$
    * $CKB_2^0 = CKB_2$
    * $CKB_2^i = eval(CKB_2^{i-1}, \psi_{\tau(\phi_i)}, \tau_{ctx}(\phi_i)\}, \tau_{ctx} + \epsilon)$ *with* $i \geq 1$*, where* $\psi_{\tau(\phi_i)}$ *is the trust protection statement for* ϕ_i *w.r.t.* τ_{ctx} *and* $\epsilon > 0$
    * $CKB_2^{\psi} = CKB_2^n$
  - $CKB_1' = CKB_1$
  - τ_{ctx}' *is such that* $ADTA'$ *is an anti-manipulation Dialogical Trust-Based Agent (i.e.* τ_{ctx}' *is defined as the last bullet in Definition 45).*

Although properties showing the usefulness of the definitions above shall not be formally proved [9], I will informally exemplify how the ideas in the previous paragraphs may be interesting. For instance, if agent α from the example in the last subsection observed the move $claim(\beta, \mathcal{A}_\beta)$ where $\mathcal{A}_\beta = \{claim(\beta, \underline{A}), \underline{\phi} \in Formulae(\underline{A}) \rightsquigarrow \tau(\underline{\phi}) = 1\}$ is the example argument at the beginning of this subsection, we could use the update rule in Definition 47: let us assume that $(CKB_2, \tau_{ctx}, CKB_1)$ represents the current state of α, then we would evaluate $\mathcal{A}_\beta$ by $CKB_2' = eval(CKB_2^{\phi}, \mathcal{A}_\beta, \tau_{ctx})$. Notice now that in the resulting meta-level argumentation theory, i.e. $\mathcal{AT}(CKB_2')$, the argument $\mathcal{A}_\beta$ cannot be justified (assuming a past history of dialogue moves as in the example in the previous subsection) since we added corresponding trust protection statements $\psi_{\tau(\phi)}$ [10]. These contradict the conclusion in $\mathcal{A}_\beta$ and their trust levels are strictly greater (see first set of sub-bullets in the meta-level case in Definion 47).

[9] Due to time constraints during the writing of this thesis

[10] This is true for any attempt of manipulation of this type as long as the assumptions of screened defeat are satisfied.

6 Argument Revision with Context

Here I will propose to extend Snaith's Argument Revision with concepts from contextualized revision.

In the previous sections I presented revision with context and how to formalize it using ASPIC+. In this section I will introduce an approach to belief revision in argumentation, called Argument Revision and I will propose how to extend this framework in order to include the use of properties of effective Contextualized Argumentation Theories.

The aim of Argument Revision [Sna12] is to guarantee a certain acceptability status of an argument in an argumentation theory. Argument Revision is divided into two broad processes: *expansion* and *contraction*. The two guarantee that arguments in some set are respectively acceptable and unacceptable in the resulting argumentation theory. This may have strategical applications in dialogical settings since it can be used for keeping consistency of commitment stores and for evaluating lying in dialogues.

Snaith presents an approach to argument revision by means of insertion/removal of formulas in a meta-argumentation theory. Since this modification can also take place at the meta-level the insertion/removal of formulas may not translate directly to adding/removing premises but to changes in the contrariness function, the type of inference rules, the addition/removal of inference rules and preferences on arguments. Given an argumentation theory and a set of arguments to expand/contract in it, there are several ways this can be done performing a sequence of operations like the ones mentioned above. Snaith propose a cost function based on structural and acceptability changes in the Argumentation Theory to minimize when choosing among the possible ways of revising.

In general, the approach to argument revision in Snaith's work is "destructive". If for instance we want to contract an argumentation theory by a set of arguments S, this can happen by increasing the preference over the defeaters, but may as well happen by means of the total removal of the arguments in S from the argumentation theory (e.g. by removing some of the premises of the arguments in S). If one applies Argument Revision to a Contextualized Argumentation Theory, it may happen that some of the desiderata on it such as the ones in Section 5 may not hold any more after an operation of expansion or contraction. My aim here is to modify Snaith's approach to Argument Revision in order to make expansion and contraction satisfy the desiderata in the previous sections. I propose to achieve this goal by imposing constraints on the operations performed according to some of the (contextual) properties of arguments.

The original description of Argument Revision in [Sna12] is not totally formalized and presents several flaws. In this work I will tackle some of these aspects described further in subsubsection 6.1.5.

Summarizing, the contributions of this section are twofold: enhancing the formalization by Snaith of Argument Revision and propose measures of impact on the base of properties of Contextualized Argumentation Theory and ways to apply these measures in argument revision and, in turn, in dialogical settings.

6.1 Preliminaries on Argument Revision

In this subsection I will introduce the inner mechanics of Argument Revision. I will first formally introduce expansion and contraction, then I will describe the basics of the actual process: formula addition and removal. The latter can be used to affect rules, preferences, contrariness and strictness is performed at the meta level. Some of the ideas behind the definitions and properties in this subsection are presented originally in [Sna12].

6.1.1 Principles of Argument Revision

Argument revision (AR) is divided into two broad processes: *contraction* and *expansion*. In both cases the goal relate to the acceptability of arguments in the resulting Argumentation Theory. As in classic AGM Theory [Gär88] there is no unique way to specify AR functions.

Definition 48 (Contraction)**.** *Given an Argumentation Theory $\mathcal{AT} = (\mathcal{AS}, \mathcal{K}, \leq)$ and a set of arguments $\mathcal{S} \subseteq Args(\mathcal{AS})$, $\mathcal{AT} \dot{-} \mathcal{S}$ is a maximal (with respect to set inclusion) set of argumentation theories where $\forall \mathcal{AT}' \in \mathcal{AT} \dot{-} \mathcal{S}$ it holds that*

$$\forall \mathcal{A} \in \mathcal{S}, \mathcal{A} \notin E(\mathcal{AT}')$$

We refer to $\mathcal{AT}^-_S$ as a contraction *of $\mathcal{AT}$ by $\mathcal{S}$.*

Definition 49 (Expansion)**.** *Given an Argumentation Theory $\mathcal{AT} = (\mathcal{AS}, \mathcal{K}, \leq)$ and a set of arguments $\mathcal{S} \subseteq Args(\mathcal{AS})$, $\mathcal{AT} \dot{+} \mathcal{S}$ is a maximal (with respect to set inclusion) set of argumentation theories where $\forall \mathcal{AT}' \in \mathcal{AT} \dot{+} \mathcal{S}$ it holds that*

$$\forall \mathcal{A} \in \mathcal{S}, \mathcal{A} \in E(\mathcal{AT}')$$

. We refer to $\mathcal{AT}^+_S$ as an expansion *of $\mathcal{AT}$ by $\mathcal{S}$.*

Informally, the definitions of expansion and contraction above, describe a set of argumentation theories such that the arguments in the revising set are respectively justified or overruled.

If it is not necessary to know what type of argument revision we are referring to, we use the notation $\mathcal{AT}^\pm_S$ to refer to an expansion or contraction of $\mathcal{AT}$ by $\mathcal{S}$.

6.1.2 The process of Argument Revision

In the remainder of this subsection I will describe how to achieve the contraction or expansion of an argumentation theory. We will loosely follow the same general idea described in Snaith [Sna12], that is, starting from an argumentation theory $\mathcal{AT}$ and in order to get $\mathcal{AT}^\pm_S$:

- define a notion of atomic change on the argumentation theory (*formula removal* and *formula addition*, see next paragraphs),

- describe all the possible sequence of changes by means of a *change graph*, where vertices are argumentation theories and there is an arc from $\mathcal{AT}_1$ to $\mathcal{AT}_2$ iff there exists a formula addition/removal change that transforms $\mathcal{AT}_1$ into $\mathcal{AT}_2$[11] (see next pararaphs),

- pick a path in the change graph from $\mathcal{AT}$ such that (*i*) the final vertex of the path is an argumentation theory in the set $\mathcal{AT}^{\pm}_{S}$, and (*ii*) the path satisfies certain constraints (e.g. minimality).

The process of change of an argumentation theory is based on modifications of the underlying argumentation system and knowledge base in the theory, so to add or remove arguments or to change their acceptability status (thus achieving expansion or contraction).

Since we are using meta-argumentation, adding or removing arguments can be achieved by adding and removing formulas from a meta-knowledge base. We will see how this enables us to obtain changes in rules, preferences and contrariness in the argumentation theory: recall that for each of these there is a corresponding (possibly atomic) meta-argument for it in the meta-level knowledge base.

The ultimate goal of AR is to guarantee a certain acceptability status for arguments. But what are the possible ways changes in acceptability can occur? Snaith identifies three:

- Adding a new argument defeater of the original argument;

- Removing an argument which acts as a defender or defeater of the original argument;

- Revising the meta-arguments for contrariness and preference relations, which involve the original argument.

Snaith proposes to achieve the changes above by modifying an argumentation system by adding and removing formulas, thus yielding new argumentation theories. The general idea of the AR process is then that of a *search* through the space of the possible Argumentation Theories obtainable by this set of operations, the *change graph*.

Once described all the possible changes obtainable from the starting argumentation theory, we may choose a sequence of changes that leads to one revised argumentation theory in $\mathcal{AT}^{\pm}_{S}$. There may be several of such sequences in the change graphs, thus in [Sna12] it is proposed to choose a shortest path with respect to a certain measure of minimal change. In this thesis we will perform this last step in a different fashion. We will still define a measure of impact for each argumentation theory or for each change but with two main differences:

[11]Note that the formal definition of change graph in the next paragraphs will be slightly diffent from this, since the actual vertices will be Argument Revision Bases (defined in this subsection) and not argumentation theories. However, informally speaking, this description and the actual definition are equivalent and share the same intuitions. Note that also the actual definition of formula addition/removal will be on Argument Revision Bases rather than argumentation theories.

- these measures will be based on properties of contextual information, rather than on structural properties as in [Sna12],
- I will propose additional criteria than shortest paths to choose among different paths according to their measures of impact.

6.1.3 Adding and removing formulas

Before introducing the change graph, I will formally describe the ways we can modify an argumentation system during the AR process.

Starting from a (meta) argumentation theory $\mathcal{AT}$, given a set of arguments S, we aim at defining changes on it that would lead us to $\mathcal{AT}'$ with $\mathcal{AT}' \in \mathcal{AT}^{+}_{S}$ ($\mathcal{AT}^{-}_{S}$). Changes on argumentation theories will be defined as changes on a description of theirs that makes use of meta-argumentation. For this purpose in the following paragraphs I will introduce argument revision bases on which actual notions of atomic change will be defined: formula addition and removal.

Argumentation systems for Argument Revision are instances of object-level argumentation systems (argumentation systems on the basis of which we assume a meta-argumentation system $\mathcal{AS}'$) that are described by a knowledge base at the meta-level. The way these argumentation systems are described is given by the notion of constructive conclusion. These concepts are inspired by the definition of meta-argumentation system in [Sna12].

Definition 50 (Constructive conclusion)**.** *Given an argumentation system $\mathcal{AS}$ and a knowledge base $\mathcal{K}$ in $\mathcal{AS}$, a formula $\phi \in \mathcal{L}(\mathcal{AS})$ may be constructively concluded in $(\mathcal{AS}, \mathcal{K})$ if there exists an argument $\mathcal{A} \in Args(\mathcal{AS})$ s.t. $Premises(\mathcal{A}) \subseteq \mathcal{K}$ and $Conc(\mathcal{A}) = \phi$*

Definition 51 (Argumentation system for Argument Revision)**.** *Let $\mathcal{AS}'$ be a meta-argumentation system on the basis of $\mathcal{AS}$ and let $\mathcal{K}'$ be a knowledge base in $\mathcal{AS}'$. $\mathcal{AS}$ is an argumentation system for Argument Revision w.r.t. $\mathcal{AS}'$ and $\mathcal{K}'$:*

- *$r \in \mathcal{R}$ iff $in(\underline{r}, \underline{\mathcal{R}(\mathcal{AS})})$ may be constructively concluded in ($\mathcal{AS}'$, $\mathcal{K}'$)*
- *$\phi \in Not(\psi)$ iff $in(\underline{\phi}, \underline{Not}(\underline{\psi}))$ may be constructively concluded in ($\mathcal{AS}'$, $\mathcal{K}'$)*

The definition above says that, in an Argumentation system for Argument Revision, contrariness and rules are determined by the existence of an argument at the meta-level whose conclusion represents respectively a contrariness relation or a rule.

The actual process of Argument Revision will be defined on Argument Revision Bases. Their definition resembles the definition of argumentation theories in [vdWD12, Sna12].

Definition 52 (Argument Revision Base)**.** *An Argumentation Revision Base (ARB) is a tuple $(\mathcal{AS}', \mathcal{AS}, \mathcal{K}', \mathcal{K})$ where:*

- *$\mathcal{AS}$ is an argumentation system for argument revision w.r.t $\mathcal{AS}'$ and $\mathcal{K}'$,*
- *$\mathcal{K}'$ is a meta-knowledge base in $\mathcal{AS}'$ on the basis of $\mathcal{K}$.*

Given an Argument Revision Base, we define the Argumentation Theory described by it.

Definition 53 (Argumentation theory w.r.t. an Argument Revision Base)**.** *Let $ARB = (\mathcal{AS}', \mathcal{AS}, \mathcal{K}', \mathcal{K})$ be an argument revision base. An argumentation theory w.r.t. ARB, denoted with $\mathcal{AT}(ARB)$, is the object-level argumentation theory based on $\mathcal{AT}'$, where $\mathcal{AT}'$ is a meta-argumentation theory w.r.t. $\mathcal{AS}'$ and $\mathcal{K}'$.*

Definition 54 (Formula removal function)**.** *Let $ARB_0 = (\mathcal{AS}', \mathcal{AS}_0, \mathcal{K}'_0, \mathcal{K}_0)$ be an argument revision base. Let $\phi \in \mathcal{L}(\mathcal{AS}')$:*

$$ARB_0 - \phi = (\mathcal{AS}', \mathcal{AS}_1, \mathcal{K}'_1, \mathcal{K}_1)$$

where:

- $\mathcal{K}'_1 = \mathcal{K}'_0 \setminus \{\phi\}$
- *$\mathcal{AS}_1$ is an argumentation system for argument revision w.r.t. to $\mathcal{AS}'$ and $\mathcal{K}'_1$ and $\mathcal{L}(\mathcal{AS}_1) = \mathcal{L}(\mathcal{AS}_0)$, and*
- *$\mathcal{K}_1$ is such that $\mathcal{K}'_1$ is a meta-knowledge base in $\mathcal{AS}'$ on the base of $\mathcal{K}_1$*

What the rule above does is to remove a formula from the meta-knowledge base in the argument revision base. The purpose of this is to modify the underlying argumentation systems for argument revision $\mathcal{AS}$ (by changing its contrariness function $Not(\mathcal{AS})$ or the set of rules $\mathcal{R}(\mathcal{AS})$) or the knowledge base $\mathcal{K}$ (by removing a predicate $op(\cdot)$ or $np(\cdot)$). This is why $\mathcal{AS}_1$ and $\mathcal{K}_1$ are needed to replace their old counterparts. Notice that the argumentation system for argument revision $\mathcal{AS}_1$ is uniquely determined by $\mathcal{AS}', \mathcal{K}'_1, \mathcal{L}(\mathcal{AS}_0)$, while the knowledge base $\mathcal{K}_1$ is uniquely determined by $\mathcal{K}'_1$. Other types of changes for the purpose of Argument Revision may occur too, for example changes in ordering. A schematic view of the possible changes with be provided in definition 57. Notice that the result of removing a formula from an argument revision base is an argument revision base.

Similarly, formula addition is defined as follows:

Definition 55 (Formula addition function)**.** *Let $ARB_0 = (\mathcal{AS}', \mathcal{AS}_0, \mathcal{K}'_0, \mathcal{K}_0)$ be an argument revision base. Let $\phi \in \mathcal{L}(\mathcal{AS}')$:*

$$ARB_0 + \phi = (\mathcal{AS}', \mathcal{AS}_1, \mathcal{K}'_1, \mathcal{K}_1)$$

where:

- $\mathcal{K}'_1 = \mathcal{K}'_{1,p} \cup \mathcal{K}'_{1,n}$
  - $\mathcal{K}'_{1,n} = \mathcal{K}'_{0,n}$

- $\mathcal{K}'_{1,p} = \mathcal{K}'_{0,p} \cup \{\phi\}$

- *$\mathcal{AS}_1$ is an argumentation system for argument revision w.r.t. to $\mathcal{AS}'$ and $\mathcal{K}'_1$ and $\mathcal{L}(\mathcal{AS}_1) = \mathcal{L}(\mathcal{AS}_0)$, and*

- *$\mathcal{K}_1$ is such that $\mathcal{K}'_1$ is a meta-knowledge base in $\mathcal{AS}'$ on the base of $\mathcal{K}_1$*

For sake of notation ease, I will introduce formula removal/addition in argumentation theories. In order to do that we assume, in the remainder of this section, that each argumentation theory $\mathcal{AT}$ has a corresponding argument revision base ARB s.t. $\mathcal{AT} = \mathcal{AT}(ARB)$. We denote this argument revision base by $ARB(\mathcal{AT})$. This assumption is plausible since, when carrying out AR on an argumentation theory $\mathcal{AT}$, we will always deal with argumentation theories obtained by modifications of argument revision bases.

Definition 56 (Formula removal/addition in an Argumentation Theory)**.** *Let $\mathcal{AT}$ be an argumentation theory and let $ARB = ARB(\mathcal{AT})$ be its corresponding argument revision base, then*

$$\mathcal{AT} - \phi = \mathcal{AT}(ARB - \phi)$$

$$\mathcal{AT} + \phi = \mathcal{AT}(ARB + \phi)$$

I mentioned earlier in this section how one of the advantages of using meta argumentation for AR was that it would allow us to specify a unique method of revision. In fact, AR simply uses formula addition and removal to deal with several types of changes. According to the type of formula we are adding or removing we may obtain different types of changes:

Definition 57 (Types of formula change)**.** *Given an argumentation theory $\mathcal{AT}$ and a formula ϕ, we call the operation $\mathcal{AT} - \phi$ ($\mathcal{AT} + \phi$):*

- premise-based Argument Revision*, if $\phi = np(\underline{\psi})$ or $\phi = op(\underline{\psi})$ for some formula ψ;*

- preference-based Argument Revision*, if there exists an argument $\mathcal{A} \in E(\mathcal{AT})$ ($E(\mathcal{AT} + \phi)$) s.t. $\phi \in Prem(\mathcal{A})$ and $Conc(\mathcal{A}) = \underline{A_1} \preceq_{Args} \underline{A_2}$ for some arguments A_1 and A_2;*

- contrariness-based Argument Revision*, if there exists an argument $\mathcal{A} \in \mathcal{AT}$ ($\mathcal{AT}+\phi$) s.t. $\phi \in Prem(\mathcal{A})$ and $Conc(\mathcal{A}) = Not(\underline{\psi})$ for some formula ψ;*

- rule-based Argument Revision*, if there exists an argument $\mathcal{A} \in \mathcal{AT}$ ($\mathcal{AT}+\phi$) s.t. $\phi \in Prem(\mathcal{A})$ and $Conc(\mathcal{A}) = in(\underline{r}, \mathcal{R}(\mathcal{AS}))$ for some rule r.*

The taxonomy above is not provided to guide the process of Argument Revision but rather to formally describe the type of changes (in the sense of formula addition/removal) that define a significant impact on an argumentation theory, that is changes that may lead to drop or gain of constructible or acceptable arguments, as discussed in subsubsection 6.1.2.

These definitions suggest how we can use addition/removal of formula on an argumentation theory to obtain changes that involve not only premises, but also rules, contrariness and preferences. Premises and rules determine which arguments are constructible, contrariness determines attack and preferences determine defeat among conflicting arguments. Notice that the structure of the types of changes above differs: premise-based AR depends only on the removal/addition of a particular type of formula (i.e. a formula describing the object-level knowledge base), preference-based AR depends on the *drop/gain in acceptability* of arguments with a certain type of conclusion (i.e. a conclusion on argument ordering), contrariness and rule-based AR depend on the *mere existence* of arguments with a certain type of conclusion (i.e. a conclusion describing respectively contraries and rules in the object level argumentation system). The limitations and the rationale of these definitions will be discussed further in the discussion at the end of this subsection.

6.1.4 Change graphs and minimal impact

The two functions for formula addition and removal above are used to model expansion and contraction of Argumentation Theories by a structure called *change graph.*

The revision of an Argumentation Theory is carried out by modifying knowledge bases in the argumentation systems within the theory, with modifications continuing to take place untile the properties of the chosen revision (i.e. revision or contraction) are satisfied. One such sequence of modifications can be seen as a path from the initial Argumentation Theory to another that has been revised with respect to the set of arguments in input. A change graph models the possible ways we can either expand or contract an argumentation theory.

Definition 58 (Difference set between argumentation theories)**.** *Let $\mathcal{AT}_1, \mathcal{AT}_2$ be argumentation theories, let $ARB_1 = ARB(\mathcal{AT}_1), ARB_2 = ARB(\mathcal{AT}_2)$, then the difference set between $\mathcal{AT}_1$ and $\mathcal{AT}_2$ is given by*

$$\mathcal{AT}_1 \Delta \mathcal{AT}_2 = \mathcal{K}'(ARB_1) \Delta \mathcal{K}'(ARB_2)$$

where Δ on the right-hand side is the symmetric difference between sets.

Definition 59 (Change graph)**.** *A change graph $CG(\mathcal{AT}, \Pi')$ for the revision of $\mathcal{AT}$ to a set of argumentation theories $\Pi' = \{\mathcal{AT}^{\pm}_{S,1}, ..., \mathcal{AT}^{\pm}_{S,n}\}$ is a directed acyclic graph (Υ, Ω) where:*

- $\Upsilon \subseteq \Pi$ *(where Π is the set of all possible argumentation theories),*
- *(Atomic change) $\Omega \subseteq \Upsilon \times \Upsilon$ where $\forall \omega \in \Omega$ such that $\omega = (\mathcal{AT}', \mathcal{AT}'')$, we have that $|\mathcal{AT}'' \Delta \mathcal{AT}'| = \pm 1$,*
- $\Pi' \subseteq \Upsilon$ *is such that $\forall \mathcal{AT}' \in \Pi'$, $\nexists \mathcal{AT}'' \in \Pi', \mathcal{AT} \Delta \mathcal{AT}'' \subset \mathcal{AT} \Delta \mathcal{AT}'$.*

The second bullet defines the arcs of the graph as the ones between argumentation theories such that one is obtained as the result of formula addition/removal from the other and this results in a change in the corresponding meta-knowledge base; however, this constraint is not expressed merely as "$\mathcal{AT}'' = \mathcal{AT}' \pm \phi$ for some ϕ" because what we are actually aiming for is a stronger property: that the formula addition/removal leads to a different argumentation theory. In fact, if the constraint were not expressed as in the definition but as above, there could exist an arc from an argumentation theory $\mathcal{AT}'$ to itself since $\mathcal{AT}' = \mathcal{AT}' - \phi$, $\forall \phi \notin \mathcal{K}'(ARB(\mathcal{AT}'))$.

The third bullet rules out from the set of the vertices any of the argumentation theories in Π' (the set of target revised argumentation theories) for which there exists another revised argumentation theory which is "less distant" from the starting one ($\mathcal{AT}$). This constraint is mostly for pruning purposes.

One of the aims of Argument Revision is to provide a way to choose a revised argumentation theory that has the *least impact* among the possible ones (this is close to the concept of entrenchment in AGM). Once we have defined the ways we can modify argumentation theories as paths in a graph we can define impact as weights on the edges of the change graph. Snaith [Sna12] provides a definition of impact that takes into account structural properties of arguments. These measures of impact (of which I will not provide a formal definition) are: argument gain, argument loss (which constructable arguments have been gained/lost), acceptability gain, acceptability loss (which acceptable arguments have been gained/lost). In [Sna12] a revised Argumentation Theory has undergone minimal impact if the sum of the measures above on the change graph path is minimal. We will not use any of these measures in this work and in the remainder of this section we will follow a different approach. In the next subsections I will describe measures of minimal impact that define an entrenchment based on context values. Before then, I will provide a short discussion of some of the changes in the original work by Snaith.

6.1.5 Limitations of Snaith's meta-argumentation approach to Argument Revision

The framework presented above is based on the original work by Snaith in [Sna12]. In the next paragraphs I will outline some of the differences of my approach with the original presentation of Argument Revision and some of the current limitations in both formalizations.

One of the main differences in the definition of Argument Revision is in the assumed framework of meta-argumentation. Snaith proposes a variant of the definition of meta-argumentation system provided in [vdWD12] where contrariness, preferences and rules in the argumentation system at the object level depend on the existence of an argument at the meta-level (similarly to the approach to contrariness and rules in the current thesis) and preferences over arguments is not expressed by the introduction of a predicate $\preceq_{Args}$ at the meta-level but rather by atomic formulae (I will explain below why such a thing represents a limitation). Snaith also makes use of towers of argumentation of arbitrary size [vdWD12]. On the other hand, my approach relies on the meta-argumentation framework presented in [vdW11]. I restrict my attention to towers of argumentation systems of size 2, since these show to be sufficiently powerful for most of the applications this thesis is concerned with. Ordering on arguments is here expressed as a predicate at the meta-level and axioms for ensuring admissible orderings are provided in the meta-knowledge bases. The fact that these axioms are not present in Snaith's definitions makes it unclear to see how admissible orderings can be achieved. In fact, in his thesis it seems to be suggested that only argumentation theories which, out of the revision process, show admissible orderings would be considered in the change graph. However, even if this were the case, it would lead to an explosion in the size of the change graph because of the presence of many argumentation theories with ill-defined orderings. Thus ensuring an admissible ordering would be a significant enhancement; this is achieved in the current thesis: in my formalization of argumentation revision I use object level argumentation theories in the sense of Definition 41 where the ordering is derived from the justified conclusions at the meta-level and it is ensured to be admissible because of the axioms in the meta-knowledge bases.

Also, in the current work, this ordering comes from the argumentation semantics (at the meta-level) and does not depend merely on the existence of arguments whose conclusion express preferences (over rules, premises or arguments). This allows to solve some limitations in the original formalization of Argument Revision; since these limitations are still present in this thesis for rules and contrariness, I will clarify them using rules as an example: consider Definition 51 (Argumentation system for argument revision), if $\mathcal{AS}$ is an argumentation system for argument revision, then r is a rule in $\mathcal{R}(\mathcal{AS})$ iff there exists an argument at the meta-level having $\phi_r = in(\underline{r}, \underline{\mathcal{R}(\mathcal{AS})})$ as a conclusion. Let us now assume that the arguments $\mathcal{A}_1$ and $\mathcal{A}_2$ are constructible at the meta-level with $\mathcal{A}_1 = (\{\psi, \psi \rightsquigarrow \phi_r\}, \phi_r)$ and $\mathcal{A}_2 = \neg\phi_r$. Given these two arguments, we satisfy the requirements of Definition 51 and thus $r \in \mathcal{R}(\mathcal{AS})$. However, if $\mathcal{A}_2$ defeated $\mathcal{A}_1$ in the meta-level argumentation theory, it would still be the case that $r \in \mathcal{R}(\mathcal{AS})$ despite ϕ_r being overruled (at least this is the case if no other argument can reinstante it). This is counterintuitive because it implies that defeat and argument semantics play no role for argument revision. In the formalization I propose in this thesis, this problem is solved for argument ordering, but my definitions still make use of constructive conclusions (see Definition 50) for rules and contrariness. Future work should address such an issue. One possible way to tackle this limitation would be to change the definition of argumentation systems for argument revision bases so that rules and contrariness relations depend on justified conclusions at the meta-level.

The definition of change graphs has also undergone modifications from its original formalization since it is now based on Argument Revision Bases — this notion is not present in Snaith's work.

Formula removal and addition have been made a bit more precise and at the same time simplified: now they do not require to operate at different meta-levels of knowledge bases, whereas in [Sna12] this occurred even in the case of towers of meta-argumentation systems of size 2.

One final enhancement in the formalization in this thesis in comparison to that in [Sna12] is that the process of argument revision has been made a bit more precise by introducing a formal definition of type of changes and by explicitely stating the use of the semantics of meta-argumentation. In [Sna12] Extended Argumentation Frameworks [Mod09] seem to have been proposed for this purpose but the way they are used in the Argument Revision framework is not formalized nor clear. In this work, the definition of object-level argumentation theory (see Definition 41) [vdW11] makes explicit how meta-level and object-level are connected by argumentation semantics. In turn this solves another issue in Snaith's work: in fact in [Sna12] the definition of argument expansion and contraction require explicitly that the revised argumentation theory is well-formed [Pra10]. As shown in [vdW11] Definition 41 ensures that the resulting argumentation theories are well-formed and it is not necessary to include this constraint in the definition of argument revision.

6.2 Differences between Argument Revision and Contextualized Revision and its properties

The following is a short discussion of the differences between Argument Revision and Contextualized Revision with the three desired properties of effective Contextualized Revision (namely Total Memory, Consistency with external ordering and Screened Defeat).

Despite the existence of several differences between Argument Revision and Contextualized Revision, there are also important similarities: both aim at establishing logical properties of arguments and both modify elements of arguments to do that, they are both inspired by Belief Revision.

6.2.1 Goals

First, let us recall again what are the main general goals of Argumentation Revision and Contextualized Argumentation Theories. Argument Revision is about accepting/unaccepting arguments and it does this by means of changing other arguments in the Knowledge Base. Contextualized revision, on the other hand, injects into an Argumentation Theory parameters on arguments. These parameters are external to the Argumentation Theory itself and are used to enable arguments with certain properties to be able to defeat others (Consistency with context ordering) and not to be defeated by others (Screened defeat). Moreover, no argument in the Knowledge Base is removed to ensure these properties (Total memory property).

Expanding an Argumentation Theory with a set of arguments can guarantee those arguments are acceptable whereas revising in a Contextualized Argumentation Theory does not guarantee that, even if an argument has the maximum context value in the CAT. On the other hand when extending the Knowledge Base nothing guarantees that arguments made acceptable by expansion keep being acceptable, whereas this can be assured to some extent in Contextualized Argumentation Theories (namely when no revision with argument with higher context occurs). Also, in Contextualized Revision strong (context-wise) arguments can be enabled to defeat weaker ones that may threaten them in the future. Conversely, Argument Revision does not allow anything like that.

Summarizing, Argument Revision can guarantee the acceptability or overruling of arguments, whereas Contextualized Revision does not. Also, Contextualized Revision can tell us something about the future of the interactions with other arguments whereas Argument Revision can deal only with the present: arguments are made acceptable immediately and without concern for the future.

6.2.2 Approach

Another important difference is in the way the two types of revision achieve their goals. In Argument Revision an Argumentation Theory is modified by adding and removing structural parts of arguments themselves so that the set of arguments we expand with becomes acceptable. In Contextualized Revision[12] we only change preferences in the underlying argumentation system, without changing other parts. In this sense Argument Revision has a more general type of change than Contextualized Revision.

Another important difference however is how changes are directed: in Contextualized Revision (by Contextualized Knowledge Bases) there is a single set of changes in preferences that has to be carried out to achieve desired Properties on CATs. Also, these changes strictly depend on the context of the argumentation theory. In Argument Revision they are not deterministic: the result, whether it be expansion or contraction, can be achieved with different sets of changes and the only criteria to choose them is minimal change as discussed in Section 6.1.

6.2.3 What Contextualized Revision can do for Argument Revision

Part of the rationale behind the definition of minimal change in Argument Revision is to provide a way to determine an entrenchment of arguments based only on acceptability criteria without any predetermined entrenchment. In settings where forms of preferences and entrenchment are available however, Argument revision may still be used and these preferences may guide the processes of expansion and contraction. Contexts are additional information on arguments, as such they can be used to define a notion of minimal change.

In the remainder of this section I will extend AR on Contextualized Argumentation Theories and present concepts of minimal change based on properties of effective CATs.

6.3 Guiding Argumentation Revision with CATs

The following is an extension of Argument Revision to Contextualized Argumentation Theory:

[12]here I am referring to the instantiation of CAT presented in this work by means of CKBs

Definition 60 (Expansion for Contextualized Argumentation Theories). *Given a Contextualized Argumentation Theory* $CAT = (Ctx_A, \mathcal{AT})$ *and a set of arguments* S *the expansion of* CAT *by* S *is a new Contextualized Argumentation Theory* $CAT' = CAT\dot{+}S = (Ctx_A, \mathcal{AT}')$ *where* $\mathcal{AT}' \in \mathcal{AT}\dot{+}S$.

Contraction may be defined in a similar way.

If, in the definition above, the Contextualized Argumentation Theory is effective, the process of expansion or contraction may make it not be effective anymore. Or there can be other changes that may make the resulting argumentation theory less desirable with respect to the context values. From this we may define a measure of the impact of the changes.

There are two main ways this can happen. Preference-based revision (see Definition 57) may cause preferences to change; this may happen in ways that violate the properties of effective CATs. The other types of formulae addition/removal in Definition 57 cannot violate properties of CATs. However, they may also remove trustworthy arguments/formulas.

Therefore, if we are expanding/contracting a Contextualized Argumentation Theory we may end up "harming" a CAT in two ways:

- violated Screened Defeats and Consistency with Ordering
- removal of trusted arguments/formulas

The first case occurs whenever the resulting Argumentation Theory has preferences between arguments that ensure that properties of effective CATs hold (see Definition 32) and these are changed by prefererence-based revision. The second case occurs whenever a trusted formula is removed by premise or rule-based revision.

I propose that the observations above are used to define functions that in turn will be used to guide the Argument Revision process. Thus I will now introduce related cost functions and then I will describe two approaches to define minimal impact by them.

Cost functions

The cost functions in this part will be defined combining several cases of "undesired" consequences in Contextualized Argumentation Theories. Recall that an Argumentation Revision process works by iteratively applying cahnges to an argumentation theory thus obtaining a change graph, where modified argumentation theories are nodes in the graphs. In the original work by Snaith [Sna12] minimal impact is defined by providing each edge in the change graph with a weight, thus reducing the problem to finding a minimum path in the change graph. In the following paragraphs I will describe how to apply the cost functions defined in this subsubsection; the approach taken in this thesis will be different from Snaith's minimum paths.

In the previous paragraphs we saw that Contextualized Argumentation Theories can be impacted by changes in preferences during the revision process.

Thus, one parameter to look at is how many of these changes occur. This justifies the following definition:

$$Violations_{SD}(\tau_A, \mathcal{AT}) = \{(A_1, A_2) : \underline{A_1} \prec \underline{A_2}\rceil \in \mathcal{K}', \tau_A(A_2) < \tau_A(A_1)\}$$

where $\mathcal{K}'$ is the meta-level knowledge base in $ARB(\mathcal{AT})$ (see Definition 52).

The set above represents the pairs of arguments where A_1 has a relative strength higher than A_2 despite $\mathcal{T}_A(A_1) < \mathcal{T}_A(A_2)$, that is those pairs of arguments that violate the property of Screened Defeat in the Contextualized Argumentation Theory.

Another case is when trustworthy arguments get "weakened" because the changes during the Argumentation Revision process may remove their preferences over less trustworthy arguments. This justifies the following definition:

$$Violations_{CTO}(\tau_A, \mathcal{AT}) = \{(A_1, A_2) : A_1, A_2 \in Args(\mathcal{AS}), \tau_A(A_2) \leq \tau_A(A_1) \wedge \underline{A_1} \preceq \underline{A_2} \notin \mathcal{K}'\}$$

where $\mathcal{K}'$ is defined as above and $\mathcal{AS}$ is thhe object-level argumentation system in $ARB(\mathcal{AT})$.

These two functions above return set of pairs of arguments, we may use them to define a cost measure for an argumentation theory by the following:

$$V_v(\tau_A, \mathcal{AT}, \mathcal{AT}\prime) = |Violations_{SD}(\tau_A, \mathcal{AT})| + |Violations_{CTO}(\tau_A, \mathcal{AT})|$$

I will now propose another function, V_{ctx} to measure the impact of expansion or contraction which is deafined on pairs of argumentation theories $(\mathcal{AT}, \mathcal{AT}')$, thus giving a form of distance between argumentation theories. This new cost function is specific to trust as a context for it makes assumptions on the bounds of the context, however it can be generalized to any other bounded context function. This function V_{ctx} uses context levels of specific formulas: let us assume that we are carrying out a removal of a formula ϕ in an argumentation theory $\mathcal{AT}$. Let us also assume that we have a context value $\tau(\phi)$ associated with the formula. If we remove the formula ϕ we are giving away information that may be trustworthy. At the same time if we carry out a change that introduce at new formula ϕ, the change may be less advantageous if ϕ is not very trustworthy. The function V_{ctx} expresses these intuitions:

$$V_{ctx}(\tau_A, \mathcal{AT}\,\mathcal{AT}\prime) = \sum_{A \in DA(\mathcal{AT}, \mathcal{AT}\prime)} \tau_A(A) + \sum_{A \in DA(\mathcal{AT}\prime, \mathcal{AT})} (1 - \tau_A(A))$$

where DA returns the differences in arguments between the two argumentation theories: $DA(\mathcal{AT}, \mathcal{AT}\prime) = E(\mathcal{AT}) \setminus Args(\mathcal{AT}\prime)$. Intuitively we sum the trustworthiness of the lost arguments and the non-trustworthiness of the arguments we acquired. The function above may be considered a weighted version of the argument drop function in [Sna12].

These two functions define a cost vector $CV(\mathcal{AT}, \mathcal{AT}') = (V_v, V_{ctx})$.

The following example is adapted from [Sna12] and shows the use of these functions.

Example 3. *Let us consider a Contextualized Argumentation Theory CAT. For sake of concreteness let us assume that $CAT = CAT(CKB)$, where CKB is the Contextualized Knowledge Base below (CKB may have been obtained by repeatedly evaluating formulae in an empty CKB as in Example 1):*

- $\mathcal{K}(CKB) = \{a_1, a_2, b_1, b_2, c_1\}$
- $\mathcal{R}(CKB) = \{r_1 : (a_1, a_2 \rightsquigarrow a), r_2 : (a_1 \rightsquigarrow x), r_3 : (a_2 \rightsquigarrow y), r_4 : (b_1, b_2 \rightsquigarrow b), r_5 : (c_1 \rightsquigarrow c)\}$

We also have that $a_2 \in \bar{b_1}$ *and* $b \in \bar{c}$ *and the context is given by* $\tau = \{(a_1, .2), (a_2, .6), (c_1, .4), (b_1, .6), (b_2, .2), (r_1, .6), (r_2, .5), (r_3, .9), (r_4, .8), (r_5, .8)\}$

The arguments with their contexts (assuming τ_{min}*) in* $Args(CAT)$ *are:*

- $\mathcal{A}_1 : a_1,\ \tau_A(\mathcal{A}_1) = .2$
- $\mathcal{A}_2 : a_2,\ \tau_A(\mathcal{A}_2) = .5$
- $\mathcal{A}_3 : b_1,\ \tau_A(\mathcal{A}_3) = .6$
- $\mathcal{A}_4 : b_2,\ \tau_A(\mathcal{A}_4) = .2$
- $\mathcal{A}_5 : c_1,\ \tau_A(\mathcal{A}_5) = .4$
- $\mathcal{A}_6 : \mathcal{A}_1, \mathcal{A}_2 \rightsquigarrow a,\ \tau_A(\mathcal{A}_6) = .2$
- $\mathcal{A}_7 : \mathcal{A}_1 \rightsquigarrow x,\ \tau_A(\mathcal{A}_7) = .2$
- $\mathcal{A}_8 : \mathcal{A}_2 \rightsquigarrow y,\ \tau_A(\mathcal{A}_8) = .6$
- $\mathcal{A}_9 : \mathcal{A}_3, \mathcal{A}_4 \rightsquigarrow b,\ \tau_A(\mathcal{A}_9) = .2$
- $\mathcal{A}_{10} : \mathcal{A}_5 \rightsquigarrow c,\ \tau_A(\mathcal{A}_{10}) = .4$

Assume that we wish to achieve $CAT \dot{-} \{A_6, A_{10}\}$ *using only premise-based revision, that is we ignore, for the time being, preferences on arguments and formula revision on them.*

Let us consider the removal of the premises of $\mathcal{A}_5$*,* a_1 *and* a_2*, the dropped arguments* (Δ_A) *and the values of* V_{ctx} *are shown in the table below:*

ϕ	Δ_A	V_{ctx}
a_1	$\{\mathcal{A}_1, \mathcal{A}_6, \mathcal{A}_7\}$	.6
a_2	$\{\mathcal{A}_2, \mathcal{A}_6, \mathcal{A}_8\}$	1.3

Let us now consider preference-based revision. Notice that given their context values — and the consequent values in $pref(CKB)$ *— it holds that* $A_9 \prec A_{10}$*. However, consider the argumentation theory* $\mathcal{AT}\prime$ *obtained by removing the formula* $\underline{b_2} <_K \underline{c_1}$ *and adding* $\underline{b_2} >_K \underline{c_1}$ *Let* $\phi_1 = \underline{b_2} <_K \underline{c_1}$ *and* $\phi_2 = \underline{b_2} >_K \underline{c_1}$*, we have:*

$$Violations_{SD}(\mathcal{AT} - \phi_1 + \phi_2) = \{\underline{\mathcal{A}_4} \prec \underline{\mathcal{A}_5}, \underline{\mathcal{A}_9} \prec \underline{\mathcal{A}_{10}}\}$$

$$Violations_{CTO}(\mathcal{AT} - \phi_1 + \phi_2) = \{(\mathcal{A}_4, \mathcal{A}_5), (\mathcal{A}_9, \mathcal{A}_{10})\}$$

We can see that $V_v(\mathcal{AT}, \mathcal{AT} - \phi_1 + \phi_2) = 4$

In the following I will describe how the cost functions above can be used to reduce impact on AR in Contextualized Argumentation Theories.

6.4 Limiting impact with context

In his work [Sna12] Snaith proposes *minimal change*, as inspired by traditional AGM Theory, as a way to guide the argument revision process. His concern is with a measure of minimal change that would take into account only structural and acceptability properties of arguments. One of the assumptions of this work is that external preferences are available from the agents' setting and that they should be used as a form of entrenchment for arguments and should be used to guide agents' reasoning with argumentation; thus in the following paragraphs I will describe how context can be used to limit impact during Argument Revision. Once again, the discussion will focus on trust as a context, but this choice is mostly for sake of concreteness and the same observations can be applied to other settings. I describe two different approaches: thresholds and budgeting. Thresholds are simple way to constrain the types of possible changes (i.e. which formulae we may add or remove) to go from $\mathcal{AT}$ to $\mathcal{AT}^{\pm}_{S}$. Budgeting uses the functions defined in the previous subsection to limit the total cost of a change from $\mathcal{AT}$ to $\mathcal{AT}^{\pm}_{S}$. Thus, thresholds limit impact by looking at *local* changes (edges in the change graph) whereas budgeting consider *global* changes (paths in the change graphs).

6.4.1 Thresholds

As we saw in Subsection 6.3, information with high preferences can be lost during the Argumentation Revision process. This is due to the fact that removal of formulae does not discriminate between more and less trustworthy ones. An approach to reduce loss of trustworthy information when removing formulas is the use of *thresholds*. These thresholds can be used to *protect* trustworthy arguments from being ruled out from the Argumentation Theory identifying a basic level of trust above which a formula should not be removed from the Argumentation Theory. On the other hand, in the setting of formula addition, thresholds can be used to guarantee that only arguments that are trustworthy enough enter the Argumentation Theory.

A threshold can be defined by any bound on the context values.

Definition 61 (Thresholds for argument revision)**.** *Let $t \in [0,1]$, τ a context function and $(\tau_A, \mathcal{AT})$ a Contextualized Argumentation Theory (where $\tau_A = \mathcal{T}(\tau)$ for some transformer $\mathcal{T}$) we say that an argumentation revision process:*

- *t-forbids formula addition if for any change $\mathcal{AT} + \phi$ it must hold that:*
    - *$\tau(\psi) \geq t$ when $\mathcal{AT} + \phi$ is a premise-based AR change (where ψ is defined as in the first bullet of Definition 57);*
    - *$\tau(r) \geq t$ when $\mathcal{AT} + \phi$ is a rule-based AR change (where ψ is defined as in the fourth bullet of Definition 57);*

- *t-forbids formula removal if for any change $\mathcal{AT} - \phi$ it must hold that:*
  - $\tau(\psi) < t$ *when $\mathcal{AT} - \phi$ is a premise-based AR change (where ψ is defined as in the first bullet of Definition 57);*
  - $\tau(r) < t$ *when $\mathcal{AT} - \phi$ is a rule-based AR change (where ψ is defined as in the fourth bullet of Definition 57).*

The definition above induces a subgraph of the change graph where all the edges respect the conditions above.

There are two issues that arise by the use of thresholds. First, fixed a certain threshold, it is not guaranteed if it is possible to reach a desired expansion or contraction given that threshold. Second, it would be useful to be able to measure the impact of a certain threshold, thus evaluating its "quality", e.g. how much advantage an agent obtains from the information it is protected from.

One shortcoming of thresholds is that they impose a sharp separation between "good" and "bad" information. This emerging separation may be hard to tune or estimate and it may create a systematic bias in agents that would rule out always the same type of information; in some settings this may not be desirable, e.g. with trust, where this would lead to excluding always the same sources. The next approach, budgeting, partly solves these problems.

6.4.2 Budgeting

In this section I will describe how to use *budgeting* to guide the argument revision process with context. Budgeting is a technique described in [PSM12] to deal with Trust in argumentation theories. The basic idea is to allow a certain amount of total trustworthiness to invest during the process of change. The original idea by Parsons et al. is used in a different setting[13] and is here adapted for our purposes.

Each agent has a certain budget $\mathcal{B} = (B_v, B_{ctx})$ which represents a pair of values. Informally this pair of values is an upper bound of the cost vector CV defined previously. We will use the notion of Pareto domination [SLB09].

If we have an Argumentation Theory $\mathcal{AT}$ to which we are applying Argument Revision, let $\mathcal{AT}'$ be the resulting Argumentation Theory resulting from the adding or removal of formulas. We consider $\mathcal{AT}'$ to be *in budget* if $\mathcal{B}$ Pareto-dominates the cost vector CV:

Definition 62 (Budgeting for argument revision)**.** *Let $(\tau_A, \mathcal{AT})$ be a Contextualized Argumentation Theory, $\mathcal{AT}' \in \mathcal{AT}^{\pm}_{S}$, $\mathcal{B} = (B_v, B_{ctx})$ be a budget, $CV = CV(\tau_A, \mathcal{AT}, \mathcal{AT}')$ we say that $\mathcal{AT}'$ is a revision in budget $\mathcal{B}$ for $\mathcal{AT}$ if it holds that*

$$B_v \geq CV_v \wedge B_{ctx} \geq CV_{ctx} \wedge (B_v > CV_v \vee B_{ctx} > CV_{ctx})$$

[13]In [PSM12] budgeting is applied during reasoning with arguments that have trust information rather than to modifications of Argumentation Theories

To some extent this approach is an extension of thresholds; an important difference is that trust threshold rules out information from agents below a certain level, whereas budgeting is more flexible. One example where this approach may be useful is when that may be worth including or not excluding (respectively if we consider formula addition and removal) are close to a certain threshold level.

When discussing thresholds in the previous paragraphs I observed how it can be hard to know in advance useful thresholding values or how to tune these values; budgeting presents similar problems. Another limitation of budgeting is that it may be not suitable for all types of metalogical information. Some context functions may not provide much information when accumulated or budgeted, for example because they have no upper bound. One such example is the freshness context function seen in preliminary sections.

In general, how well-behaved the notion of budgeting is in settings of argumentative or trusted agents is still subject of research [PSM12, DHM$^+$11].

6.5 Applications

In [Sna12] Snaith presents possible applications of argument revision for strategical formula retraction and lying in dialogues. In the next paragraphs I will discuss how the extensions to AR described in Subsection 6.3 may improve the strategical capabilities of agents in these dialogical settings.

In argumentation dialogues, it is common to speak of agents incurring *commitments* to their assertions [MP09]. The various commitments of the participants are usually tracked in a publicly-readable database, called a commitment store. For example, after an agent claims the arguments $(\{\phi\}, \phi)$ and $(\{\psi, \psi \rightsquigarrow \phi\}, \phi)$, its commitment store may be $CS = \{\phi, \psi, \lceil \psi \rightsquigarrow \phi \rceil\}$. In certain dialogues, an agent may be forced to retract an assertion if the latter cannot be defended against a counter-argument from its opponent. If we assume that a commitment store is closed, however, retracting the statement alone may be insufficient. In fact, they may still hold commitments to other statements from which the retracted one is a consequence (notice for example how removing ϕ would be insufficient in the commitment store above to retract ϕ). Walton and Krabbe [WK95] term the process of ensuring that a retracted statement can no longer be inferred in a commitment store a *stability adjustment*, and illustrate one possible method to achieve it. One limitation in their approach is that they do not explain how to choose between possible adjustments. Stability adjustments can be modelled by Argument Revision (we can contract by the arguments having the formula to retract as a conclusion) and by using the minimal impact functions in [Sna12] on the different revisions we have a method to choose among the possible adjustments.

Lying is the process of uttering a statement believed to be false [SCH10]. There are several reasons an agent may lie during a dialogue [Sna12], the scenario I will focus on in this discussion is when an agent lies because it possesses no acceptable argument that defeats a previous argument from its opponent. Snaith [Sna12] proposes to use argument revision to model lies: through uttering a lie, the speaker must in effect, for the purposes of the dialogue at hand, update its beliefs to accommodate it, otherwise it risks being exposed. Hence making up a lie can be seen as a process of Argument Revision. Moreover, we can use definitions of minimal impact to define a lie which is as small as possible[14].

Above I explained how Argument Revision may enable agents to choose which retraction to carry out or which lie to utter. Besides, Argument Revision may provide tools to strategical agents to choose between honestly retracting or dishonestly claiming something they do not believe as an alternative. Consider for example the case of an agent that cannot defend a counterargument from an adversary; if not necessarily honest the agent may want to go for covering its argument with a lie rather than retracting it, to the agent, if that is the best course of action. The "best course of action" can be decided by looking at the minimal impact in the argumentation theory. In [Sna12] impact is defined as a measure of the structural and acceptability changes on the argumentation theory (how many arguments have been dropped after the change, how many of these were acceptable...). The extensions of Argument Revision presented in Subsection 6.3 allow to consider impact of retraction and lying based on meta-logical information. Moreover, the techniques presented in the last subsection allow agents to have more sophisticated strategies, since they are may be provided with reference parameters to use in their choices (respectively thresholds and budgets).

Consider an agent carrying out a dialogue and having a trust context function on the information it reasons with. The agent has just been asked to make its commitment store consistent. We may assume its strategy does not demand full honesty, thus the agent contemplates the possibility of lying rather than retracting. However, in either choice it may want to have available trustworthy arguments: if it is honest this will allow it to make a more cogent case [15]; if it decides to lie it would still want to to have available highly trustworthy arguments, so that the agent can limit the use of lying in the future (spotting a lie has been examined as a research topic in the past [PS06]).

One possible way the agent can carry out this choice is by imposing a threshold t and *t-forbid* formulas removal of agents during the revision for retraction. If such a revision were possible that would allow the agent to retract without sacrificing any argument with a trust level below t. If this were not the case the agent may fall back on lying. Alternatively, it may use a budget $\mathcal{B}$ on the revision for retraction and try to limit the trust levels and the violations of properties of effective Contextualized Argumentation Theories in the process, the agent might rely on these properties as part of his strategies. These two potential scenarios describe how thresholds and budgeting may be used to select the best course of action between retracting and lying, but notice that the agent may also refine the process by which it chooses a lie to utter by imposing thresholds or budgets on the corresponding revision process.

[14] Keeping small lies has been proposed in the logical account of lying in [SCH10] according to which small lies are easier to maintain.

[15] I assume that, in this specific scenario, a high trust level of an argument is a plausible heuristic for estimating its potential cogency.

6.6 Summary and Contributions to AR

In this section I described Snaith's Argument Revision and compared it to Contextualized Revision, presenting a formalism that connected the two. Contextualized revision is focused on propagating metalogical information during reasoning, whereas Argument Revision has the goal of modifying Argumentation Theories to ensure acceptability (or inacceptability) of arguments. In the original approach by Snaith, Argument Revision, as a process for modifying argumentation theories, is guided by notions of minimal change. These notions, inspired by traditional AGM, are formalized by looking at structural and semantic changes in the resulting Argumentation Theories (gain/loss of arguments and gain/loss of acceptable arguments). In certain settings however agents may benefit from using contextual values of information for Argument Revision. The techniques presented here may be used to define strategic policies in dialogues, for example employing commitment retraction policies and lying.

This section also contained several contributions to the original definition of Argument Revision. I introduced a two-level variant of the definition of meta-argumentation system in [Sna12] basing it on [vdW11] , resolving some issues unclarified in the original work by Snaith — such as the transitivity of the predicate $\preceq$ at the meta-level. I also made more precise the definition of extensions used to define argument expansion and contraction by proposing to use argumentation theories based on meta-argumentation in [vdW11]. I formalized the types of changes in an argumentation theory (premise, rule, preference and contrariness-based argument revision), presented only informally in [Sna12]. Differently from the original work in [Sna12], I propose a definition of expansion and contraction that ensures that the resulting argumentation theories are well-formed (these properties are ensured as a consequences of using the meta-argumentation approach in [vdW11]) and that enables the ordering at the resulting object level argumentation theory to be determined by justified conclusions at the meta-level. The latter point represents a significant enhancement to the original definition of Argument Revision since in [Sna12] orderings are not guaranteed to be admissible and are determined only by the existence of arguments in the argumentation theory (similarly to current contrariness and rule-based changes) rather than by argumentation semantics.

7 Conclusions

7.1 Summary

In this thesis I proposed an abstract operator for updating argumentation theories with arguments with contextual information (contextualized revision). I discussed desirable properties for arguments with context in an argumentation theory and formalized these properties in a framework that combines an ASPIC+ argumentation theory with contextual knowledge: Contextualized Argumentation Theories. I presented an abstract operator of contextualized revision where revising the system with an argument together with a context value makes the following invariants hold:

- arguments are protected from others with lower context;
- arguments are enabled to defeat others with lower context;
- enabling an agent to recall all the arguments he acquired;

Then I presented Contextualized Knowledge Bases, a simple system to update ASPIC+ argumentation systems and knowledge bases. I proved that, under certain circumstances, Contextualized knowledge bases allow to specify Contextualized Argumentation Theories with the properties above, i.e. effective Contextualized Argumentation Theories. I described plausible ways we can build context values on arguments upon context values on their specific elements, i.e. formulae. I generalized these notions by operators called context transformers.

I showed how Contextualized Argumentation Theories can be used in dialogical settings and exemplified a meta-argumentation theory to express changes on context itself.

I showed how the properties of Contextualized Argumentation Theories can be used for other approaches to Belief Revision and Argumentation, in particular I proposed a new form of minimal change measures based on contexts in Argument Revision employing techniques from literature on Argumentation and Trust [PSM12]. These ideas use an enhanced formalization of the Argument Revision framework originally described in [Sna12].

7.2 Answers to the Research Questions

In the following paragraphs I answer the Research Questions of Subsection 1.3.

RQ 1: How can we define an abstract formal system for properties of dynamics of arguments with contextual information?

RQ 1.a: What is the formal definition of a system able to express arguments, contextual properties and constraints on these?

In Section 3 I defined Contextualized Argumentation Theories as pairs of a context function and Argumentation Theories. A context function expresses metalogical information on arguments. We are able to express connections between metalogical and logical properties of arguments by defining properties on arguments in a Contextualized Argumentation Theory.

RQ 1.b: What is the formal definition of an abstract operator for describing the acquisition of arguments with a certain context?

In Section 3 I defined an abstract operator $\otimes$ for *contextualized revision* that takes a context value and an argument and revise a Contextualized Argumentation with that argument and context. Dynamic properties of Contextualized Argumentation Theories can be expressed as properties of the abstract operator.

RQ 1.c: What general properties should we require for the arguments in the system?

Always in Section 3, I proposed two simple properties that respectively enable arguments with a strong context to defeat arguments with lower context and arguments to be protected from defeat from weaker arguments. A third property was proposed that guarantees that the properties above hold without arguments being removed from the argumentation theory so that agents may always make use of arguments they acquired in the past. These properties make no assumption on the underlying metalogical information.

RQ 1.d: How can we formally describe these properties in the system?

The properties above were formalized as properties on arguments in a Contextualized Argumentation Theory and as invariants of the operator $\otimes$ (see Subsection 3.3). I defined effective Contextualized Argumentation Theories as those respecting these properties.

RQ 2: How can we define context on arguments?

This question was answered using specific aspects of ASPIC+.

RQ 2.a: On which elements of arguments can we assume context is defined?

In Subsection 3.4 I observed how metalogical information may be aggregated from the structure of arguments in ASPIC+, in particular I used premises and rules in arguments (that I generically call formulae). I proposed that we assume context on these elements and use it to build context on arguments by means of tranformers (see *RQ 2.b*). This assumption (which is also used in [PTS+11]) was inspired from the fact that several dialogue protocols allow utterances of formulae [Pra06].

RQ 2.b: What functions to use to aggregate context of arguments from different elements that are general with respect to the nature of the metalogical information?

I proposed two types of context transformers (see Section 3.4), $\mathcal{T}_{min}$ and $\mathcal{T}_{hom}$ for aggregating contexts on formulae. These operators combine contextual information following the structure of ASPIC+ arguments, that is they allow to define a context that is compatible with admissible orderings (see Section 4). These simple transformers makes no assumption of the nature of the context and capture the intuitions respectively for which arguments should be assigned the possible lowest context and arguments having subarguments with similar contexts may assume the context of their subarguments.

RQ 3: How can we define a way to modify an agent's knowledge base and an argumentation system so that the properties described in the abstract system of RQ 1 are respected?

In Section 4 I described Contextualized Knowledge Bases and their update operator *eval*. I used the latter — based on a simple way to manipulate the preferences of premises and rules in ASPIC+ argumentation systems — to obtain effective Contextualized Argumentation Theories under certain assumptions depending on the type of arguments involved (strict or defeasible) and the transformer used. These formal results offer an analysys of update of argumentation systems and its effect on argumentation theories in ASPIC+.

RQ 4: Can this system be used in realistic settings of agent interaction?

RQ 4.a: Can this system be concretely used in dialogues?

RQ 4.b: Can the system model changes in the context values due to revision itself?

In Section 5 I described how to combine Contextualized Knowledge Bases and the meta-argumentation approach in [vdW11] for observing dialogue moves. I showed how this allows to use contextualized revision to reason during dialogue settings when the context depends on the acquired arguments themselves.

RQ 4.c: Are the properties defined in RQ 1 flexible enough to formalize and implement useful properties in dialogues or other agent settings?

In Subsection 5.4 I described an enhancement formalization of Dialogical Trust-based Agents where agents could be protected from naive manipulations of trust-levels.

In Section 6 I described an enhancement of the original framework of Argument Revision in [Sna12] and showed how the properties of effective Contextualized Argumentation Theories can be used to define minimal impact during the Argument Revision process. I suggested that this may help defining domain specific entrenchment and thus make Argumentation Revision more effective in some of the settings described in [Sna12], such as commitment retraction and lying.

7.3 Discussion and future work

Multi-Context Argumentation Theories and integration with other preferences

The Contextualized Argumentation Theories presented in this work allow an argumentation theory to be revised with only one context function. Thus an agent reasoning with the corresponding argumentation theory is dealing with only one type of meta-logical information at a time. In realistic settings agents may need to consider arguments that have more than one type of contextual information at the same time [Pag04]. Therefore contextualized argumentation theories with more than one context should be explored. Technically, one potential way to define them without significant changes to the existing system would be the following: given n context functions on formulae $Ctx^1, ...Ctx^n$ and the respective context functions on arguments $Ctx^1_A, ...Ctx^n_A$ we can build an aggregate context on formulae Ctx' and on arguments Ctx'_A as respectively a combination of all the Ctx^i and Ctx^i_A. Now we can define a new Contextualized Argumentation Theory $(Ctx'_A, \mathcal{AT})$. An important desirable property would be that if the CATs $(Ctx^1_A, \mathcal{AT}), ...(Ctx^n_A, \mathcal{AT})$ are all effective, then also the new aggregate CAT $(Ctx'_A, \mathcal{AT})$ is effective. I worked on one instance of this approach[16]: if all the Ctx^i_A are defined by $\mathcal{T}_{min}$ we can take the mininimum among the Ctx^i values and define $Ctx'_A = \mathcal{T}_{min}(Ctx')$. I proved that, with these definitions, the CAT $(Ctx'_A, \mathcal{AT})$ is effective if all $(Ctx^i_A, \mathcal{AT})$ are effective. One potential drawback of this approach is that in some cases some of the specificity of the meta-logical information may be lost if combined with others; as a consequence, properties such as screened defeat can lose significance.

A well-behaving system of effective Contextualized Argumentation Theories with more than one context may also overcome part of the limitations of some contexts on defeasible arguments, for example by combining (in the trust case) τ_{min} and τ_{hom} (see following paragraphs for a more detailed discussion on contexts and defeasible arguments).

Context functions for defeasible argumentation

One of the reasons for the choice of ASPIC+ as a framework was to overcome some of the limitations of the argumentation system in [PTS+11], which allows only strict inference. In Section 4 I showed that if we define context only on homogeneous arguments we can use CKBs to have effective contextualized argumentation theories with defeasible arguments. I also showed that this is not true for contexts on arguments obtained with the transformer $\mathcal{T}_{min}$. The latter limitation raises the question of whether it is possible to use $\mathcal{T}_{min}$ to build effective CATs. This would require further work since, as Example 2 shows, the current definition of contextualized knowledge bases does not allow this straightforwardly. This happens because weakest-link and last-link in ASPIC+ carry out different preference comparisons between premises and (defeasible) rules, whereas the simple approach in CKBs loses every information on the formula from which the context on argument derives (i.e. which formula had the minimum context value). One potential solution may investigate more sophisticated versions of $\mathcal{T}_{min}$ that, still capturing the intuition of taking the weakest context, accounts for differences between rules and premises.

Since $\mathcal{T}_{hom}$ is the only transformer we can use for effective contextualized revision with defeasible arguments, another question is how limiting it is for

[16]This work was not described further in this thesis for time constraints

an agent to carry out revision only by homogenous contexts. Let us ask the question: how expressive are homogeneous arguments? The intuition behind two homogeneous arguments A, B being such that $Ctx(A) < Ctx(B)$ is that the subarguments in A and B all have quite the same context strength; also, this context strenght is lower for A's subarguments (or, more specifically, formulae) than for B's. In Section 3 the latter intuition was formalized in τ_{hom} by strictly imposing equality on contexts on formulae. However, this definition can be extended by considering any equivalence relation among context values. One example of this would be trust values that are in the same range. This point also shows how effective Contextualized Argumentation Theories can be flexible and how the requirement of real numbers as a codomain of the context functions may be easily adapted to different settings.

The ways CAT propagate properties on metalogical information on properties of arguments require further investigation. The work in this thesis tried to contribute to theoretical research on connections between properties of multi-agent systems and argumentation. The connections between the metalogical information of the system and the agent's arguments are defined by *(i)* context tranformers and *(ii)* desired properties of CAT. Context tranformers gather metalogical information on formulae in the arguments and synthesize them in a (numerical) parameter on the argument itself. Given these parameters we can check certain desired properties on argumentation theories. The transformers proposed in this thesis were simple; more sophisticated ways to synthesize context functions can be considered, such as those in [Pag04]. One possible direction is exploiting the structure of arguments defining context on subarguments too; simple properties of this type have been explored in [PTS+11] (see Subsection 1.1.1). The transformers in this work are also general, making no assumption on the arguments they will be applied on. Hopefully this can make them more widely applicable; however further work may aim at investigating transformers suitable for specific domains.

Applicability to MAS and experimental work on Argumentation-Based Agents

In Sections 5 and 6 I outlined the connections between Context Revisions and dialogues. In Section 5 I proposed that meta argumentation may be used to enable agents to reason about dynamic contexts by means of a meta-argumentation system to observe dialogue moves and whose object level argumentation system is derived from a contextualized knowledge base. In Section 6 I extended Snaith's Argumentation Revision approach to include measure of minimal impact based on context. I suggested that this extension can let agents make use of a richer entrenchment for Argument Revision.

The fact that the revision operator $\otimes$ takes a pair (Argument, Value) is an useful abstraction if we want to focus on the context value of arguments and their properties in the argumentation theory. Later on in this thesis I proposeed transformers to allow constructing context for arguments out of contextual values of their elements. Although this further layer of abstraction helps in expressing more complex features of the MAS setting, I have not illustrated how we may use transformers in a setting where agents receive one formula at the time (such as several dialogical protocols) but they still have to revise with whole arguments. One way to achieve this would be to define a less abstract[17] operator on pairs of formulae and context values that would generalize Contextualized Argumentation Theories and their properties. This or other potential formalizations should be explored to allow wider applicability.

In this thesis, the connection between dialogues and contextualized argumentation theories were outlined using a simplified dialogue protocol that has *claim* as only location type. Richer dialogue protocols [Pra05] may have more locution types, e.g. the *why* locution by which agents can ask for supporting arguments. I propose that further work would investigate the use of effective Contextualized Argumentation Theories in this richer setting to help define useful dialogue policies. Informally, a dialogue policy is a function from the state of a dialogue to the set of possible next moves. For example, agents may use a policy to decide whether it is the case to attack or accept arguments from other agents and, in the former case, which arguments to build as replies. Context functions that associate values according to the dialogue structure may be used to define heuristics or to enrich existing dialogue policies, such as those in [Dij12]. Another example of contexts to define heuristics for dialogical agents would be functions that exploit the subjective mapping proposed in [vdWD12] or the structure of the dialogical tree [Pra05]

Finally, experimental work would help validating how useful effective CATs can be to improve the performance of argumentative agents. In [KMPV12] Kok et al. proposed an experimental framework for the evaluation of agents that make use of structured argumentation in dialogues. For generating scenarios, the authors consider several quantitative input parameters. Contextualized argumentation theories are suitable to an experimentation setting since they are easily parametrizable. Examples of parameters are:

- adopted context functions;
- context transformers;
- threshold and budget levels (for the Argument Revision extensions).

Thus effective CATs may be used in [KMPV12] to generate more complex and realistic scenarios.

Enhancement and theoretical validation of applications to dialogues

At the end of Section 5 I formalized agents that are protected against naive manipulation by claims of other agents; this thread of the work in this thesis is close to the research in [PSM12]. I suggested that the formalization I proposed may make agents skeptical towards other agents claiming more trust they deserve

[17] in the sense that it would make stronger assumptions on the structure of the arguments

(given the current information); my definitions also modelled a skeptical attitude towards agents talking about other agents' trustworthiness if these claims are not supported by a sufficient level of trust. In the future, these potential applications may be validated by connecting the notions I proposed to logical accounts of information acquisition and trust [Lia03]. Another direction may explore how to formalize and to which extent it is useful to let changes in trust levels retroactively be propagated; this may be pursued together with enhancing the current framework with a logical account of time. These potential lines of research suggest useful applications for Contextualized Knowledge Bases and Contextualized Argumentation Theories as formal abstractions for the relation between agents' reasoning capabilities (e.g. argumentation) and properties of agents related to their application domain (e.g. dialogues and trust).

References

[AC98] Leila Amgoud and Claudette Cayrol. On the acceptability of arguments in preference-based argumentation. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, pages 1–7. Morgan Kaufmann Publishers Inc., 1998.

[AGM85] Carlos E Alchourrón, Peter Gärdenfors, and David Makinson. On the logic of theory change: Partial meet contraction and revision functions. *Journal of symbolic logic*, pages 510–530, 1985.

[Bre01] Gerhard Brewka. Declarative representation of revision strategies. *Journal of Applied Non-Classical Logics*, 11(1-2):151–167, 2001.

[CRL00] Daniela V Carbogim, David Robertson, and John Lee. Argument-based applications to knowledge engineering. *The Knowledge Engineering Review*, 15(2):119–149, 2000.

[DHM+11] Paul E Dunne, Anthony Hunter, Peter McBurney, Simon Parsons, and Michael Wooldridge. Weighted argument systems: Basic definitions, algorithms, and complexity results. *Artificial Intelligence*, 175(2):457–486, 2011.

[Dij12] Pieter Dijkstra. *Automated support of regulated data exchange: a multi-agent systems approach.* PhD thesis, University of Groningen, 2012.

[Dun95] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial intelligence*, 77(2):321–357, 1995.

[FGKIS11] Marcelo A Falappa, Alejandro J Garcia, Gabriele Kern-Isberner, and Guillermo R Simari. On the evolving relation between belief revision and argumentation. *Knowledge Engineering Review*, 26(1):35–43, 2011.

[FH99] Eduardo L Fermé and Sven Ove Hansson. Selective revision. *Studia Logica*, 63(3):331–342, 1999.

[FKIS09] Marcelo Alejandro Falappa, Gabriele Kern-Isberner, and Guillermo Ricardo Simari. Belief revision and argumentation theory. *Argumentation in artificial intelligence*, pages 341–360, 2009.

[Gär88] Peter Gärdenfors. *Knowledge in Flux: Modeling the Dynamics of Epistemic States*, volume 66. MIT press Cambridge, 1988.

[HH11] LJ He and HK Huang. Trust and reputation mechanism in multi-agent system. *Computer Science*, 38(4):1, 2011.

[Hun13] Anthony Hunter. Modelling uncertainty in persuasion. In *Scalable Uncertainty Management*, pages 57–70. Springer, 2013.

[KMPV12] Eric M Kok, John-Jules Ch Meyer, Henry Prakken, and Gerard AW Vreeswijk. Testing the benfits of structured argumentation in multi-agent deliberation dialogues. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 3*, pages 1411–1412. International Foundation for Autonomous Agents and Multiagent Systems, 2012.

[Lia03] Churn-Jung Liau. Belief, information acquisition, and trust in multi-agent systems—a modal logic formulation. *Artificial Intelligence*, 149(1):31–60, 2003.

[Mak97] David Makinson. Screened revision. *Theoria*, 63(1-2):14–23, 1997.

[Mod09] Sanjay Modgil. Reasoning about preferences in argumentation frameworks. *Artificial Intelligence*, 173(9):901–934, 2009.

[MP09] Peter McBurney and Simon Parsons. Dialogue games for agent argumentation. In *Argumentation in artificial intelligence*, pages 261–280. Springer, 2009.

[Pag04] Fabio Paglieri. Data-oriented belief revision: Towards a unified theory of epistemic processing. In *Proceedings of STAIRS*, pages 179–190, 2004.

[PC05] Fabio Paglieri and Cristiano Castelfranchi. Revising beliefs through arguments: Bridging the gap between argumentation and belief revision in mas. In *Argumentation in Multi-Agent Systems*, pages 78–94. Springer, 2005.

[Pol87] John L Pollock. Defeasible reasoning. *Cognitive science*, 11(4):481–518, 1987.

[Pra05] H. Prakken. Coherence and flexibility in dialogue games for argumentation. *Journal of logic and computation*, 15(6):1009–1040, 2005.

[Pra06] H. Prakken. Formal systems for persuasion dialogue. *Knowledge Engineering Review*, 21(2):163, 2006.

[Pra10] Henry Prakken. An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1(2):93–124, 2010.

[PS06] Simon Parsons and Elizabeth Sklar. How agents alter their beliefs after an argumentation-based dialogue. *Argumentation in Multi-Agent Systems*, pages 297–312, 2006.

[PSM12] Simon Parsons, Elizabeth Sklar, and Peter McBurney. Using argumentation to reason with and about trust. In *Argumentation in Multi-Agent Systems*, pages 194–212. Springer, 2012.

[PTS+11] Simon Parsons, Yuqing Tang, Elizabeth Sklar, Peter McBurney, and Kai Cai. Argumentation-based reasoning in agents with varying degrees of trust. In *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*, pages 879–886. International Foundation for Autonomous Agents and Multiagent Systems, 2011.

[PWA03] S. Parsons, M. Wooldridge, and L. Amgoud. On the outcomes of formal inter-agent dialogues. In *Proceedings of the second international joint conference on Autonomous agents and multiagent systems*, pages 616–623. ACM, 2003.

[RMF+08] Nicolás D Rotstein, Martín O Moguillansky, Marcelo A Falappa, Alejandro Javier García, and Guillermo Ricardo Simari. Argument theory change: revision upon warrant. *COMMA*, 172:336–347, 2008.

[SCH10] Chiaki Sakama, Martin Caminada, and Andreas Herzig. A logical account of lying. In *Logics in Artificial Intelligence*, pages 286–299. Springer, 2010.

[SLB09] Yoav Shoham and Kevin Leyton-Brown. *Multiagent systems: Algorithmic, Game-theoretic, and Logical Foundations.* Cambridge University Press, 2009.

[Sna12] Mark Ian Snaith. *Argument revision and its role in dialogue.* PhD thesis, University of Dundee, 2012.

[SS05] Jordi Sabater and Carles Sierra. Review on computational trust and reputation models. *Artificial Intelligence Review*, 24(1):33–60, 2005.

[TSP12] Yuqing Tang, Elizabeth Sklar, and Simon Parsons. An argumentation engine: Argtrust. In *Ninth International Workshop on Argumentation in Multiagent Systems*, 2012.

[vdW11] Thomas L van der Weide. Arguing to motivate decisions. *SIKS Dissertation Series*, 2011, 2011.

[vdWD12] Thomas L van der Weide and Frank Dignum. Reasoning about and discussing preferences between arguments. In *Argumentation in Multi-Agent Systems*, pages 117–135. Springer, 2012.

[vdWDM+12] Thomas L van der Weide, F Dignum, J-J Ch Meyer, H Prakken, and GAW Vreeswijk. Multi-criteria argument selection in persuasion dialogues. In *Argumentation in Multi-Agent Systems*, pages 136–153. Springer, 2012.

[WK95] Douglas Neil Walton and Erik CW Krabbe. *Commitment in dialogue: Basic concepts of interpersonal reasoning.* SUNY press, 1995.